



University
of Glasgow

Mathis, Florian (2023) *Moving usable security research out of the lab: evaluating the use of VR studies for real-world authentication research*. PhD thesis.

<https://theses.gla.ac.uk/83723/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>
research-enlighten@glasgow.ac.uk

MOVING USABLE SECURITY RESEARCH
OUT OF THE LAB: EVALUATING THE USE
OF VR STUDIES FOR REAL-WORLD
AUTHENTICATION RESEARCH

FLORIAN MATHIS

SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF
Doctor of Philosophy



University
of Glasgow



THE UNIVERSITY
of EDINBURGH

SCHOOL OF COMPUTING SCIENCE
COLLEGE OF SCIENCE AND ENGINEERING
UNIVERSITY OF GLASGOW

JULY 2023

© FLORIAN MATHIS

Abstract

Empirical evaluations of real-world research artefacts that derive results from observations and experiments are a core aspect of usable security research. Expert interviews as part of this thesis revealed that the costs associated with developing and maintaining physical research artefacts often amplify human-centred usability and security research challenges. On top of that, ethical and legal barriers often make usability and security research in the field infeasible. Researchers have begun simulating real-life conditions in the lab to contribute to ecological validity. However, studies of this type are still restricted to what can be replicated in physical laboratory settings. Furthermore, historically, user study subjects were mainly recruited from local areas only when evaluating hardware prototypes. The human-centred research communities have recognised and partially addressed these challenges using online studies such as surveys that allow for the recruitment of large and diverse samples as well as learning about user behaviour. However, human-centred security research involving hardware prototypes is often concerned with human factors and their impact on the prototypes' usability and security, which cannot be studied using traditional online surveys.

To work towards addressing the current challenges and facilitating research in this space, this thesis explores if – and how – virtual reality (VR) studies can be used for real-world usability and security research. It first validates the feasibility and then demonstrates the use of VR studies for human-centred usability and security research through six empirical studies, including remote and lab VR studies as well as video prototypes as part of online surveys.

It was found that VR-based usability and security evaluations of authentication prototypes, where users provide touch, mid-air, and eye-gaze input, greatly match the findings from the original real-world evaluations. This thesis further investigated the effectiveness of VR studies by exploring three core topics in the authentication domain: First, the challenges around *in-the-wild shoulder surfing* studies were addressed. Two novel VR shoulder surfing methods were implemented to contribute towards realistic shoulder surfing research and explore the use of VR studies for security evaluations. This was found to allow researchers to provide a bridge over the methodological gap between lab and field studies. Second, the ethical and legal barriers when conducting *in situ usability research* on authentication systems were addressed. It was found that VR studies can represent plausible authentication environments and that a prototype's *in situ* usability evaluation results deviate from traditional lab evaluations. Finally, this thesis contributes a novel evaluation method to *remotely study interactive VR replicas of real-world prototypes*, allowing researchers to move experiments that involve hardware prototypes out of physical laboratories and potentially increase a sample's diversity and size. The thesis concludes by discussing the implications of using VR studies for prototype usability and security evaluations. It lays the foundation for establishing VR studies as a powerful, well-evaluated research method and unfolds its methodological advantages and disadvantages.

DECLARATION AND CONTRIBUTING PUBLICATIONS

The research presented in this thesis is entirely the author's own work. This thesis comprises several research papers (**Publication 1 to 9**) that were published at international venues:

The narrative of the thesis was published in Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR 2021) and presented and discussed at IEEE VR 2021's Doctoral Consortium and at SOUPS 2022 in form of a Lightning Talk.

[Publication 1] **Mathis, F.** (2021). [DC] VirSec: Virtual Reality as Cost-Effective Test Bed for Usability and Security Evaluations. In Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR 2021). IEEE, DOI: [10.1109/VRW52623.2021.00235](https://doi.org/10.1109/VRW52623.2021.00235)

[Publication 2] **Mathis, F.** (2022). Moving Usable Security and Privacy Research Out of the Lab: Adding Virtual Reality to the Research Arsenal. In Symposium on Usable Privacy and Security (SOUPS 2022) Lightning Talks. Usenix Association, URL: http://fmathis.com/publications/LightningTalk_FlorianMathis_SOUPS2022.pdf

The results of the expert interviews in Chapter 3 were published as a journal paper in the International Journal of Human-Computer Interaction (IHIC 2022).

[Publication 3] **Mathis, F.**, Vaniea, K., & Khamis, M. (2022). Prototyping Usable Privacy and Security Systems: Insights from Experts. In International Journal of Human-Computer Interaction. Taylor & Francis, DOI: [10.1080/10447318.2021.1949134](https://doi.org/10.1080/10447318.2021.1949134)

The research in Chapter 4 was published as a full paper in Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI 2021) and as a full paper in Proceedings of the International Conference on Academic Mindtrek (Mindtrek 2021).

[Publication 4] **Mathis, F.**, Vaniea, K., & Khamis, M. (2021). Replicueauth: Validating the use of a lab-based virtual reality setup for evaluating authentication systems. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI 2021). ACM, DOI: [10.1145/3411764.3445478](https://doi.org/10.1145/3411764.3445478)

[Publication 5] **Mathis, F.**, Vaniea, K., & Khamis, M. (2021). Observing Virtual Avatars: The Impact of Avatars' Fidelity on Identifying Interactions. In Academic Mindtrek 2021 (Mindtrek 2021). ACM, DOI: [10.1145/3464327.3464329](https://doi.org/10.1145/3464327.3464329)

The research in Chapter 5 was published as two full papers in Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR 2022).

[Publication 6] **Mathis, F.**, O’Hagan, J., Khamis, M., & Vaniea, K. (2022). Virtual Reality Observations: Using Virtual Reality to Augment Lab-Based Shoulder Surfing Research. In Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR 2022). IEEE, DOI: [10.1109/VR51125.2022.00048](https://doi.org/10.1109/VR51125.2022.00048) (**Best Paper Award Nominee**)

[Publication 7] **Mathis, F.**, Vaniea, K., & Khamis, M. (2022). Can I Borrow Your ATM? Using Virtual Reality for (Simulated) In Situ Authentication Research. In Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR 2022). IEEE, DOI: [10.1109/VR51125.2022.00049](https://doi.org/10.1109/VR51125.2022.00049)

The research in Chapter 6 was published as a full paper in Proceedings of the International Conference on Advanced Visual Interfaces (AVI 2022) and as a position paper at CHI’s first XR remote research workshop (CHI 2021).

[Publication 8] **Mathis, F.**, O’Hagan, J., Vaniea, K., & Khamis, M. (2022). Stay Home! Conducting Remote Usability Evaluations of Novel Real-World Authentication Systems Using Virtual Reality. In Proceedings of the International Conference on Advanced Visual Interfaces (AVI 2022). ACM, DOI: [10.1145/3531073.3531087](https://doi.org/10.1145/3531073.3531087)

[Publication 9] **Mathis, F.**, Zhang, X., O’Hagan, J., Medeiros, D., Saeghe, P., McGill, M., Brewster, S., & Khamis, M. (2021) Remote XR Studies: The Golden Future of HCI Research?. In CHI 2021 Workshop on XR Remote Research, URL: http://fmathis.com/publications/chi2021_workshop_remoteXR.pdf

Other Publications: The author of this thesis collaborated with researchers from the United Kingdom, France, Belgium, Germany, and Austria in several research projects. A selected list of publications is listed below (cf., [Google Scholar](#) for a full list). The research outputs of the collaborations are not part of this thesis, but they inspired some of the conducted PhD work, expanded the author’s knowledge, and kept him motivated throughout his journey.

Mansour, S., Knierim, P., O’Hagan, J., Alt, F., **Mathis, F.** (2023). BANS: Evaluation of Bystander Awareness Notification Systems for Productivity in VR. In Symposium on Usable Security and Privacy (USEC 2023). Internet Society, DOI: [10.14722/usec.2023.234566](https://doi.org/10.14722/usec.2023.234566)

Gruenefeld, U., Auda, J., **Mathis, F.**, Schneegass, S., Khamis, M., Gugenheimer, J., Mayer, S. (2022). VRception: Rapid Prototyping of Cross-Reality Systems in Virtual Reality. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI 2022). ACM, DOI: [10.1145/3491102.3501821](https://doi.org/10.1145/3491102.3501821) (**Honourable Mention Award**)

Mathis, F., Williamson, J. H., Vaniea, K., & Khamis, M. (2021). Fast and secure authentication in virtual reality using coordinated 3d manipulation and pointing. In ACM Transactions on Computer-Human Interaction (TOCHI). ACM, DOI: [10.1145/3428121](https://doi.org/10.1145/3428121)

ACKNOWLEDGEMENTS AND THANKS

The research in this thesis was funded by the University of Edinburgh and the University of Glasgow. Additional funding to disseminate this PhD research was received from the Scottish Informatics and Computer Science Alliance (SICSA), the European Institute of Innovation & Technology (EIT), the College of Science & Engineering, the Advanced Computing Systems Association (USENIX), and the National Science Foundation (NSF).

Below are some more personal words about the people who were part of my time in Munich, Glasgow, Edinburgh, Toronto, and Vorarlberg. Before moving on to a new chapter of my life, I want to say “Thank you!” to the people who have believed in me and supported me.

First and foremost, I want to thank those closest to me: **Reinelde, Gerhard, Patricia, Bettina, Hilda, and Benny**. You enabled me to pursue an academic career and have always supported me and my decisions. I also want to thank **Lydia** for motivating me, keeping me sane, and helping me through the last few months. Thank you for continuously making me a better human. I prefer to write the following few sentences in German.

Mama und Papa, ihr wisst beide, dass ich ohne eure Hilfe nie nach München, Aarhus, Glasgow und Toronto gehen hätte können. **Danke** für eure unendliche und unbezahlbare Unterstützung. **Danke**, dass ihr meinen Schwestern und mir immer alles ermöglicht habt. **Danke**, dass ihr mir den Weg, den ich von 2014 bis 2023 gegangen bin, geebnet habt und Teil meiner Träume und Reisen gewesen seid. Ich bin euch unendlich dankbar! **Danke** Patricia für deine Unterstützung, deine motivierenden Worte, und für alles was du und Bernd mir ermöglicht haben während meines Aufenthalts in Vorarlberg. Danke an euch alle, ohne euch wäre diese Arbeit nie zustande gekommen!

Working at the University of Glasgow and at the University of Edinburgh was an incredible experience. I want to thank all the fantastic people who created an amazing and welcoming work atmosphere and shared their knowledge with me. I will never forget my time in Glasgow and Edinburgh – I will always look up to you and learn from you. Thank you **Stephen, Euan, Mark, Julie, Joseph, Dong-Bach, Patrizia, Gözel, Shaun, and Yosuef** for all your support! I also want to thank my G132 office mates, **Siwei and Jingmin**, for creating a welcoming work atmosphere and spending time with me! Thank you also to **Gang, Graham, Ilyena, Jamie, Katharina, Kieran, Max, Melvin, Robin, Thomas, Vilma**, and many others for making my PhD journey enjoyable and fun! Of course, I also want to thank all my PhD friends from TULIPS: **Adam, Nadin, Mohammad, Nicole, Dilara, Kholoud, Sarah, Tarini, Duncan**, and many others. Your feedback and kindness tremendously helped me to shape my PhD research and get through the COVID-19 lockdowns – thank you! Thank you also to my Annual

Progress Review panel – **Dr Helen Purchase, Dr Ron Poet, and Dr Joseph Maguire** – for the interesting discussions and the feedback on my research.

I also want to thank my previous lecturers and mentors at LMU Munich, who introduced me to academic research and inspired me to pursue a research career. Thank you **Prof Dr Heinrich Hußmann, Dr Tobias Seitz, and Prof Dr Florian Alt!** I am deeply thankful for your mentorship throughout the many years in Munich. Thank you also to the Usable Security and Privacy group in Munich for making my research visit, the CHI 2022 conference trip, and the AVI 2022 conference trip a pleasant (and fun!) experience. Special thanks to **Sarah P., Lukas, Felix, Sarah D., Pascal, Yasmeen, and Heike.** I also want to say thank you to all my friends in Munich who made my life easier and with whom I shared exciting and fun experiences. It was always a pleasure for me to spend time with you! Special thanks to **Bill, Marco, Michi, Xuesong, and Fe.**

Furthermore, I want to thank all the people with whom I worked on various HCI, VR/AR, and USEC challenges. Thank you **John, Hassan, Adalberto, Jolie, Jamie, Uwe, Jonas, Stefan, Jan, Sven, Kieran W., Yomna, Axel, and Shady.**

I would also like to thank my thesis committee – **Prof Dr Enrico Rukzio, Dr Mathieu Chollet, and Dr Sean MacAvaney** – for the interesting discussions during the viva. Thanks also for travelling to Glasgow and reading my thesis!

Last but not least, I would like to thank my advisors, **Kami Vaniea and Mohamed Khamis,** for their continuous support and for giving me the freedom to shape my own research. I have learned so much about research and academia – thank you!

There are many more people who were part of my PhD journey and deserve to be mentioned. My personal webpage fmathis.com/thanks continues with extended acknowledgements.

To my parents, my sisters, my grandma, my girlfriend, and my cat. I love you.

EDUCATION USE CONSENT

*“The education of young people in science is at least as important, maybe more so, than the research itself.” – **Glenn T. Seaborg***

I hereby give my permission for this thesis to be shown and distributed to other students at, and beyond, The University of Glasgow and The University of Edinburgh.

Florian Mathis

Table of Contents

1	Introduction	11
1.1	Motivation	11
1.2	Thesis Statement	13
1.3	Main Contributions and Research Questions	14
1.3.1	Insights from USEC Experts: Challenges and Opinions	15
1.3.2	Validation of VR Studies for Usable Security Research	15
1.3.3	VR Studies' Potential for Usable Security Research	16
1.3.4	Synthesised Thesis Research	17
1.4	Thesis Walkthrough	17
1.4.1	Overview of Studies	19
2	Literature Review	22
2.1	Introduction of Literature Review	22
2.2	Human-Computer Interaction Research: History and Research Paradigms	24
2.2.1	Lab Studies in HCI	26
2.2.2	Field Studies in HCI	27
2.2.3	Online Studies and Online Surveys in HCI	28
2.2.4	Interview Studies and Focus Groups in HCI	30
2.2.5	Virtual Reality Studies in HCI	31
2.2.6	Summary	32
2.3	Usable Security Research: HCI's Sibling?	33
2.3.1	Lab Studies in USEC	34
2.3.2	Field Studies in USEC	35

2.3.3	Online Studies and Online Surveys in USEC	36
2.3.4	Interviews and Focus Groups in USEC	37
2.3.5	Summary	38
2.4	Virtual Reality: An Overview	39
2.4.1	In Contrast to HCI and USEC: Empirical VR Research	39
2.4.2	Prerequisites for Applied Empirical VR Research in USEC	41
2.4.3	Summary	45
2.5	Authentication Prototypes and User Evaluations	45
2.5.1	Authentication Research: An Overview	46
2.5.2	Design Space of Authentication Prototypes	47
2.5.3	Authentication Prototypes and Empirical Research	52
2.5.4	Research Challenges in the Authentication Research Field	53
2.5.5	Summary	57
2.6	Summary and Conclusions of Literature Review	57
3	Scoping the Problem: Key Challenges in Usable Security Research	61
3.1	Introduction	61
3.1.1	Chapter Structure	62
3.2	Methodology	62
3.2.1	Recruiting USEC Experts	62
3.2.2	Interview Structure	64
3.2.3	Research Approach and Data Analysis	65
3.2.4	Methodological Limitations	66
3.3	Results	67
3.3.1	Threat Modelling is not Straightforward	67
3.3.2	Prototyping USEC Research Artefacts	69
3.3.3	Sample Size and Selection Process in USEC	71
3.3.4	Evaluation Methodologies in USEC	73
3.3.5	USEC's Research Culture	75
3.3.6	Academia and Industry in USEC	78
3.4	Discussion	81

3.4.1	There is No One Best Way for Doing USEC Research	81
3.4.2	Selecting Sample Sizes in the Presence of Constraints	84
3.4.3	Problem-Scoping and Problem-Solving USEC Research	85
3.5	Chapter Conclusion	86
3.5.1	Research Question 1 (RQ ₁) and Ways Forward	87
3.5.2	Contributions	88
4	Validating the Use of VR Studies for Evaluating USEC Prototypes	90
4.1	Introduction	90
4.1.1	Chapter Structure	92
4.2	Ethics and Compensation	92
4.3	The Selection of the USEC Prototype	93
4.3.1	CueAuth and RepliCueAuth: An Overview	93
4.4	Overview of Studies	95
4.4.1	Overview: Usability Evaluation of RepliCueAuth	95
4.4.2	Overview: Security Evaluation of RepliCueAuth	96
4.5	Data Analysis of the Usability and Security Study	96
4.5.1	VR Usability and Security Study Data Analysis	97
4.5.2	Validation Data Analysis	97
4.6	Usability Evaluation of RepliCueAuth	97
4.6.1	Apparatus and Implementation	98
4.6.2	Methodology and Study Design	99
4.6.3	Demographics	100
4.6.4	Results	100
4.6.5	Discussion of RepliCueAuth's VR Usability Evaluation	106
4.7	Pre-Security Study: Defining the VR Avatar	109
4.7.1	Apparatus and Implementation	110
4.7.2	Methodology and Study Design	111
4.7.3	Demographics	113
4.7.4	Results	113
4.7.5	Discussion of the Pre-Security Study and Its Implications	116

4.8	Security Evaluation of RepliCueAuth	117
4.8.1	Methodology and Study Design	118
4.8.2	Demographics	119
4.8.3	Results	119
4.8.4	Discussion of RepliCueAuth’s VR Security Evaluation	127
4.9	Limitations	128
4.10	Chapter Conclusion	129
4.10.1	Research Question 2 (RQ ₂)	130
4.10.2	Research Question 3 (RQ ₃)	131
4.10.3	Ways Forward	132
4.10.4	Contributions	133
5	Using VR Studies to Advance USEC Research	135
5.1	Introduction	135
5.1.1	Chapter Structure	137
5.2	Ethics and Compensation	137
5.3	VR Studies to Augment Lab-Based Shoulder Surfing Research	138
5.3.1	Introduction	138
5.3.2	Authentication Scenarios, Apparatus, and Implementation	140
5.3.3	Methodology	143
5.3.4	Demographics	145
5.3.5	Results	145
5.3.6	Discussion	151
5.3.7	Limitations	155
5.3.8	Conclusion	155
5.4	VR Studies for Simulated In Situ USEC Research	157
5.4.1	Introduction	157
5.4.2	Methodology	159
5.4.3	Apparatus and Implementation	165
5.4.4	Demographics	167
5.4.5	Results	167

5.4.6	Discussion	175
5.4.7	Limitations	177
5.4.8	Conclusion	178
5.5	Chapter Conclusion	179
5.5.1	Research Question 4 (RQ ₄)	180
5.5.2	Ways Forward	183
5.5.3	Contributions	184
6	Remote Evaluation of Real-World USEC Prototypes Using VR Studies	186
6.1	Introduction	186
6.1.1	Chapter Structure	188
6.2	Ethics, Compensation, and Data Collection	189
6.3	Investigated USEC Prototypes and Context	189
6.4	Implementation and Apparatus	191
6.5	Methodology	194
6.5.1	Independent and Dependent Variables	194
6.5.2	Study Design and Task	195
6.5.3	Data Analysis	196
6.5.4	Demographics	196
6.6	Results	196
6.6.1	Input Times	197
6.6.2	Number of Corrections	197
6.6.3	Number of Incorrect PIN Entries	197
6.6.4	Perceived Workload (NASA-TLX) and User Experience (UEQ)	198
6.6.5	System Usability Scale (SUS)	199
6.6.6	Usability/Security Ranking and 5-Point Likert Scales	199
6.6.7	Sense of Presence (IPQ and TPI)	200
6.6.8	Semi-structured Interviews	200
6.7	Discussion	204
6.7.1	RVR ³ : A Complementary Research Method	204
6.8	Limitations	205

6.9	Chapter Conclusion	206
6.9.1	Research Question 5 (RQ ₅) and Research Applications	206
6.9.2	Contributions	208
7	Summary and Reflection on Thesis Research	210
7.1	Introduction	210
7.2	Central Contributions	212
7.2.1	Identify: Insights from USEC Experts	212
7.2.2	Validate: VR Studies for USEC Research	212
7.2.3	Advance: VR Studies' Potential for USEC Research	213
7.2.4	Synthesised Thesis Research: Research Recommendations	213
7.3	Limitations	220
7.3.1	The Focus on Authentication Research	220
7.3.2	The Impact of VR Technology on Research	220
7.3.3	The User Study Samples: How Representative Are They?	221
7.3.4	The Lengths of the User Studies	221
8	The End of a Journey: Conclusion, Future Research, and Final Remarks	223
8.1	Conclusion and Ideas for Future Work	223
8.1.1	Conclusion	223
8.1.2	Future Research Directions	224
8.2	A (More Personal) Final Thought: Are VR Studies Taking Over?	227
A	Supplemental Materials	230
B	Appendix for Chapter 3	232
C	Appendix for Chapter 4	239
D	Appendix for Chapter 5	304
E	Appendix for Chapter 6	370
	Bibliography	402

List of Figures

1.1	Overview of the research contributions in this thesis: Identifying the existing USEC Challenges, Validating the use of VR Studies for USEC, and Advancing USEC research using VR studies.	14
2.1	Common research methods in HCI, including virtual field studies.	25
3.1	Schematic figure of USEC research and its links to the real world. The figure is based on interviews with USEC domain experts.	86
4.1	Picture of the replicated real-world authentication system and a screenshot of the replication in virtual reality.	91
4.2	Detailed overview of the authentication system's input methods and the study environment in VR.	94
4.3	Physical surface in the real world that was location-mapped to the virtual screen in VR. The image shows the study setup.	98
4.4	Comparison of the entry times between the input methods, including touch, mid-air, and eye-gaze input.	101
4.5	Comparison of the participants' perceived workload when providing touch, mid-air, and eye-gaze input. Both the VR and the real-world data are visualised.	103
4.6	The participants' ratings of the input methods on various 5-point Likert scales from 1-Strongly Disagree to 5-Strongly Agree.	104
4.7	Overview of the avatars and a human in the real world used in the study in combination with touch, mid-air, and eye-gaze input.	110
4.8	The participants' interaction identification performance and the perceived workload when observing the VR avatars and the human in the real world performing touch, mid-air, and eye-gaze input.	114
4.9	Pictures taken by the participants that showcase some of their observation strategies.	125

5.1	Overview of the shoulder surfing scenarios used in the user study, including ATM authentication, pattern authentication on a smartphone, and PIN authentication on a smartphone.	139
5.2	Overview of the two authentication environments used in the study, including an ATM and a bus station to simulate real-world contexts.	140
5.3	Comparison of the IPQ scores and the NASA-TLX dimensions between the 2D, 3D, and VR observation method.	148
5.4	Observation positions of the participants when observing ATM authentications using the 3D and VR observation method.	152
5.5	Overview of the study environments in VR, including a more traditional laboratory environment and a close-to-reality ATM environment for more realistic in situ research.	159
5.6	Example of a ColorPIN input which was used during the study to introduce the participants to the prototype.	161
5.7	Overview of the authentication scenarios used in the study, including two in the real world and three in virtuality.	162
5.8	Hardware during the user study: a physical keyboard, a greenscreen, and a camera to blend the keyboard and the user’s hand into virtuality.	166
5.9	Comparison of the authentication times, the number of ColorPIN corrections, and the number of errors between the various environments simulated in the real world and VR.	168
5.10	Comparison of the participants’ sense of presence and perceived workload across the different authentication contexts and environments.	171
5.11	One-dimensional visualisation of the location of VR studies on a theoretical ecological validity continuum, ranging from controlled lab studies to more naturalistic field studies.	180
6.1	Overview of the study environment and the authentication prototypes, i.e., traditional 4-digit PIN, Hand Menu, Tap, and Glass Unlock, for the remote VR study.	187
6.2	First person perspective when interacting with the different authentication prototypes implemented for the remote VR study.	190
6.3	Aerial perspective of the VR 3D environment and the participants’ position when interacting with the prototypes and the ATM during the study.	192

- 6.4 Screenshots of the in-VR menu that guided the participants through the remote VR study. 193
- 6.5 Comparison of the participants' user experience across the different authentication prototypes. 198

- B.1 Screenshot of the anonymised interview request sent to the experts. 237

- D.1 Observation positions of the participants when observing PIN authentications using the 3D and VR observation method. 350
- D.2 Observation positions of the participants when observing pattern authentications using the 3D and VR observation method. 350

I

INTRODUCTION

Chapter 1

Introduction

Begin at the beginning and go on till you come to the end; then stop.

– Lewis Carroll –

1.1 Motivation

More than 20 years ago, in 1999, “*Why Johnny Can’t Encrypt: A Usability Evaluation of PGP 5.0.*” by Whitten and Tygar [542] and “*Users are not the enemy*” by Adams and Sasse [9] beautifully demonstrated that security systems will not be secure if the usability of these systems is neglected or even ignored. Since the birth of the usable security and privacy (USEC) community in around 1995 [155, 568], researchers have argued that systems are only secure as long as users can (and know how to) securely interact with them. At the point where systems can be used in a more usable but less secure way than initially designed for, they will, sooner or later, intentionally or unintentionally, put the users’ security and privacy at risk. Security mechanisms that are confusing to people will be misused [568], and in the worst-case scenario, they will not find widespread adoption.

“A big lie of computer security is that security improves as password complexity increases. In reality, users simply write down difficult passwords, leaving the system vulnerable. Security is better increased by designing for how people actually behave.” – Jakob Nielsen [358]

As early as 1996, Zurko and Simon [568] emphasised the need to apply usability evaluations and techniques to secure systems, develop user-centred security for user-friendly systems, and consider user needs as a primary design goal instead of adding it as an additional layer afterwards. The USEC community has brought forth a plethora of novel usable, secure, and privacy-preserving artefacts that extend state-of-the-art and facilitate new insights. However,

despite the substantial body of USEC research and the academic, industrial, and societal interest in human-centred security and privacy, little progress has been made on fundamental security and privacy issues. There is still no silver-bullet replacement for passwords (and probably will not be soon [203]). Despite repeated and sustained efforts, phishing is still one of the most common security threats to UK organisations [73], affecting millions of users [524]. Many new technologies and novel threats to people’s privacy and security put even more pressure on human-centred security researchers [17]. Regardless of the USEC community’s success in the last decade, progress often feels slow:

“Issues of security and usability are no longer the province of military cryptographers but of software developers, system administrators, and the user community. Nevertheless, progress in usable security research and design has been slow, due in part to the need to master a large amount of (usually) mutually exclusive, yet necessary, skills and knowledge.” – Fléchais and Faily [137]

Whilst the USEC community has inherited many research methods from the broader Human-Computer Interaction (HCI) field, as the Literature Review in chapter 2 will show, USEC has methodological challenges that differ from those in other HCI subdomains. Security often plays a secondary role in reality [267], but recreating authentic use of USEC prototypes in the lab is challenging. Ethical and legal constraints further complicate (and sometimes even prohibit) detailed USEC research in the field [100, 529]. Traditional empirical research methods such as interviews, observations, and online studies are valuable for the USEC research community and have been successfully applied in the broader USEC field to study, for example, the users’ interaction behaviour on automated teller machines (ATMs) [100, 529], the strategies and decisions of users when encountering possibly suspicious emails [117], and the security of smartphone unlock PINs [298]. However, many of these research methods are unsuitable for adequately evaluating the usability and security of novel USEC research prototypes as they do not allow participants to interact with and experience the prototypes in real time. Hands-on experience with prototypes is vital as studies have found discrepancies between what people say and how they think and behave [159, 485].

An essential next step must be to facilitate and advance USEC research concerned with hardware prototype solutions and provide researchers with methods that allow them to bridge the methodological gaps between lab studies and more realistic field studies. The USEC community has already made use of online platforms such as Amazon Mechanical Turk [298], virtual study environments [210], or field observations [82, 100] to provide valuable insights into people’s security behaviour in realistic contexts. Whilst the challenges of conducting ecologically valid and impactful USEC research in real-world contexts have been recognised and are partially addressed by the community through, for example, role-playing scenarios to increase the realism of password studies [133, 140], simulating real-world contextual factors in

the lab [118], and accessing resources within a broader university infrastructure [82], research methods that combine the strengths of controlled lab studies with research in the field are still missing. To date, no research has attempted to employ technologies such as virtual reality (VR) to implement and advance the evaluation of real-world USEC prototypes.

The lack of research methods suitable for empirical evaluations of hardware USEC prototypes and the constant development of new ubiquitous technologies creates a need to adjust and advance USEC's methodological research arsenal [17]. Contextual factors in which USEC prototypes are evaluated, the users' interaction behaviour, and the realism of users' tasks are important factors that impact a prototype's usability and security evaluation results. As this thesis will show in chapter 3, *how* USEC researchers evaluate research artefacts is often not in line with reality. For example, the study environments often do not represent the actual usage scenarios in which USEC research artefacts are eventually deployed in the real world.

This thesis addresses the lack of empirical research methods suitable for evaluating real-world USEC prototypes, specifically in the authentication field, a major research domain in USEC [155] and the most addressed research topic in the broader USEC field [110]. Research on novel authentication prototypes often involves hardware prototypes (for example, [98, 323]) as well as complex study setups to assess a system's security (for example, [27, Fig. 3], [239, Fig. 2]), making authentication prototypes a suitable candidate for the first validation of the use of VR studies for USEC research. Therefore, this thesis considers the use of VR studies as a novel, well-evaluated research method for usability and security research on VR replicas of real-world USEC prototypes.

1.2 Thesis Statement

This thesis explores the suitability of virtual reality (VR) studies in supporting human-centred usability and security research. VR studies enable researchers to augment human-centred research methodologies that are constrained to conditions that can be physically replicated in the lab. This thesis presents a novel complementary research method for human-centred usability and security research by exploiting VR's characteristics to expand the possibilities of evaluating USEC prototypes. It first identifies existing research challenges in usable security (USEC) (**identify**, chapter 3). It then validates the use of VR studies for human-centred usability and security research (**validate**, chapter 4). This thesis concludes with investigations on how VR studies can augment and move USEC research out of the lab (**advance**, chapter 5 and chapter 6), unfolding the advantages and disadvantages of VR studies for human-centred usability and security evaluations of real-world prototypes.

1.3 Main Contributions and Research Questions

This thesis aims to balance the desire for the highest-possible-quality research with the existing research constraints. It advances human-centred usability and security research by facilitating prototype-driven USEC research using VR studies. It is important to highlight that the aim of this thesis is not to contribute an “*all-in-one*” empirical research method for the HCI and USEC communities. The research conducted here must be seen as complementary to the many existing research methods commonly applied in these areas, including interviews, surveys, focus groups, laboratory studies, field studies, and many more.

The research presented in this thesis, including the use of VR studies for simulating real-world research and the remote VR research method to move research on prototypes out of the lab, are valuable contributions to both the USEC and the HCI communities due to USEC’s interdisciplinary nature and the methodological links of this thesis to HCI and USEC. This thesis makes novel contributions to the human-centred usability and security research fields by applying a three-stage research arc (cf., Figure 1.1): It first unfolds the existing challenges when designing, implementing, and evaluating USEC prototypes (**identify**). It then validates the suitability of VR studies for substitutional real-world USEC research (**validate**). Finally, it showcases how VR studies contribute to advanced USEC research (**advance**). The three central contributions of this thesis are summarised in the next sections.

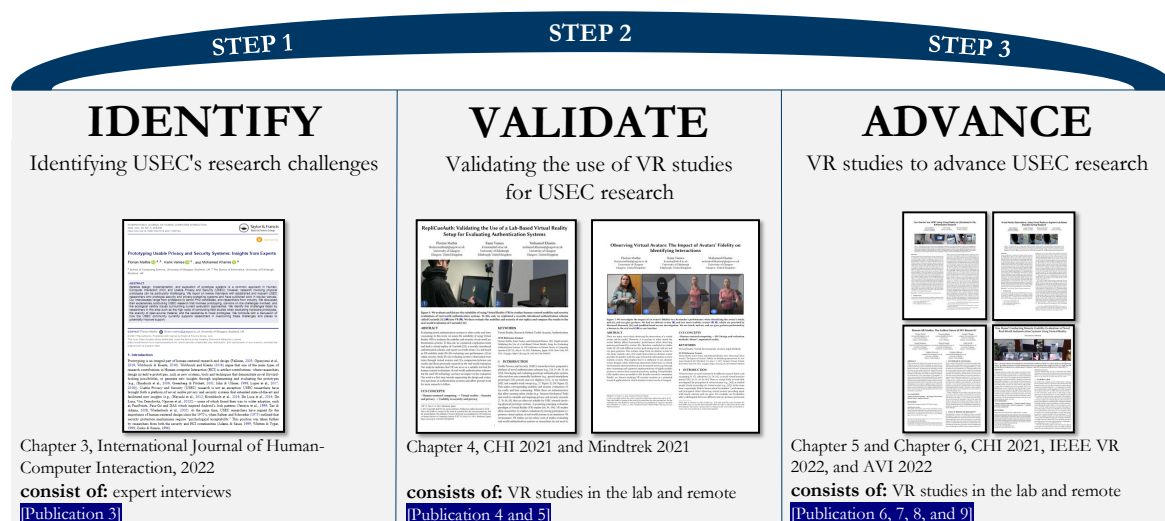


Figure 1.1: This thesis first identifies the challenges USEC experts experience when conducting USEC research involving prototypes. It then validates the use of VR studies for empirical USEC research. Finally, it demonstrates the potential of VR studies for USEC research through several user studies that augment USEC’s methodological research landscape.

1.3.1 Insights from USEC Experts: Challenges and Opinions

This work first contributes USEC expert interviews on the design, implementation, and evaluation challenges that experts face when conducting usable security and privacy research involving prototypes (cf., chapter 3 and RQ₁). Whilst other works have touched on what makes USEC research laborious in general, there has not yet been a structured attempt to elicit obstacles experienced by experts that use prototypes in their work. This thesis puts forward such a compilation of experienced challenges. The interviews provide, for the first time, insights into the USEC experts' opinions on the current research culture and the applied research methods within the broader USEC research field.

RQ₁ What are the challenges that USEC experts experience when designing, implementing, and evaluating security and privacy-enhancing prototypes?

1.3.2 Validation of VR Studies for Usable Security Research

Many of the challenges identified in the expert interviews, such as the lack of appropriate hardware, were considered critical obstacles in the broader USEC field. This thesis aims to address a subset of the challenges in subsequent empirical evaluations investigating the use of VR studies for substitutional real-world USEC research. Whilst all the key challenges identified in the interviews are essential to address, some of those require large-scale community input (cf., Key Challenge 1 in chapter 3) or highlight the lack of strong collaborations between industry and academia (cf., Key Challenge 9 in chapter 3), which are undoubtedly intriguing to tackle, but were outside the scope of this thesis.

The first VR study contributes a usability investigation of a real-world authentication prototype through a comparison to its real-world counterpart evaluation, providing answers to RQ₂. Two experiments then formulate an investigation of the representation of a user in VR that is required to allow for sequential empirical security evaluations and a validation of the use of VR studies for security evaluations of research artefacts, both leading to answers to RQ₃.

In summary, the usability and security studies in Section 4.6, Section 4.7, and Section 4.8 contribute the first empirical comparison of human-centred usable security research in VR and the real world, leading to the following two research questions:

RQ₂ Which findings of VR-based usability evaluations on USEC prototypes match the findings from corresponding evaluations in traditional physical lab settings?

RQ₃ Which findings of VR-based security evaluations on USEC prototypes match the findings from corresponding evaluations in traditional physical lab settings?

1.3.3 VR Studies' Potential for Usable Security Research

Two additional experiments contribute to the validation of using VR studies for real-world USEC research. They formulate novel research methods for 1) simulated shoulder surfing research in public, a common threat model when conducting security research on prototypes, and 2) simulated *in situ*¹ research on real-world prototypes. The focus on shoulder surfing in chapter 4 and chapter 5 should be seen as a sample demonstration of security evaluations in VR. Shoulder surfing was selected as security evaluation due to its common use when evaluating USEC prototypes (for example, [55, 56, 245, 530, 544]), and the impact it had on the USEC community as a whole, leading to a plethora of works on USEC prototypes that aim to mitigate shoulder surfing on situated displays [61, 101, 242], mobile devices [243, 402], and XR headsets [150, 158, 308], among others. Other security threats such as smudge attacks [28], thermal attacks [3], and guessing attacks [164] were beyond the scope of this thesis, but they surely deserve attention in the future.

Both studies in chapter 5 endorse the use of VR studies for USEC research and advance existing research methods. Whilst *in situ* studies, more formally known as “field studies”, are well known within the broader HCI community [258], research of this type is often time-consuming, expensive, and introduces ethical and legal challenges [100, 291, 529]. The two experiments in chapter 5 are the first empirical works that use VR studies for simulating real-world USEC research and aim at combining the internal validity of controlled lab experiments with the external validity of field experiments (RQ₄).

RQ₄ Can substitutional *in situ* studies using VR provide a bridge over the methodological gap between lab and field studies?

Finally, this thesis contributes a novel remote VR research method, referred to as *Remote Virtual Reality for simulating Real-world Research* (RVR³). RVR³ combines traditional out-of-the-lab VR research [340, 400] with using VR for simulating real-world research [149, 291]. Empirical evaluations on real-world prototypes are often conducted in a physical lab, which this thesis advances through the introduction and validation of the use of VR studies. To contribute towards establishing the use of VR studies as a complementary research method within USEC research involving prototypes and to advance research in this field, this thesis concludes by showcasing how RVR³ enables researchers to move their research out of the lab and evaluate novel real-world prototypes in a remote setting (RQ₅).

RQ₅ Can the use of VR studies move traditional USEC research on real-world prototypes out of physical labs?

¹*In situ* refers to a (simulated) environment for which a system is intended to be used, similar to [528, 539].

1.3.4 Synthesised Thesis Research

The findings and the lessons learned from all studies conducted as part of this thesis contribute answers to RQ₆ and highlight the advantages and disadvantages of using VR studies for usability and security evaluations. In contrast to the previous research questions, i.e., RQ₁ to RQ₅, the last research question of this thesis is answered in chapter 7 and reads:

RQ₆ What are the advantages and disadvantages of using VR studies for USEC research involving prototypes compared to traditional USEC research in physical laboratories and in the field?

To answer RQ₆, the thesis synthesised the overarching findings of chapter 4 to chapter 6 in five *Research Recommendations* in Section 7.2.4 to support researchers in their work and facilitate and advance future research that is concerned with USEC research prototypes.

1.4 Thesis Walkthrough

This thesis has been organised in the following way:

Chapter 2, *Literature Review*, first provides an overview of empirical evaluations and how research involving user study participants is generally conducted. It reviews the landscape of practical research methods in three research areas relevant to this thesis: Human-Computer Interaction, Usable Security and Privacy, and Virtual Reality. The literature review discusses the various research methods applied and highlights the differences and similarities in conducting human-centred research between one and another area. Finally, this chapter reviews the broader authentication research domain, a major theme in USEC research [155] which “has always been one of the main challenges in usable security” [17]. The review briefly touches on different “*authentication types*” to set the context of this thesis and then provides an overview of the current literature on novel forward-looking authentication prototypes. It provides insights into the research methods that are applied when evaluating USEC prototypes and discusses some of the evaluation challenges in this field. The literature review summarises the lessons learned from existing works and unfolds the identified research gaps which motivated the research questions to which this thesis provides answers.

Chapter 3, *Scoping the Problem: Key Challenges in Usable Security Research*, synthesises the key challenges identified in the expert interviews, addressing **RQ₁**. This chapter presents and discusses, for the first time, the results of interviews with USEC experts who draw on significant expertise on the design, implementation, and evaluation of USEC prototypes. The following chapters, chapter 4, chapter 5, and chapter 6, build on the key challenges identified in this chapter and contribute towards potential solutions.

Chapter 4, *Validating the Use of VR Studies for Evaluating USEC Prototypes*, contributes the first validation of using VR studies for real-world usability and security research. This chapter is the first step towards providing solutions to the key challenges identified in chapter 3 by validating the use of VR studies for usability and security research on USEC prototypes, which forms the foundation for the follow-up evaluations in chapter 5 and chapter 6. Two experiments compare the usability (**RQ₂**) and security (**RQ₃**) evaluation findings of a prototype replicated and evaluated in VR with the original real-world study findings. Before contributing answers to **RQ₃**, this chapter defines the avatar level, i.e., abstract or high fidelity, that is required to simulate human gestural behaviour (for example, mid-air gestures) in VR. The findings from these validation studies supply the first evidence of the suitability of VR studies for simulating real-world USEC research, which is further supported through two additional user studies in chapter 5.

Chapter 5, *Using VR Studies to Advance USEC Research*, presents and investigates, for the first time, 1) a novel VR-based shoulder surfing research method to augment traditional shoulder surfing research in the lab, and 2) the use of VR studies to assess the impact of (simulated) in situ studies on a USEC prototype's usability evaluation outcome.

The first experiment in this chapter describes *3D Observations* and *VR Observations*, two novel shoulder surfing methods that provide user study participants with a more realistic shoulder surfing experience than traditional lab studies are capable of, contributing answers to **RQ₄** and **RQ₆**. The second experiment in this chapter describes a simulated in situ usability study that evaluates the impact of an environment on an authentication system's usability findings and users' authentication behaviour. This experiment reports on an evaluation and comparison of: lab studies in the real world, lab studies in VR, in situ studies in the real world, and in situ studies in VR, contributing answers to **RQ₄** and **RQ₆**. Both experiments in chapter 5 highlight the potential of using VR studies for substitutional real-world USEC research and unfold the associated advantages and disadvantages.

Chapter 6, *Remote Evaluation of Real-World USEC Prototypes Using VR Studies*, describes the combination of traditional remote VR research [340,400] with using VR as a proxy for real-world research [139,291,305] and introduces RVR³, a novel research method to remotely evaluate VR implementations of real-world prototypes. It presents the final experiment of this thesis and showcases the potential of VR studies for human-centred usability and security research through a remote usability and social acceptability evaluation of two novel real-world authentication prototypes (**RQ₅**). The study in this chapter is the first one that moves the use of VR studies for USEC research entirely out of the lab, highlighting the full potential of VR studies for USEC research involving VR replicas of real-world prototypes and scenarios.

Chapter 7, *Summary and Reflection on Thesis Research*, summarises the research conducted in this thesis and revisits the RQs outlined in Section 1.3. The research in the empirical

chapters resulted in a set of *Research Recommendations* to support researchers in their decision on when (and when not) to use VR studies for human-centred usable security research. The recommendations are synthesised in this chapter and contribute answers to **RQ₆**. Whilst the research recommendations of this thesis are more tailored towards authentication research, the inherently disciplinary nature of this thesis allows researchers from the broader HCI and USEC communities to learn from the recommendations outlined in this thesis. This chapter concludes with a discussion of the limitations of this work.

Chapter 8, *The End of a Journey: Conclusion, Future Research, and Final Remarks*, discusses open research topics and promising ideas for future research directions when applying VR studies for usability and security research. This chapter completes the thesis with a more personal remark by the author of this PhD research.

1.4.1 Overview of Studies

This thesis discusses seven user studies (N = 145), including one expert interview study (N = 12), three VR lab studies (N_{Study2} = 20, N_{Study5} = 18, N_{Study6} = 20), two remote non-immersive VR studies (N_{Study3} = 28, N_{Study4} = 22), and one remote immersive VR study (N_{Study7} = 25). Each study contributes to at least one research question outlined in Section 1.3. Table 1.1 provides an overview of the studies and how they relate to the research questions. Each study in Table 1.1 has an additional column describing the nature and the aim of the research.

Research Topic	User Study	Research Questions	Nature of Experiment and Purpose	
identify	Study 1	RQ ₁	Expert Interviews (Interviews, remote, N = 12) <ul style="list-style-type: none"> ◦ Elicitation of Challenges Experienced by USEC Experts ◦ Compilation of Key Challenges of Prototype-focused USEC Research ◦ “Ways Forward” to Address USEC’s Existing Key Challenges 	[Publication 3]
	Study 2	RQ ₂ , RQ ₆	Usability Study (VR study, in the lab, N = 20) <ul style="list-style-type: none"> ◦ Validation of the Use of VR Studies for Usability Evaluations on Real-World USEC Prototypes 	[Publication 4]
validate	Study 3	RQ ₃	Pre-Security Study (Survey with video prototypes, online, Pre-study of Study 4, N = 28) <ul style="list-style-type: none"> ◦ Exploration of the Impact of Avatar Fidelity on Identifying Interactions and Gestures 	[Publication 5]
	Study 4	RQ ₃ , RQ ₆	Security Study (Survey with video prototypes, online, N = 22) <ul style="list-style-type: none"> ◦ Validation of the Use of VR Studies for Security Evaluations on Real-World USEC Prototypes 	[Publication 4]
advance	Study 5	RQ ₄ , RQ ₆	Shoulder Surfing Study (VR study, in the lab, N = 18) <ul style="list-style-type: none"> ◦ A Sample Demonstration of Security Evaluations in VR ◦ Advancing Shoulder Surfing Research Using VR 	[Publication 6]
	Study 6	RQ ₄ , RQ ₆	Simulated In Situ Study (VR + real-world study, in the lab, N = 20) <ul style="list-style-type: none"> ◦ Investigation of In Situ Usable Security Research Using Real-World and VR Prototypes 	[Publication 7]
	Study 7	RQ ₅ , RQ ₆	Remote VR Study (VR Study, remote, N = 25) <ul style="list-style-type: none"> ◦ Usability and Social Acceptability Evaluation of Novel Usable Security Prototypes Using VR ◦ Transition Prototype-focused USEC Research Out of the Lab Using VR 	[Publication 8]

Table 1.1: Summary of the user studies presented in this thesis, along with the research questions they contribute answers to, the main purpose of each study, and the related publication.

Notes on Writing Style and Definitions:

Singular “They”: This thesis uses the gender-neutral pronoun “they”, i.e., *they* instead of *he* or *she*, when referring to researchers and user study participants. The singular *they* has been commonly employed in everyday English and has been formally accepted with the move towards gender-neutral language as it is inclusive of all people². It also helps writers avoid making assumptions about gender.

“Researcher”: This thesis uses the term “researcher” when addressing individuals within the broader research community. However, not everyone within the research community defines themselves as a “researcher”. Whenever this thesis uses the term “researcher” it covers any individual who contributes to research, be it a researcher who researches, a student who undertakes a research project, or a practitioner who practices a profession. Furthermore, the term “researcher” does not distinguish between different seniority levels – the research community requires and benefits from all levels of expertise.

“Prototype”: Researchers often have different expectations of what a prototype is [209]. For example, is a brick a prototype? Presenting one definition that covers all research fields and prototype variants is challenging. Whenever this thesis uses the term “prototype” (or “research artefact”), it refers to Wobbrock’s and Kientz’s definition of artefact contributions [556]: *“Artifacts, often prototypes, include new systems, architectures, tools, toolkits, techniques, sketches, mockups and envisionments that reveal new possibilities, enable new explorations, facilitate new insights, or compel us to consider new possible futures.”*

“Real-World Research”: Traditional HCI and USEC works use the term “real-world research” to refer to observations in the field, i.e., research in people’s natural environment³. However, due to the nature of this thesis, which simulates reality inside virtual reality, “real-world research” refers to either traditional research in a physical lab or to simulated field research. For example, the “real-world authentication environment” in Section 5.4.2.2 refers to research that simulates both a lab environment as well as a field environment (for example, *RW Lab* and *RW ATM* in Figure 5.7), whereas the “real-world ATM use” in, for example, De Luca et al.’s work [100], to which this thesis draws links, refers to field observations. This thesis uses the adjective “simulated” whenever research in the real world is simulated, both in a physical environment and in a VR-based environment (cf., Figure 5.7 on page 162).

“Language”: The research presented in this thesis has mainly been conducted in the United Kingdom. Therefore, all chapters of this thesis are written in British English. It also makes use of the stylistic Oxford/serial comma⁴, the final comma in a list of things.

²<https://apastyle.apa.org/style-grammar-guidelines/grammar/singular-they>, last accessed 22/01/2023

³<https://www.nngroup.com/articles/field-studies/>, last accessed 22/01/2023

⁴https://en.wikipedia.org/wiki/Serial_comma, last accessed 22/01/2023

II

LITERATURE REVIEW

Chapter 2

Literature Review

*Research is to see what everybody else has seen,
and to think what nobody else has thought.*

– Albert Szent-Gyorgyi –

2.1 Introduction of Literature Review

The opening chapter introduced the fundamental research problem that this thesis addresses: the inadequacy of empirical research methods when conducting human-centred usability and security research that involves hardware prototypes. The societal and technological transition, where mobility has become pervasive and prime to many people’s lives, has changed *how* and *where* people interact with technology. Smartphones, smartwatches, public displays, untethered VR/AR headsets, and many more technologies have contributed to a world where interacting with private data is no longer bound to people’s own four walls.

”In the mainframe era, authentication was limited to the workplace, the only location where users faced a need to authenticate with a computing system. Today, the fact that people can and do access sensitive information all the time and everywhere makes security and privacy protection an ever-present requirement.”
- Alt and von Zezschwitz [17]

In an era where Mark Weiser’s vision [538] of seamless integration of computing devices into our everyday life has already found application and where “*technology spreads from the workplace to our homes and everyday lives and culture*” [51], it must be of interest to the USEC community to revolutionise its research methods and address the existing (and emerging [17]) research challenges. Conventional laboratory studies cannot adequately simulate reality and cannot “provide for the wide range of competing activities and demands on users that might arise in a natural setting” [226]. Alt and von Zezschwitz [17] highlighted

the need to fundamentally rethink how the USEC community currently designs, implements, and evaluates research artefacts that facilitate people's lives but, at the same time, put their privacy and security at risk.

The literature review that this thesis will present does not put forward a detailed compilation of HCI's and USEC's history as other experts in the field have already covered this (for example, Carroll [69–71], Grudin [228], Myers [347], and Garfinkel and Lipford [155]). It also does not put forward a detailed review of traditional research methods as done in popular HCI books (for example, as done in "Research Methods In Human-Computer Interaction" by Lazar et al. [276]). However, the literature review here touches on HCI's and USEC's history and empirical research in general to allow the thesis to provide context for the following sections, which take a closer look at the research methods in HCI and USEC. Furthermore, the literature review will put forward an overview of VR's relevant subdomains applied when validating the use of VR studies for real-world USEC research. Reviewing the various research methods and interests within these areas helps the thesis position its contribution to the existing literature and sets the foundation of its interdisciplinary research approach.

The challenges in investigating USEC prototypes will be discussed in the context of "*Usable Security: History, Themes, and Challenges*" [155]. Garfinkel and Lipford [155] synthesised the historical context and the major themes that have emerged from USEC research. This thesis builds upon these findings and evaluates physical USEC prototypes, particularly in the authentication field. As Section 2.2 will show, the landscape of empirical research methods is extensive. Still, traditional HCI research methods often need to be adjusted to fit the requirements of USEC research, which will be appraised in Section 2.3.

The review in Section 2.4 will provide insights into the most relevant VR research components applied in this thesis. Although the VR research community has begun to adopt many research methods and questionnaires from the HCI communities (for example, the system usability score questionnaire [60] as used in [130] and the NASA-TLX questionnaire [193] as used in [567]), it has developed its own standardised questionnaires to deal with domain-specific research challenges, such as the Igroup Presence Questionnaire (IPQ) [451], the Embodiment Questionnaire [378], and the simulator sickness questionnaire (SSQ) [238]. These differences compared to the HCI and USEC communities and how VR found application in this thesis are discussed in Section 2.4.

Section 2.5 first considers the existing authentication research and then synthesises novel authentication prototypes that are accompanied by user evaluations. Reviewing the research landscape of human-centred authentication prototypes is essential as the validation of using VR studies for USEC research has been conducted on various authentication prototypes as part of this thesis (for example, on CueAuth [245] and ColorPIN [99]). The investigation focuses on knowledge-based authentication as it is often the primary authentication mechanism and

frequently used as a fallback method for biometrics. It has to be acknowledged that secrets, for example, PINs and passwords, will be part of people's lives for the near future [53, 203]. However, to acknowledge the size of the research space, this thesis chapter also explores existing biometric authentication prototypes to provide an overview of the broader user-centred authentication field. The remaining review discusses how authentication systems are currently evaluated and unfolds the challenges experienced by the USEC experts in this field. The chapter concludes with a summary and lessons learned based on the most stressing findings of this literature review, which have shaped the research presented in this thesis.

2.2 Human-Computer Interaction Research: History and Research Paradigms

The birth of the HCI community can be traced back to the first CHI conference in 1982¹. However, Carroll [69] argued that a book called “software psychology” constituted the historical foundation for current HCI research as early as 1970. Even earlier, in 1945, one of the first, albeit hypothetical, human-centred prototypes was described in “As We May Think” by Bush [64], where the idea of a microfilm-based electromechanical information-processing machine was presented. Bush's prototype enabled humans to quickly index and retrieve documents [64], which inspired computer science researchers, who have made major HCI contributions in the subsequent years. For example, Sutherland presented Sketchpad [500], a graphical communication system which opened up a new area of human-machine communication. Douglas Engelbart and colleagues presented the “on-line system” [128, 129], a computer collaboration system that employs hypertext links, the mouse, screen windowing, and other computing concepts. Their user study, which highlighted the advantages of mouse input over other input methods such as knee control or light pen, was one of the first steps towards human-centred system evaluations [129].

Until the late 1970s, the user base of interactive machines was mainly dominated by professionals. However, with one of the first affordable personal computers advertised in January 1975, i.e., the MITS Altair 8800 [2, 393], every human on this planet became a potential user. A few years later, in 1982, the CHI conference series started with the Human Factors in Computer Systems conference. Over the years, CHI has grown in popularity, and with it also the various research contributions and research methods [258, 556]. Although CHI articles alone do not depict HCI's entire research landscape and methods, they provide researchers with a good overview of the development of HCI's research field [286]. That being said, researchers often disagree on what “HCI research” actually encompasses and how correspond-

¹<https://sigchi.org/conferences/conference-history/chi/>, last accessed 22/01/2023

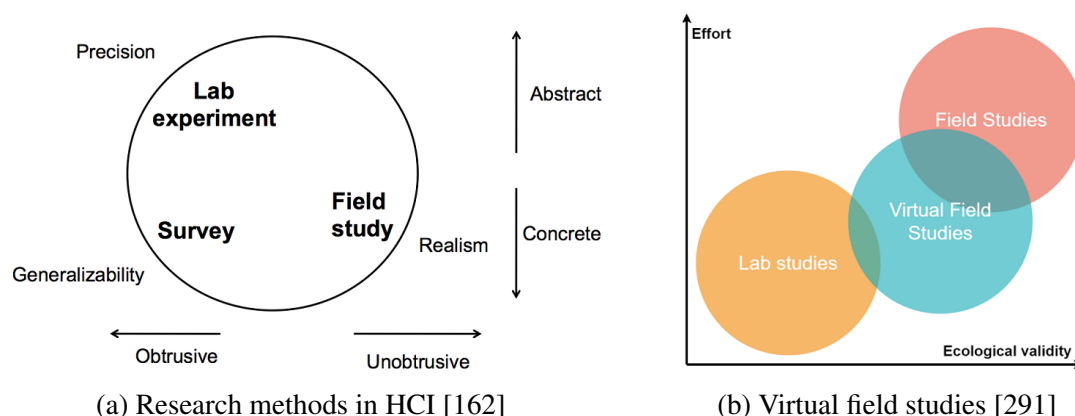


Figure 2.1: An overview of common research methods in HCI, including virtual field studies, as introduced by Mäkelä et al. [291]. Figures are original copies from [162] and [291].

ing evaluations should look like. A figure by Grudin [228, Fig. 1] shows four fields with major HCI research threads, highlighting its interdisciplinary nature and relatively broad definition:

“HCI is cross-disciplinary in its conduct and multidisciplinary in its roots, drawing on – synthesizing and adapting from – several other fields, including human factors (e.g., the roots for task analysis and designing for human error in HCI), ergonomics (e.g., the roots for design of devices, workstations, and work environments), cognitive psychology (e.g., the roots for user modeling), behavioral psychology and psychometrics (e.g., the roots of user performance metrics), systems engineering (e.g., the roots for much pre-design analysis), and computer science (e.g., the roots for graphical interfaces, software tools, and issues of software architecture).” – Hartson [412]

In an analysis of 1014 CHI papers [258], Koeman highlighted a large number of different research methods, the various user study lengths, and the differences between user study sample sizes within HCI. HCI has steadily encompassed more subfields such as Mobile HCI [252], USEC [437], Human-Computer Integration (HInt) [342], and many more. In 2023, the CHI conference counted overall eighteen subcommittees² to which researchers could submit their work (compared to eight subcommittees in 2013³), including *Critical Computing*, *Sustainability*, *Social Justice*, and *Privacy & Security*, among others. The vast amount and variety of HCI subdomains underline its interdisciplinary nature.

Empirical contributions, the act of providing new knowledge through findings based on observations and data collection, form one of the seven main research contributions in HCI and are defined as “the backbone of science” [556]. Traditional evaluation methods range from small-scale lab studies [6, 104, 145] to large-scale evaluations in the field [82, 200, 379], with

²<https://chi2023.acm.org/subcommittees/selecting-a-subcommittee/>, last accessed 22/01/2023

³<https://chi2013.acm.org/authors/call-for-participation/papers-notes/selecting-a-subcommittee/>, last accessed 22/01/2023

the most frequent ones being controlled lab experiments, field studies, surveys, and interviews [35, 66, 258, 462]. Although some of these approaches are based on expert evaluations, for example, cognitive walkthroughs [48, 387] or GOMS analysis for empirical predictions [225], others involve end-users for usability testing [111]. For instance, one of the most famous HCI experiments, the *Teledata* study by Nielsen and Molich [359], tested the user interface of a videotext system and asked students to report any usability problems they find. Evaluations of this type have become common in the design and evaluation process of HCI research artefacts. However, Poppe et al. [389] argued that many of the traditional evaluation methods are inappropriate these days due to emerging HCI applications that provide new sensing possibilities and diversify physical interfaces. Performance is no longer a key objective of HCI applications, but rather how HCI systems can be applied in everyday life [389].

The various HCI evaluation methods and the recent shift to novel research paradigms that make use of VR studies for simulating real-world research [291, 350, 450, 528] (cf., Figure 2.1) show that there is no “all-in-one” evaluation method. The lack of “universal panaceas” [173] is depicted in the famous “three-horn dilemma” that highlights the strengths and weaknesses of particular research methods and that factors such as the generalisability, rigour, and context of research cannot all be maximised simultaneously [112].

“No study, irrespective of the method used, can provide findings that are universally transferable. The study design should show a thorough consideration of what an adequate degree of transferability would be, in view of the assumptions of the research question, and present a relevant sampling strategy.” – Malterud et al. [295]

The literature review discusses in the following sections the most common research methods in HCI and puts those in contrast to each other.

2.2.1 Lab Studies in HCI

Lab studies have found widespread application in HCI and are the most popular research method in the broader HCI field. In essence, they are defined as studies that take place in controlled physical lab environments. Studies of this type are suitable for iterative evaluations of research artefacts [111] and can be employed in a variety of ways and at almost every research stage, for example, when comparing different feedback modalities [146] or in advance of deploying a prototype in the wild. Lab studies are often described as “cost-effective” as they require comparatively little planning, are suitable for early use in the research process, and allow for high control over extraneous variables. Typical lab studies isolate specific variables that will enable researchers to investigate the impact of individual variables on a research artefact’s usability, for example, the effect of different input methods on usability [380, 464].

Several researchers commented on the value, pros, and cons of laboratory studies in HCI as part of an expert panel [557]: According to Landauer, the most familiar application of laboratory studies is in “iterative developmental, or formative, evaluations” [557] where Wizard of Oz [87] prototypes or fully functional prototypes are tested on user study samples. Such evaluations in the lab allow researchers to pinpoint obvious design flaws and incorporate the observations made and comments raised by user study participants into an iterative design process. Landauer highlighted the high degree of convenience and control of studies in the lab and that a “blanket assertion that lab findings are of little value is not justified” [557]. However, whilst lab studies are inevitable for human-centred research, and the most leveraged study type in HCI [257, 258], researchers have recognised and acknowledged that studies of this type often do not represent real-life situations [111, 191, 357, 541].

“Laboratory studies are often somewhat unrealistic because they do not happen in the right context or situation intended for the application, like at home, at a specific public place or during a sight seeing tour.” – Rukzio [426]

Despite the value of lab studies, it has been argued that, whenever possible, prototypes should be evaluated with their actual intended users and within their application contexts. For example, John emphasised that lab studies have a role in HCI research, but their role is not to evaluate proposed systems and “feeding directly into design” [557]. Whiteside [557] found that their final product did not meet their expectations because the research in the lab overlooked crucial contextual factors in which the system is eventually used. They concluded that laboratory experiments provide little design guidance because they are ecologically invalid as the laboratory context is not people’s natural context in which they interact with HCI systems [541, 557].

2.2.2 Field Studies in HCI

Researchers are well aware that the evaluation context⁴ is key in HCI [421, 446]. In contrast to lab studies, field studies aim at evaluating prototypes or receiving insights into human behaviour in their intended use case in the wild (for example, when learning about real-world responses to interactive gesture-based public displays [192]). There are ongoing debates about whether or not conducting field studies is worth the inconvenience of increasing the external validity of user studies [253, 254]. One prominent example is the comparison of a lab evaluation and an equivalent field study by Kjeldskov et al. [254]. In 2003, Kjeldskov et al. [252] highlighted the dominance of laboratory evaluations in mobile human-computer interaction research and that the shortcomings of how research is conducted inhibit the development of the Mobile HCI research field as a whole. One year later, Kjeldskov et

⁴Definition from Cambridge Dictionary: “the scenario in which something exists or happens, and that can help explain it” <https://dictionary.cambridge.org/dictionary/english/context>, last accessed 22/01/2023

al. [254] compared the results of a lab-based usability evaluation on a context-aware mobile system with an evaluation in the field to investigate the studies' abilities to identify the usability flaws of a context-aware mobile prototype. To the surprise of many HCI researchers, Kjeldskov et al.'s results show that the benefits of field studies over laboratory investigations are neglectable and that the advantages of field studies over lab studies are unclear and likely to be "not worth the hassle" [254]. Ten years after the research by Kjeldskov et al., they shifted their ground, and although they still have not found a definite answer to the lab vs field question, the real question should be *when* and *how* field research is worth the hassle [253].

The discussions on lab and field studies have been a hot, controversial topic across different HCI subdomains for years. Nielsen et al. argued that "it is definitely worth the hassle to conduct usability evaluations in the field" [357], and that different contexts reveal problems with interaction style and cognitive load that are impossible to identify in the lab only. Dourish emphasised that "*the situation in which technology is used has become more variable, and so we need to understand more about it. [...] if we take 'ubiquitous computing' seriously, then we should be applying its ideas ubiquitously, not just in the relatively narrow areas of interaction with handheld and embedded devices.*" [116].

Although field studies are often considered to be the holy grail of human-centred research when aiming for evaluations in natural environments [111], they are often expensive and hard to control [254, 291]. Many potential threats to field studies have "conceivable relevance" in HCI [83, 375], including threats to internal validity, construct validity, and external validity, as results obtained in one field study might not generalise to other settings.

2.2.3 Online Studies and Online Surveys in HCI

Online studies can broadly be defined as user research conducted in remote settings. Traditionally, remote software tools such as websites were consulted for research purposes. These studies often range from more traditional online surveys that consist of people filling in questionnaires to more "*lab-style*" studies where users are asked for hands-on activities. This includes online surveys that show video prototypes [416], ask the participants to answer questions about the quality of listening experiences [67], or require the participants to download and install study applications [201, 218, 369, 566]. Whilst the traditional method of surveys is to use paper-based surveys [276], with the advent of remote research opportunities there is a notion of an increase in online platforms such as Stack Overflow⁵, Reddit⁶, XRDRN⁷,

⁵<https://stackoverflow.com/>, last accessed 22/01/2023, is a community-based space to discuss and solve technical challenges

⁶<https://www.reddit.com/>, last accessed 22/01/2023, is a platform where people can dive into their interests, hobbies, and passions.

⁷<https://www.xrdrn.org/>, last accessed 22/01/2023, is a platform where XR researchers can connect with user study participants.

Amazon Mechanical Turk [20], and Prolific [394] for online research and participant recruitment. As discussed by Lazar et al. [276] in their HCI book, there is often the question of whether responses from electronic surveys are comparable, trustworthy, and valid as in-person paper-based surveys. Although there is no evidence that people are more dishonest in online surveys compared to paper surveys, user study participants in online surveys tend to be more honest when delivering bad news [499]. Furthermore, anonymously submitted surveys can lead to an increased level of self-disclosure [318,483]. Online studies on platforms like Reddit have been shown to be inexpensive, offer data quality comparable to the responses obtained from student samples [222], and allow for rapid data collection from a large sample, whilst targeting specific populations [458].

The capabilities of websites for data collection are limited, which motivated researchers to deploy custom-built or modified software that is more tailored to their research needs [276, p. 308]. McGrenere et al. [317] built an extension for Microsoft Word 2000 to allow users to work with a simplified user interface. As part of a field study, they tested a novel interface design for heavily-featured productivity and installed it on the participants' devices along with a logging tool that allowed them to capture usage data when interacting with and navigating through the menus and toolbars. Wagener et al. [534] modified the VR painting app Open Brush⁸ to customise its functionality and deploy a self-care VR app that allowed inducing pleasant emotions and foster self-reflection and well-being. Dang et al. [89] implemented a web system to investigate how multiple sliders, with and without feedforward visualisations, influence the users' control of generative models (i.e., styleGAN2 [234]). Seitz et al. [454] deployed an online browser-game that supports users in quantifying the perception of password strength. Although their game was mainly advertised to peers and students during student orientation days, everyone who found the web application online was a potential participant in a research project [454]. Similarly, Henze et al. [202] deployed a typing game in the Android market to investigate the users' typing behaviour on the standard Android keyboard. Results based on 6,603,659 keystrokes and 13,013 installations showed that visualising the users' touched points through a dot decreased both their input speed and their error rate [202].

In conclusion, Lazar et al. [276] emphasised that HCI researchers have a "formidable resource" that can (and should) be exploited to overcome the obstacles of laborious, time-consuming, and - as a result - often error-prone HCI research. In contrast to traditional lab studies, a significant advantage of online research (be it hands-on or more survey-like) is its capability to recruit a large number of diverse participants. The widely available and custom-built software tools allow researchers to collect vast amounts of usage data, often from participants in highly realistic real-life scenarios, which is close-to-impossible to achieve in the lab.

⁸Open Brush is a fork of Tilt Brush by Google, which was made open source in 2021: <https://opensource.googleblog.com/2021/01/the-future-of-tilt-brush.html>, last accessed 22/01/2023

2.2.4 Interview Studies and Focus Groups in HCI

Interview studies help researchers to better understand and explore the users' experiences and opinions on a specific topic. One of the strengths of interviews in comparison to other research methodologies is the ability to "go deep" [276]. However, on the downside of interviews is the challenge of recruiting interviewees as they are often time-consuming and require personnel resources, which might not always be easy to access. Interviews can either be structured, where researchers strictly follow a protocol, or semi-structured, where researchers are roughly guided by key questions and add additional questions depending on the interviewees' responses. They can target end-users or experts within a field, which are particularly valuable for receiving in-depth insights into a research field and learning about the associated research challenges [264, 484]. Interviews are often conducted face-to-face in a research lab, on online video chatting platforms, like Zoom [67], or via phone calls [407]. They are often applied to complement other research methods in HCI, for example, additionally to usability testing in the lab [314, 380], or to follow up on large-scale online surveys to support, strengthen, or refute the prior research findings [494, 513].

Interview sessions with more than one interviewee, usually referred to as focus groups, are promising for collecting user needs and designing product requirements. Focus groups are often the best way to exchange viewpoints, engage in brainstorming sessions, and get insights into people's opinions and preferences in a shared, social setting. Compared to interviews, they are reasonably effective and inexpensive for gathering a broad range of opinions easily [8, 276]. There is a broad range of focus group samples and, similar to the sample size discussions in HCI [66], there is no clear answer to the "how many people per focus group?" question. Whilst some researchers suggest recruiting between eight and 12 people [419, p. 300], others prefer slightly smaller focus groups for more in-depth conversations [179, 271]. Guest et al. [179] brought up the interpersonal dynamics that potentially impact saturation: "*A few highly disruptive or vocal participants, for example, can reduce the variability of responses within a focus group*" [179], highlighting the need of good moderating skills to contribute to high reliability and validity of the findings from focus groups. Examples of focus groups span a large variety of topics. For example, Melenhorst et al. [325] conducted focus groups to investigate the context-related benefits of mobile phones for elderly people. Others applied focus groups to collect early feedback on interactive VR experiences [153] or to explore the users' perceptions and misconceptions of internet cookies [185].

Although focus groups have found application within HCI, they are often criticised for their lack of reliable and detailed data to properly ground a product design in its projected use case [423]. Focus groups are too frequently used as a usability evaluation method and users cannot accurately assess research artefacts until they interact with those [423].

2.2.5 Virtual Reality Studies in HCI

In 1997, Pierce and Aguinis raised “*awareness of the possibility of using immersive VR technology as a tool for conducting empirical research [...]*” [383]. A few years later, in the early 2000s, Blascovich [50] proposed using VR studies as a new social psychological research tool to overcome the existing problems around control–mundane realism trade-off, lack of replication, and atypical sampling. There is no single work known as the pioneer of using VR studies for empirical research; however, with these comments in mind, the HCI community has shown interest in using VR studies to simulate real-world research and cope with different study requirements and challenges. Mäkelä et al. [291] see VR studies as a good trade-off between lab and field studies as they contribute towards increased ecological validity compared to lab studies and are more affordable than traditional field research (cf., Figure 2.1). Nobel Prize winner Edvard I. Moser and colleagues highlighted the VR’s potential to facilitate studies of animal behaviour [332]. Rebelo et al. [404] argued that VR studies enable researchers to develop realistic-looking environments that exhibit greater control of the experimental conditions than lab settings, benefitting the user experience evaluations as VR allows simulating the interaction between a system, the user, and its context of use.

So far, many of the existing studies in the broader HCI field have focused on how well findings from VR studies transfer to reality. Early works provide the first evidence of the suitability of VR studies for simulating real-world research in several subdomains of HCI.

Mäkelä et al. [291] explored user behaviour in front of public displays in VR to then compare it to real-life audience behaviour. Their work highlighted the many similarities between results from VR studies and original real-world studies, such as successfully recreating the honeypot effect [59] in VR, which describes how people interacting with public displays “passively stimulate passers-by to observe, approach, and engage in an interaction” [560]. Voit et al. [528] explored the differences and similarities when using five different research methods (i.e., lab, online, VR, AR, in situ) to evaluate smart artefacts. They found that the selected research method can negatively impact the study outcome with regards to usability ratings. However, VR and in situ provided similar ratings for usability, attractiveness, and pragmatic and hedonic qualities of their tested research artefact [528]. Bruno and Muzzupappa [62] compared a product interface design evaluation in the real world and VR. Their research found that VR is a valid alternative to traditional research methods for product interface usability evaluations and that VR does not invalidate the usability evaluation findings [62]. Götz et al. [171] investigated four different methods of autobiographical recall in VR (i.e., talking to a virtual computer-controlled avatar, talking to an allegedly connected avatar to the researcher, writing/drawing with a VR pen, and thinking quietly). They found that all four autobiographical recall methods induced changes in emotional states, suggesting that their studied autobiographical recall methods can be used in VR [171].

VR-based simulations have also found application in research on user-centred automotive interfaces. Colley et al. [81] explored vehicle motion effects on interaction quality in VR using a motorised swivel seat, finding that a VR and motion effect simulation increased the participants' perceived realism of vehicle motion and their feeling of presence. However, they conclude that VR-based simulation studies should be primarily used for early explorations as there are indeed differences between simulated driving and real-world driving [417]. Similarly, Goedicke et al. [163] investigated, for the first time, an on-road VR driving simulator to support researchers in conducting "safe testing of human response and effective prediction of human performance" [163]. Although they concluded that additional validation studies are required, their initial pilot test demonstrated that an on-road VR driving simulator can indeed invoke genuine responses from participants [163].

Other works by Savino et al. [439] compared navigation methods in VR to the real world, finding that whilst the participants' perceived task load and navigation performance differed significantly between VR and reality, their route recognition (i.e., number of correct turns and directional changes) was similar across the environments. They concluded that VR studies offer a promising simulation environment to test navigation methods, but the current VR technology is not mature enough to present highly realistic real-world experiences [439]. Weiß et al. [539] evaluated do-it-yourself tasks using an online survey, as well as VR, AR, laboratory, and in situ studies. They found that the use of surrogate empirical methods such as AR/VR prototypes is valuable to infer insights about in situ studies. However, they concluded that simply transferring results from HCI subdomains to other research domains might not be possible and that researchers must verify prior results before applying substitutional research methods in different research fields [539].

2.2.6 Summary

The beginning of this section touched on the HCI's history and empirical research in general. It then provided an overview of the more specific HCI research landscape, including a review of the most common HCI research methods such as lab, online, and field studies. Applying a research method that best fits the nature of a researcher's aims, objectives, and research questions is important to fulfil the overarching goal of research: "produce information that can be shared and applied beyond the study setting" [295] and gradually add to the accumulation of human knowledge.

The review in this chapter has shown that novel research methods, for example, the use of VR studies for simulating real-world research, contribute to and facilitate empirical research. Although the use of VR studies for the simulation of real-world research has already received some attention and resulted in promising research in the broader HCI communities [291, 439, 528, 539], there have been no attempts in studying VR studies and their feasibility for usable

security research. This thesis will investigate *if* and *how* the use of VR studies can be applied within the USEC research field to support and facilitate research in this space.

The next section will discuss the links between HCI and USEC research and outlines how USEC research is generally conducted by reviewing its most common research methods.

2.3 Usable Security Research: HCI's Sibling?

So far, this thesis has reviewed common empirical research methods in HCI, many of which can be found in the broader USEC field. The close connection between the HCI and USEC communities is not a surprise, with the first formal gathering of the USEC community at ACM CHI 2003, the premier international conference of HCI research [21]. However, compared to HCI, USEC researchers are concerned with how people make decisions about their privacy and security, how they interact with security and privacy systems, and how ubiquitous technology can, and should be, designed for usable, secure, and privacy-preserving experiences. As a result, the USEC's methodological challenges are different from those in neighbouring fields. To name two examples: user study participants can comment on an HCI prototype's usability, but they are often unable to judge accurately the security of the prototype and describe security problems they have experienced [155, 399, 495]. Furthermore, in contrast to HCI, USEC requires the coordination of both the usability and the security of a research artefact, which is the key to designing, implementing, and deploying usable and secure systems [437, 542].

"[...] the process of designing and conducting security-related user studies remains extremely difficult. Users deal with security infrequently and irregularly, and most do not notice or care about security until it is missing or broken. Security is rarely a primary goal or task of users, making many traditional HCI evaluation techniques difficult or even impossible to use." – Egelman et al. [121]

Despite the different research interests within the communities, the USEC community has adopted and learned from many HCI research methods: lab studies, field studies, remote studies, and interview studies have found widespread application and resulted in forward-looking USEC prototypes and insights into people's security and privacy behaviour. Zurko and Simon [568] highlighted the need for user-centred design techniques and argued that most security systems' usability problems could be addressed by task analyses, interviews, usability testings, and iterative designs. In 1999, Whitten and Tygar [542], and Adams and Sasse [9], demonstrated that security systems will not be secure if the usability of these systems is neglected. Since then, user-centred security design techniques have begun to find application.

This thesis uses a range of USEC works and methodologies to provide an overview of the USEC's various empirical research methods and its landscape.

2.3.1 Lab Studies in USEC

Researchers often draw on evaluations in a laboratory setting regarding USEC research that involves prototypes or requires a controlled environment. In many of these lab studies, the experimenters assess forward-looking prototypes to receive insights into a prototype's usability and security and the users' interaction behaviour. However, observing realistic security behaviour in the lab is often challenging, requiring USEC researchers to adjust their study designs [133, 267, 482] or make use of alternative research methods (for example, Wizard of Oz studies [172], where approximations of fully functional prototypes are deployed). Mecke et al. [323] conducted a "Wizard of Oz study" to investigate the users' perception of physical, biometric, and behavioural authentication concepts to open doors. When role-playing an "opening door scenario" in the lab, they learned that, whilst the participants appreciated biometric authentication on doors, they valued the control they gained from the possession of physical tokens. Krol et al. [267] emphasised the use of role-playing to overcome some of the limitations of lab studies. In a study by Schechter et al. [443], the participants went through several online banking tasks using their actual credentials or simulated credentials. They found that the participants who use their own authentication credentials behaved more securely than those who received simulated ones [443].

The importance of simulating real-life contexts and providing participants with realistic research artefacts, rather than with approximations of professionally designed products, is further evidenced by De Luca et al. [98, 102]. Using two connected mobile phones to provide users with a back and front display for user authentication enabled the testing of the idea, but the prototype's weight negatively impacted its usability [102]. Furthermore, whilst the previously discussed work by Mecke et al. [323] simulated a real-life context in the lab and enabled learning more about the participants' preferences of door unlocking mechanisms, some participants stated that they were unsure about the functionality of the systems.

In summary, lab studies in USEC research exhibit similar pros and cons as lab studies in HCI, and have in common that they are often limited to local and homogeneous study samples and require both the experimenter and the user study participants to be physically present, which can be problematic when pandemics, such as lockdowns during COVID-19, prohibit face-to-face research [210]. Furthermore, the USEC prototypes might not accurately represent how the systems would function in a real-life scenario (for example, [102]). Many USEC prototypes allow for the user testing in controlled lab environments, but they are not robust enough for rigorous data collection in the field. In addition to the deployment issues of novel USEC prototypes, one example of the lack of robustness is the USEC prototype by Schaub et al. [441], a context-adaptive ambient calendar display that shows a person's schedule. Whilst Schaub et al.'s prototype allowed the initial testing of the idea, there were concerns about the robustness of the prototype and the presence detection and identification, likely resulting

in more conservative privacy settings than preferred by the users [441]. The lack of robust USEC prototypes and the need of controlled environments often inhibits the researchers to move their USEC research on prototypes out of the lab. Some of the challenges and examples of USEC research in the field are reviewed in the next section.

2.3.2 Field Studies in USEC

Field studies are rare in USEC compared to lab or online surveys (which will be discussed in Section 2.3.3). This is not to say that there have not been any attempts to conduct USEC research in the wild; however, USEC researchers face significant challenges when evaluating their prototypes or investigating the users' security and privacy behaviour in realistic field studies. On the paper, field studies are popular among USEC researchers due to their high ecological validity, which supports researchers in generalising research findings and predicting user behaviour beyond individual user studies. However, there are significant ethical and legal constraints when conducting research in the wild [100, 151, 311], which are often beyond the capabilities of individual researchers and negatively affect and reduce the amount of USEC research that takes place in the wild.

Despite the challenges in conducting USEC field studies, there are some remarkable examples: Felt et al. [136] evaluated different SSL warnings, i.e., warnings that indicate that a user's integrity and confidentiality on a websites is at risk, in Google Chrome and recorded overall 130,754 user reactions. Compared to a traditional lab study, Felt et al.'s research has a superior methodology because the behaviour of their participants is not simulated in the lab and represents the actual task that users naturally choose to do during their daily online activities [136]. De Luca et al. [100] and the replication study by Volkamer et al. [529] contribute cross-country insights into people's ATM interaction behaviour through observations in the field. Colnago et al. [82] explored the staff and the student's behaviour and opinions on deploying two-factor authentication in an academic environment. Paul et al. [377] conducted a 10-week field study to explore the users' perceptions of a smart card authentication system. Their field study made use of a variety of ethnographic research methods, including periodic surveys, diaries, and observations [377], finding that the participants positively experienced smartcards for user authentication and that their perceptions were influenced by personal benefits rather than the increased security. Chiasson et al. [75] conducted a large-scale field study to investigate click-based graphical passwords and how they work in practice, providing empirical evidence that relying on results from lab studies alone might be misleading. Harbach et al. [191] coordinated a one-month field study to gain insights into smartphone users' real-world (un)locking behaviour. The results confirmed the previous findings from an online survey and contributed to the generation of a "ground truth" for improving smartphone un(locking) mechanisms [191].

The reviewed literature shows that several researchers successfully conducted USEC research in the field and made significant scientific research contributions. Despite the high ecological validity of superior research methods (for example, [136]) and their advantages over more traditional lab research, the applications of field studies in USEC remain rare. The resources such as the ones used by Felt et al. [136] and Colnago et al. [82] are often not within reach of individual investigators and academic research labs, resulting in research that takes place in the lab and is not moved forward to the field. To overcome these limitations, researchers have started using online resources for USEC research, which will be reviewed in the next section.

2.3.3 Online Studies and Online Surveys in USEC

Whilst many hardware prototypes are evaluated in the lab due to logistical challenges, the combination of building software prototypes and conducting online studies has found widespread application in USEC. In this research, investigators deploy their experiments online and participants solve various tasks in their preferred location and time using their own devices.

Von Zezschwitz et al. [531] conducted a longitudinal study to gather insights into the users' performance on Android-like patterns and PINs for user authentication. Prange et al. [391] deployed an Android app to investigate how smartphone users perceive switching from their primary authentication method to a fallback one. In Gutfleisch et al.'s work [183], participants were given step-by-step instructions on establishing a remote connection to the experiment computer. This allowed the authors to remotely investigate how users interact with the security warning messages and if users are aware of the implications of their interaction [183]. Others relied on web browsers and their opt-in metrics to empirically assess the effectiveness of browser security warnings [13, 136] or conducted large-scale authentication evaluations using web applications [141]. Huaman et al. [210] proposed using a commodity browser to conduct lab-like USEC research and overcome the challenges around participant recruitment and complicated circumstances. Nguyen et al. [352] showcased through a remote user study that IDE security plugins can be effective security measures to help developers contribute secure code. Acar et al. [7] conducted a controlled experiment using Jupyter Notebook⁹ and Amazon Web Service (AWS) instances on which developers attempted common tasks involving symmetric and asymmetric cryptography. Krawiecka et al. [265] introduced an end-to-end framework for remote experimentation in cyber security and argued that remote experiments allow for "better representation of human participants and more realistic experimental environments and ensure research continuity in exceptional circumstances, such as nationwide lockdowns." [265].

Additionally to the reviewed online studies that used smartphones, browser-based environments, or end-to-end frameworks, the use of traditional online surveys has found widespread

⁹The Jupyter Notebook is a web application for creating and sharing computational documents: <https://jupyter.org/>

adoption in USEC. In fact, online surveys are one of the most popular research methods in USEC [110]. Redmiles et al. [406] even consider gathering self-reported data from users through online surveys as key in USEC for understanding and improving how human decision-making affects security. Markert et al. [298] conducted a large-scale online survey to explore users' perception of security, memorability, and ease of use of PIN-based authentication. Ion et al. [217] compared self-reported security practices of non-experts to those of security experts, and Tahaei and Vaniea [504] compared the programming skills, privacy and security attitudes, and secure development self-efficacy of participants across a computer science student mailing list and four crowd-sourcing platforms.

Despite the widespread application of online surveys in human-centred usability and security research, collecting self-reported user data is not always straightforward and comes with notable drawbacks. Redmiles et al. [406] highlighted the importance of word choice, the context of questions, the order of answer choices, the length of surveys, and the challenges around sensitive questions: survey takers tend to under-report socially-undesirable behaviours, whilst over-report desirable behaviours, such as exercising, using the library, or voting.

2.3.4 Interviews and Focus Groups in USEC

Similarly to HCI research, the use of interviews and focus groups has found application in USEC to learn about the users' security and privacy behaviours and preferences. Tahaei et al. [503] conducted expert interviews with software development teams to learn about their motivations, challenges, and strategies for protecting end-user privacy. Krombholz et al. [270] conducted expert interviews to understand the deployment process of HTTPS and its challenges. Interviews are often used in combination with online surveys to, for example, uncover how much of a role security and privacy played in people's decisions to use a mobile instant messenger [288]. Furthermore, they are often combined with observations to learn more about user behaviour in security scenarios [100]. Prange et al. [392] conducted semi-structured interviews to receive insights into the participants' experience with PriView, a concept and prototype that allows privacy-invasive devices in the users' vicinity to be visualised. Mecke et al. [323] conducted semi-structured interviews to compare three authentication prototypes for doors and collect qualitative feedback from their participants. In line with HCI, interviews and focus groups in USEC can complement each other or even build upon the initial findings from online questionnaires. For example, Prange et al. [390] combined semi-structured interviews with a focus group to learn more about the potential future of usable authentication in smart home environments. Nicholson et al. [356] conducted focus groups to learn about young people's knowledge of cybersecurity and support previous questionnaire findings.

2.3.5 Summary

As Section 2.3 has shown, a broad range of HCI research methods has found application in USEC. These methods have pushed forward USEC research, enabled USEC researchers to learn about the users' security and privacy behaviour, and contributed novel forward-looking prototypes. USEC research often adheres to traditional laboratory studies and online surveys, with some exceptions as reviewed in Section 2.3.2. Little attention has been paid to prototype-focused USEC research and how research in this space can be facilitated and moved out of the lab, potentially contributing to the generalisability of research findings.

There are early attempts to move traditional lab-based USEC research out of the lab using commodity browsers [210]. However, browser-based research methods do not allow researchers to evaluate the usability and security of hardware prototypes. At a time where new technologies are constantly becoming part of people's lives and affect how and where technologies are being used, it is important to understand how USEC research can be facilitated and advanced to cope with the emerging challenges and to mitigate potential threats before they become reality [17]. Existing research methods such as online surveys and the variety of web-based research methods (cf., Section 2.2.3) are not adequate for contributing usability and security evaluations on USEC prototypes as they do not allow participants to interact with and experience the prototypes in real-time.

In comparison to HCI where novel research methods, for example, VR studies [291, 528], have found application and contributed to noteworthy contributions to the field, there is a lack of equivalent validation research, as done by, for example, Mäkelä et al. [291] and Voit et al. [528], that would allow the USEC community to seamlessly adopt VR studies for usable security research. To contribute insights into the use of VR studies for USEC research, this thesis applies, for the first time, VR as a research method for the usability and security evaluation of virtual replicas of real-world artefacts that replicate look, feel, and functionality in VR simulations of real-world contexts where prototypes are intended to be used. To validate the results from USEC research on virtual replicas in VR, and to provide important empirical evidence of the validity of VR studies for USEC research, comparisons are drawn to studies in physical lab environments. This helps the thesis in pinpointing the similarities and potential differences between VR studies that simulate reality and more traditional studies in physical laboratory environments.

Due to the VR research approach this thesis introduces and validates for USEC research, it is necessary for this thesis to provide a summary of VR research and how empirical research is conducted within the VR research field. Additionally, the most stressing prerequisites and VR subdomains that were consulted for the validation of the use of VR studies for empirical USEC research are reviewed in the next section.

2.4 Virtual Reality: An Overview

The history of VR goes back to the 1960s with the VR research field being grounded in the computer graphics field [313,501] and initial contributions by Morton Heilig, *Sensorama* [199], and Ivan Sutherland, *The Sword of Damocle* [501]. As highlighted by Steuer [491], several definitions of VR are available and accepted within the community. A suitable definition for the scope of this thesis is: “*Virtual Reality is electronic simulations of environments experienced via head-mounted eye goggles and wired clothing enabling the end user to interact in realistic three-dimensional situations.*” [491]. The “three-dimensional situations” are computer-simulated, multi-sensory environments [383], where a human loses awareness of the HCI interface and their physical co-presence and, instead, feels present or fully immersed in a virtual environment [120].

Since the 1960s, there have been investigations on the affordances, applications, and potential socio-technological implications of VR [123, 181, 227, 290]. The VR research field has grown significantly in the past few years, with one prominent and aspiring application being simulating reality and real-world research artefacts in VR (for example, for HCI research [291,528], neuroscience research [52,507], and fieldwork in design education [149]). The interest in simulating real-world contexts and research artefacts in VR is motivated by ever-increasing display fidelity, i.e., “the objective degree of exactness with which real-world sensory stimuli are reproduced” [319], interaction fidelity, i.e., “the objective degree of exactness with which real-world interactions are reproduced in an interactive system” [401], and near-perfect tracking, leading towards perceptual realism and the recreation of interactions, events, and places that increasingly represent accurate simulations of reality [475].

Reviewing the VR field as a whole is beyond the scope of this thesis as its aim is not to provide an exhaustive overview of the VR literature, the scientific and technological challenges [120], and how research is being conducted in the broader VR domain. However, as this thesis applies VR as an empirical method for USEC research, the next sections provide an overview of empirical VR research to highlight the similarities and differences between VR, HCI, and USEC. Furthermore, the two VR subdomains and prerequisites that have been consulted when applying and validating the use of VR studies for simulating real-world USEC research will be discussed.

2.4.1 In Contrast to HCI and USEC: Empirical VR Research

Many of the common user-centred research methods in HCI and USEC have been adopted by VR, including traditional lab studies [182, 188, 314, 366], remote/online studies [339, 340, 400, 430, 466], online surveys [182, 366], and in-the-wild investigations [230, 486, 521]. However, due to the hardware-heavy VR research prototypes (for example, [36, 76, 261, 461, 535]), VR

research often takes place in controlled laboratory environments. For instance, testing a wrist-worn haptic device that renders virtual objects into the user's hand on demand [261] requires inviting user study subjects to the lab to experience and interact with the prototype. Yet, with the widespread adoption of affordable VR headsets such as the Meta Quest 2 [328], which has been sold millions of times and is powerful enough for research purposes, researchers have found new powerful research methods to move VR research outwith the traditional laboratory context. The two most prominent directions of VR research methods beyond traditional laboratory settings are studies in the wild [339, 379, 521] and online studies [340, 431, 546]: Mottelson and Hornbæk [339] explored pointing, 3D tracing, and body illusions and compared their findings from a laboratory setting with an in-the-wild investigation. Williamson et al. [546] conducted a remote VR workshop to learn more about the size of virtual space and how it influences group formations, shared attention, and personal space. Saffo et al. [431] conducted a remote study on a social VR platform to investigate the users' behaviours when collaborating on a tabletop. Steed et al. [486] conducted a remote VR study on presence and embodiment with the participants already possessing VR headsets. The availability and affordances of VR headsets motivated Rivu et al. [400] to conduct remote VR research on participant-owned VR headsets, contribute a framework for remote VR studies, and derive best practices when moving traditional VR research out of physical laboratory settings.

The advancements of VR research and its research methods have opened many more opportunities for the research community to benefit. Steed et al. [488] emphasised that, although the COVID-19 pandemic has had a significant impact on the VR research field, it enabled the fields of HCI and VR/AR to continue to forge forward with experimental work and investigate distributed user studies which *“will be useful for other areas of HCI and, indeed, any field that relies on human experimentation”* [488].

However, if the use of VR user studies finds widespread adoption for human experimentation beyond traditional VR research, it is essential to understand its feasibility and the differences between the various research fields before broadly applying VR studies in neighbouring research domains. For example, in contrast to HCI and USEC, where empirical evaluations often involve usability testing and contributing answers to common usability questions, the core research interests of VR researchers are quite different. Novel VR input techniques still raise typical HCI questions, like how usable the system is or how quickly users can provide input [346, 465]. Yet, VR research is additionally concerned with avatars [168], 3D and volumetric display and projection technologies [211, 220], locomotion [109], and the sense of presence [208, 453], among others. Two of these fields, namely research on the VR users' sense of presence and VR avatars, were applied in this empirical thesis research. First, to inform the design of a user representation in VR for security evaluations in Section 4.7 and Section 4.8 in chapter 4. Secondly, to comment on the participants' sense of presence and perceived realism when interacting in simulated VR environments with virtual artefacts of

real-world prototypes in chapter 5 and chapter 6.

The following section provides an overview of the research on the **concept of presence**, including place illusion, plausibility illusion [475], interactional realism [549], and **VR avatars**. Both VR domains have been consulted and utilised when validating the use of VR studies for usability and security research. Other common VR research domains, for example, research on locomotion and navigation [468] or body ownership [170], are beyond the scope of this thesis; however, they surely deserve attention when aiming to establish the use of VR studies as a research method across various human-centred research disciplines.

2.4.2 Prerequisites for Applied Empirical VR Research in USEC

Although the contributions of this thesis are situated in the HCI and USEC fields, it is important to review the VR concepts that have been consulted and applied when validating the use of VR studies for simulating real-world USEC research. This thesis will show in chapter 5 and chapter 6 the importance of the user study participants' sense of presence in virtual environments when immersing them into simulations of plausible real-world experiences. The research in Section 4.7 in chapter 4 will show that avatars play an important role when using shoulder surfing as an example of a security evaluation of USEC prototypes. Therefore, the existing literature about the concept of sense of presence (cf., Section 2.4.2.1) and user representations in VR (cf., Section 2.4.2.2) is reviewed.

2.4.2.1 Sense of Presence: Being Part of the VR Environment

The concept of presence, the “feeling of being there”, is rooted in teleoperator systems where users have the feeling of being at the place of a remote physical robot they are operating. Presence can be defined as the “subjective experience of being in one environment (there) when physically in another environment (here)” [554]. The factors that elicit a sense of presence, including genuine cognitive, emotional, and behavioural responses [415], together with the various questionnaires that allow researchers to collect data about the users' sense of presence [453], highlight the complexity (and inconsistencies [474]) of the concept, terminology, and application of presence.

“if one thinks that the level of social presence created by interaction with an mixed reality application is of interest, one should absolutely measure social presence using a widely-used instrument [...] However, one should not take this to mean that social presence is a construct relevant to every mixed reality experience, nor should one assume that an experience that elicits more social presence than another is superior. The constructs that determine the effectiveness

of a given application or experience are intrinsically linked to its purpose.” – Skarbez et al. [471]

Questionnaires such as the IPQ [410, 451, 452] or the Slater-Usoh-Steed questionnaire [478, 520] are intended to be used for virtual environments (cf., “Appendix. Comparison of Presence questionnaire attributes” in [287]). However, reality is quite different: various works have compared the sense of presence across different realities, for example, across AR and VR [523, 558], when blending in real-world snippets [314], or for non-immersive desktop VR settings [37, 65]. Despite the different interpretations of the term “sense of presence” and *how* and *where* presence questionnaires should (and can) be applied, when people behave in a virtual environment similarly to how they behave in an analogous scenario in the real world, then it can be argued that there is some sign of presence [474]. Sense of presence can be divided into two orthogonal components that contribute to VR users’ realistic behaviour [475]: *Place Illusion* (PI) and *Plausibility Illusion* (Psi). PI refers to a human’s sense of “being there” and how they perceive such a virtual environment, despite being certain that they are not (physically) there. In contrast to PI, Psi refers to the content that is actually perceived. It is an illusion that what is happening within VR is actually happening, although you can certainly tell it is not [475]. Assuming that a human feels “being there” (PI) and perceives the events that are happening within a virtual environment as happening (Psi), they are exposed to this experience and likely to behave as they would in reality [475, 476].

Existing research showed that improving visual and interactional realism leads to increases in presence [86]. For example, mimicking the tactile sensations and forces of genuine actions contributes towards interactional realism. Over the years, different VR prototypes that provide VR users with haptic feedback, therefore contributing towards interactional realism, have been proposed and evaluated, including stationary prototypes [249, 469], haptic proxies through handheld controllers [167, 261, 460], and haptic gloves [511, 516]. As this thesis will show, a location-mapped physical surface, similar to the one developed by Kim et al. [249], already contributes to some sense of perceived interactional realism (cf., chapter 4). In contrast, the absence of tactile feedback can negatively impact a user’s perception of authenticity when simulating real-world environments and prototypes in virtuality (cf., chapter 6).

The concept of the “sense of presence” has been applied in this thesis at several places to allow the thesis to comment on the participants’ sense of being part of realistic real-world scenarios and investigate their perception of reality when being exposed to VR replicas of real-world contexts and research artefacts. Section 5.3 in chapter 4 will apply the IPQ [451] to learn more about the users’ sense of presence when observing authentications using three different observation methods, including traditional 2D video observations [29, 308] and non-immersive/immersive VR observations (cf., Section 5.3.2.2). Similarly, Section 5.4 in chapter 4 will use the IPQ to compare the participants’ sense of presence across five different authentication scenarios and environments, including different realities (VR

and reality), simulated laboratory environments, and simulated public spaces. Finally, the research in chapter 6 will use the IPQ [451] and the TPI social realism questionnaire [287] to comment on the participants' sense of presence when experiencing a replicated real-world ATM authentication scenario in VR. The IPQ questionnaire [451] and its subscales, including spatial presence, involvement, realism, and sense of being there, have been consulted and applied in this thesis because the IPQ is a standard, widely-used measure for a user's sense of presence that is well understood and has found widespread application [453]. However, its deficiencies in learning more about social realism have led the research community to commonly supplement it with additional qualitative feedback and scales, which this thesis supports through additional qualitative data (for example, Section 5.3.5.4 and Section 6.6.8).

2.4.2.2 User Representations in VR

Several researchers have investigated the impact of avatars on VR users, including the perception of *self-avatars*, i.e., a 3D representation of a human that is aligned with their body and movements in reality, and of *others' avatars*, i.e., virtual bystanders that are represented through VR avatars. First, this section provides a short overview of research on VR avatars and the perception of self. It then reviews the literature on the impact of co-located VR avatars on a VR user (for example, the effect of avatars on co-presence, the "subjective experience of being together with others in a computer-generated environment, even when participants are physically situated in different sites." [564]).

Slater et al. [479] argued that the way humans and their bodies are represented in VR plays a primary role in virtual environments as it is the representation of self and lays the foundation of an interaction model for body-centred interactions. Several works found that a minimum level of self-avatar fidelity is required to change the user's perceptual judgements (for example, spatial perception [334]) and contribute towards realistic interaction in VR. Roth et al. [424] argued that the lack of behavioural cues such as eye gaze and facial expressions of VR avatars can be partly compensated, enabling VR researchers to use abstract avatars to provide universal generic representations of humans in a relatively affordable and straightforward way. Others found that a personalised cartoon-like virtual character exhibited a higher sense of body ownership and presence compared to a realistically reconstructed avatar from real imagery of the owner [223]. Steed et al. [489] concluded that active self-avatars enhance a VR user's ability to perform specific cognitive tasks and that self-avatars are essential for direct manipulation tasks. Gonzalez-Franco et al. [165] found that whenever the virtual body of a VR user does not align with the physical body, the user tends to unconsciously follow their avatar movements and compensate for the technological gap. Indeed, user performance and experience (for example, the sense of presence, embodiment, or danger) are not degraded by abstract visual user representations compared to avatars that manifest high

visual fidelity [289].

When it comes to the impact of VR avatars on others (for example, virtually co-located VR users), Piumsomboon et al. [386] showed that the presence of an adaptive avatar, a “Mini-Me” avatar which represents the VR user’s gaze direction and body gestures while it transforms in size and orientation, significantly improved social presence and the overall experience of collaborations in virtual environments. Saffo et al. [431] emphasised the importance of avatar selection in their guidelines for social VR studies and argued that it is important to have answers on the avatar design early on in the design process. Others suggested that abstract avatars with, for example, head and hands only, already produce an increased feeling of co-presence and behavioural interdependence [198]. Jo et al. [224] showed that a cartoon-like avatar created stronger co-presence than a conventional 2D video-based avatar but was less trustworthy than a realistic avatar. Casanueva and Blake [72] found significant differences between co-presence generated by realistic human-like avatars, cartoon-like avatars, and unrealistic avatars, with the human-like ones achieving the highest sense of co-presence. Others argued that avatars are the basis for creating realistic scenarios in VR [475] and can be used to “recreate and evaluate humans responses to realistic scenarios” [221].

Overall, the VR community has put in notable effort in defining how VR users can (and sometimes should) be represented in virtual environments. The choice of avatar often depends on the context, the investigators’ research questions, and the resources available [166, 289], which prohibits the community from defining a universal avatar suitable for all the various research interests, contexts, and applications. Virtual social environments like Mozilla Hubs [341] or Ubiq [148] use abstract avatars that show eyes and hands only. In contrast, Microsoft relies on highly realistic human-like avatars as part of their Rocketbox avatar library for research and educational use [168, 169], with follow-up work on animating motion-captured movements into the Microsoft Rocketbox library to provide researchers from various domains with interactive avatars [166].

As a result of the large variety of VR avatars and their effect on the VR users’ sense of presence and co-presence, Section 4.7 will explore the impact of different avatar fidelities on the bystanders’ interaction identification performance. In other words, it complements previous research that investigated the effect of avatars on the VR users’ behaviour and co-presence and compares, for the first time, how well different avatars are observable when performing human gestural movements, including touch gestures, mid-air gestures, and eye-gaze movements. The research in Section 4.7 will show that abstract avatars and their gestural movements are rich enough when the aim is to evaluate a USEC prototype’s security against observations, which Section 4.8 will then put forward through an in-depth security evaluation and comparison to a real-world study. To contribute towards a VR user’s co-presence and perceived social presence [47], and to provide user study participants with realistic virtual simulations of real-world contexts, research in chapter 5 and chapter 6 will make use of

animated human-like avatars as part of Adobe’s Mixamo avatar library [11] for simulated real-world bystanders. These avatars will be used in chapter 5 and chapter 6 as previous work highlighted their advantages over abstract avatars when aiming to elicit a sense of co-presence in VR [72, 224]. The avatars from Adobe’s Mixamo avatar library [11] were animated when simulating real-world contexts in Section 5.4 of chapter 5 and in chapter 6 as movement realism can create a strong sense of social presence, despite the absence of photorealism [175].

2.4.3 Summary

This literature review section first reviewed empirical VR research and how it differs from the HCI and USEC research fields. It then provided a short overview of two relevant VR concepts for the empirical VR research chapters in this thesis: the sense of presence, the key concept when studying user interactions in VR [79], and the different virtual representations of users in VR. The review has shown that empirical evaluations in VR range from traditional lab studies to remote VR studies where researchers make use of participant-owned VR headsets [400]. Research is now required to show how the discussed VR concepts are applied in VR studies that simulate real-world USEC research and allow for the usability and security testing of VR replicas of real-world contexts and research artefacts. This thesis will put forward such an investigation and considers the sense of presence and VR avatars – along with the usability and security research in the empirical research chapters in chapter 4, chapter 5, and chapter 6 – for the evaluation of VR replications of real-world environments and prototypes.

2.5 Authentication Prototypes and User Evaluations

So far, the literature review has touched on the broader USEC research field, its challenges, and on common research methods in HCI, USEC, and VR. It then discussed the VR concepts that will be consulted and utilised for the validation of VR studies for USEC research. Now that the thesis has reviewed and discussed how research is done in HCI and USEC and reviewed the VR literature relevant to this thesis, the final section of the literature review investigates the landscape of human-centred authentication prototypes and the empirical evaluation of authentication prototypes. Reviewing the authentication literature helps the thesis to narrow its scope from the larger USEC field to a more specific subdomain whilst putting the findings into the context of the broader field. For the research in this thesis, authentication, which is a major theme in USEC research [155] and the most addressed USEC research topic [110], remains in the spotlight and is reviewed in the next sections.

2.5.1 Authentication Research: An Overview

“The role of security is a supporting one—to protect assets and activities that users care about [...]” [437], a quote by Sasse and Fléchaïs that expresses the importance, but at the same time the supporting role, of authentication in people’s daily lives. Authentication, the process or action of verifying the identity of a user or process, is a recurring task. Florêncio and Herley [138] found that people use, on average, 8.11 passwords per day. However, going through an authentication process is usually not people’s primary task [437]. The secondary role of authentication can be illustrated with the following example: Imagine the scenario where a student wants to access course materials from the university’s online platform. In this case, the student’s primary goal is not to authenticate, but to access the course material.

To design unobtrusive, usable protection mechanisms whilst providing people with secure systems, researchers have put forward a large body of prototypes that aim to improve the usability and the security of authentications. These authentication methods are commonly classified into: 1) something the user knows (a secret), 2) something the user is (a property), and 3) something the user has (a secret token). In 2012, Bonneau et al. [53] highlighted that no authentication method retains the benefits that legacy passwords provide, hinting at the fact that the quest to replace secrets (for example, passwords) has not been solved [53].

As this review will show, biometric authentication methods (for example, Apple’s Touch ID) are often considered to be the holy grail for authentication. Systems of this kind are fast, easy to use, widely applied in commercial products (for example, fingerprint to unlock Android devices or Optic ID to unlock Apple’s Vision Pro reality headset [23]), and do not require users to recall secrets. However, despite their advantages over traditional knowledge-based authentication, biometric systems introduce privacy concerns [12, 84, 203, 559]. Furthermore, there are many situations in which biometrics do not work (for example, when users have wet fingers or due to low lighting conditions [38, 391]). Herley and van Oorschot [203] emphasised that it is challenging to establish a minimally privacy-invasive biometric solution that protects users to the same extent as knowledge-based authentication:

“Repeated and sustained effort has failed to uncover a silver-bullet replacement for passwords. It’s time to admit that this is unlikely to change. No single alternative technology is likely to possess the combination of security, usability, and economic features that meet all goals in all situations” – Herley and van Oorschot [203]

Given the biometrics’ existing shortcomings, it is unlikely that knowledge-based authentication mechanisms will disappear entirely from the surface, at least not soon. The research presented in this thesis tackles the authentication problem at its source, which is authentication based on the knowledge factor [53]. It reviews the entire authentication research, including (behavioural) biometric prototypes, but the focus of validating the use of VR studies is on

Table 2.1: A set of scientific contributions on authentication research that involve at least one human-centred evaluation and at least one prototype published in ACM TOCHI, ACM CHI, and USENIX SOUPS. Contributions are sorted by venue and publication year.

Authors	Venue	Year	Study Type			# of Participants	Main Prototype Implementation	Device	Authentication Type	Evaluation	
			Lab	Online	Field					Usability	Security
Everitt et al. [132]	CHI	2009	–	✓ _{lab}	–	N = 110, N = 69	software	traditional desktop	knowledge	✓	–
De Luca et al. [101]	CHI	2009	✓	–	–	N = 24	hardware	public display	knowledge	✓	–
Kim et al. [248]	CHI	2010	✓	–	–	N = 21	software	multi-touch device	knowledge	✓	✓
De Luca et al. [96]	CHI	2012	✓	–	✓	N = 48, N = 34	software	smartphone	knowledge	✓	✓
Sae-Bae et al. [429]	CHI	2012	✓	–	–	N = 34	software	multi-touch device	knowledge and biometrics	✓	✓
De Luca et al. [102]	CHI	2013	✓ _{lab}	–	–	N = 20, N = 24	hardware	smartphone	knowledge	✓	✓
De Luca et al. [98]	CHI	2014	✓ _{lab}	–	–	N = 32, N = 4	hardware	smartphone	knowledge	✓	✓
Buschek et al. [63]	CHI	2015	✓	–	–	N = 28	software	smartphone	knowledge and biometrics	–	✓
von Zezschwitz et al. [530]	CHI	2015	✓ _{lab}	–	–	N = 18, N = 3, N = 16, N = 3	software	smartphone	knowledge	✓	✓
Holz et al. [207]	CHI	2015	✓	–	–	N = 12	software	smartphone	biometric	–	✓
Hamdy et al. [187]	ToCHI	2011	✓	–	–	N = 274	software	traditional desktop	biometric	✓	✓
Mathis et al. [308]	ToCHI	2020	✓	–	–	N = 23, N = 21, N = 15	software	VR headset	knowledge	✓	✓
Wiedenbeck et al. [543]	SOUPS	2005	✓	–	–	N = 32, N = 83	software	traditional desktop	knowledge	✓	–
Tari et al. [506]	SOUPS	2006	✓	–	–	N = 20	software	traditional desktop	knowledge	–	✓
Chiasson et al. [75]	SOUPS	2007	✓	–	✓	N = 43, N = 191	software	traditional desktop	knowledge	✓	–
Forget et al. [140]	SOUPS	2008	✓	–	–	N = 83	software	traditional desktop	knowledge	✓	✓
Hayashi et al. [196]	SOUPS	2008	✓ _{lab}	–	–	N = 6, N = 54, N = 45	software	smartphone	knowledge	✓	–
De Luca et al. [95]	SOUPS	2009	✓ _{lab}	–	–	N = 10, N = 21, N = 1	software and hardware	traditional desktop	knowledge	✓	✓
Zakaria et al. [565]	SOUPS	2011	✓ _{lab}	–	–	N = 68, N = 30	software	personal digital assistant	knowledge	✓	✓
Hayashi et al. [195]	SOUPS	2013	✓	–	✓ _{lab}	N = 128, N = 32, N = 18	software	smartphone	knowledge	✓	✓
Harbach et al. [191]	SOUPS	2014	–	✓	✓	N = 260, N = 52	software	smartphone	any	✓	✓
Xu et al. [562]	SOUPS	2014	–	–	✓	N = 32	software	smartphone	biometric	✓	✓

knowledge-based authentication methods. Yet, investigating and validating the use of VR studies for empirical research on (behavioural) biometrics surely deserves some attention in the future and can build upon the findings of this thesis.

2.5.2 Design Space of Authentication Prototypes

The design space of authentication prototypes covers novel authentication methods for multiple devices, including novel authentication schemes for public displays [101, 118, 537], smartphones [180, 239, 243], smartwatches [177, 353], smart homes [323, 390], and VR/AR headsets [150, 158, 308]. A query-based¹⁰ review of scientific contributions within the ACM TOCHI, ACM CHI, and USENIX SOUPS proceedings from 2001 until 2022 (inclusive) has been conducted to provide a glimpse into the design space of authentication research that involves prototypes and user evaluations. ACM TOCHI is the flagship journal of HCI research, CHI is the premier international conference of HCI research, and SOUPS is the premier venue for interdisciplinary research on HCI, security, and privacy. These proceedings were reviewed because a large percentage of USEC papers are published at these venues, particularly at SOUPS and CHI [110].

The query-based review resulted in 157 eligible papers for CHI, 68 for SOUPS, and six for TOCHI. The papers were then ordered from *most cited* to *least cited* and filtered based on two criteria: including at least one prototype and at least one user evaluation. The ten most cited papers from each venue that involve at least one prototype and at least one user

¹⁰The search query was: “”query”: Title:(authentication) OR Keyword:(authentication) OR Abstract:(authentication) ”filter”: Conference Collections: CHI: Conference on Human Factors in Computing Systems, E-Publication Date: (01/01/2001 TO 06/30/2022) ” for CHI in June 2022. The same query-based review was conducted within the TOCHI and the SOUPS proceedings. For SOUPS, the proceedings were additionally manually browsed to cross-validate the review.

evaluation are reported in Table 2.1. Note that there are only two articles published in ToCHI that comprised an authentication prototype and an empirical study. The list in Table 2.1 is by no means exhaustive, and it is not the aim of the review to provide a holistic overview of the entire authentication landscape. However, Table 2.1 sets the context for the following sections by showcasing the study types and evaluations of well-cited authentication works that involve prototypes and user testing. As Table 2.1 shows, a large body of authentication prototype research took place in the lab. From the table presented, two works were conducted online, five in the field, and the remaining studies were conducted in conventional laboratory settings. The majority of the works discuss software prototypes, and most of them employ knowledge-based authentication. Sample sizes range from one participant [95] up to 274 participants [187], depending on the nature of the study.

In the following sections, the literature in all three knowledge-based authentication categories defined by De Luca et al. [101] is reviewed: software-focused authentication solutions, hardware-focused authentication solutions, and solutions based on the users' personal devices. Authentication methods that make use of a variety of input methods such as eye-gaze input [6, 243, 245], mid-air gestures [26, 245], and touch gestures [96, 102, 245] are discussed. Finally, a review of biometric authentication methods, the study types commonly applied in authentication research, and domain-specific research challenges is put forward.

2.5.2.1 Knowledge-Based User Authentication

This section discusses the variety of knowledge-based authentication prototypes, including software authentication solutions, hardware authentication solutions, and authentication methods based on hardware owned by the users.

Software-Focused Authentication Solutions. Software solutions aim at solving the authentication problem on a software level. They often strive to make observations, like shoulder surfing [125], more challenging even if attackers can observe both the input and the output of an authentication procedure. De Luca et al. [99] proposed ColorPIN, a highly secure software-based authentication scheme for ATMs that maps a colour and a PIN element to a key on a traditional keyboard. Kim et al. [248] evaluated different software-based authentication methods for tabletops in public spaces that exploit multi-touch interaction to inhibit shoulder surfing. In their prototype, the user increases the pressure on one finger per hand in predefined coloured pressure zones to communicate an (x,y) coordinate and select the corresponding object in the grid. Tan et al. [505] proposed an onscreen virtual keyboard authentication method that allows for private input by utilising a randomised keyboard layout and cycling through shift states to indicate the user's intended input. Roth et al. [425] proposed cognitive trapdoors as a highly secure PIN-entry method. Users are presented with digits in black or

white colour. Depending on their PIN, they need to press either the black or the white button several times. The system then determines the correct PIN digit by intersecting multiple distinct PIN element sets [425]. All these prototypes utilise software implementations to enhance authentication and provide end-users with usable and secure authentication methods.

Hardware-Focused Authentication Solutions. In contrast to software solutions, hardware solutions make use of additional hardware for improved security and usability. Dunphy et al. [118] employed eye-tracking hardware to evaluate gaze-contingent passwords at ATMs, highlighting the use of eye gaze for user authentication in public and its resistance against observations. Sasamoto et al. [434] proposed an authentication system that requires users to identify the location of their password element by moving a trackball to different positions. While a graphical display presents several images, the trackball simultaneously informs the user about the mapping of the buttons, making observations close to impossible [434]. Deyle et al. [107] proposed a tactile PIN entry mechanism that can be operated without any visual output. Their hardware prototype utilises solenoids with pins that can be raised/lowered by applying electric current to an embedded electromagnet [107]. Bianchi et al. [42] protected user authentications from observations using tactile cues. Their prototype hardware encodes passwords as a sequence of randomised vibration patterns, making it impossible for an attacker to observe the input. Marky et al. [299] presented 3D-Auth, 3D-printed physical hardware tokens for advanced user authentication. Similarly, Hwang et al. [214] proposed passive control widgets for tangible interaction on and around mobile devices, with user authentication as one potential application.

Authentication Solutions Based on the Users' Personal Devices. The third category, solutions based on the users' devices, makes use of hardware the users typically already own. Guerar et al. [176] proposed an authentication system that relies on the user's smartphone and uses a QR code to match colours to digits. Sharp et al. [457] used the user's personal device to view a one-time password for authentication at public displays. Nyang et al. [361] proposed utilising the user's smartphone to obtain a random permutation of a keyboard layout to authenticate in public. De Luca et al. [101] used the user's mobile device for secure authentication based on shared "lies" on a public terminal. Their prototype utilises the user's phone for tactile feedback and provides secret information when to add an overhead of "lies" to the input [101]. Khan et al. [247] used a mobile device to allow for obfuscated PIN template input. Users receive a PIN template (for example, $[48^{**}29^{**}]$) on their device that they then combine with their PIN to authenticate. Winkler et al. [553] tested a private near-eye display to communicate keypad layouts to users when authenticating on a mobile device, whilst Yi et al. [563] proposed leveraging the user's smartwatch and acoustic tones to unlock mobile devices.

In summary, the discussed prototypes rely on the knowledge factor (i.e., something the user knows); however, each prototype is unique in its design and may involve software of hardware components to a greater or lesser extent in contrast to other prototypes. The range of knowledge-based authentication methods highlights the variety of authentication prototypes, which are further supplemented by biometric authentication methods.

2.5.2.2 Biometrics for User Authentication

Although the classification of knowledge-based authentication into software-centred, hardware-centred, and solutions based on users' devices can, theoretically, also be applied to (behavioural) biometric authentication methods, biometric systems rely on the inherence factor, i.e., something the user is. Therefore, contrary to knowledge-based authentication systems, biometric authentication systems are often classified into two groups: physiological biometrics and behavioural biometrics. Whilst the former is about "something the user is", like fingerprint or heart rate, behavioural biometrics are concerned about "how the user behaves", involving parameters such as a person's keystrokes when typing [335] or their movements [381].

One of the first smartphones with a biometric authentication method, i.e., fingerprint sensor, dates back to the year 2007 [32]. Since then, the human-centred security community has proposed and investigated users' perceptions of a plethora of biometric authentication mechanisms, some of which are reviewed in the following paragraphs.

Physiological Biometrics for Authentication. De Luca et al. [97] conducted an online survey to better understand people's reasons for using or not using Apple's Touch ID and Android's Face Unlock mechanism on mobile devices. They found that usability and security were the most mentioned reasons for using, not using, or deactivating Touch ID and Face Unlock [97]. Schneegass et al. [448] proposed "SkullConduct", a biometric system that uses bone conduction of sound that traverses through the user's skull. Osadchy et al. [374] introduced SCiFI, an authentication system for secure and privacy-preserving computation of face identification. Singh et al. [470] investigated biometric electrocardiogram (ECG) for user authentication and combined their ECG-based biometric system with face biometrics and fingerprint biometrics for increased security.

Despite the promising results and applications of biometric authentication systems, Prange et al. [391] found that the usability of biometric authentication systems often depends on the context. For example, wet fingers impact a biometric system's functionality [266, 391]. Matsumoto et al. [310] pointed out the problem of impersonation by using artificial "gummy fingers". There are many more works on physiological biometrics for authentication, including works that highlight challenges regarding accessing sensor data at a large scale [77, 422] and data sharing practices that put users' privacy at risk [203, 327, 388].

Behavioural Biometrics for Authentication. As a result of the long-lasting password problem [53], and the drawbacks of physiological biometrics, a plethora of behavioural biometric authentication mechanisms have been proposed. Pfeuffer et al. [381] studied how different body motions can be used for behavioural biometric authentication in VR. Shen et al. [459] used gait recognition to authenticate users in mixed reality applications. Similarly, Mustafa et al. [345] used head, hand, and body movements for user authentication while participants had to move towards virtual balls that randomly appeared within a virtual environment. A recent work by Liebers et al. [282] investigated the feasibility of implicitly identifying users by their hand tracking data in virtual environment. They found that users are identifiable with up to 95% accuracy across sessions using an explainable machine learning approach on the users' hand tracking data [282].

Despite the promising examples of using behavioural biometrics for authentication, there is research that highlighted the users' reluctance in adopting them and that behavioural biometrics may not work in the various authentication contexts. For example, Mecke et al. [323] investigated the users' perceptions of physical, biometric, and behavioural authentication concepts to open doors, finding that participants value being in control of the authentication through the possession of a physical token (i.e., a key). Miller et al. [331] highlighted a noticeable drop in the authentication accuracy when considering cross-system behaviour-based authentication mechanisms (for example, when authenticating across various VR/AR headsets).

In summary, whilst biometric authentication is considered to be fast and implicit, it often requires sharing personal data with third parties, which can be (and have been [113, 508]) stolen. It is challenging to change biometric passwords [492] and users can be forced into using their physiological biometrics without consent. Additionally, not all users are willing to use biometric authentication [388] and applications would require long-term, permanent access to sensor data at a huge scale [422], raising ethical questions as the collection and storage of sensitive data can result in a leak of personal data [12, 84, 422, 559].

Some of the disadvantages of traditional biometrics motivated Liebers et al. [283] to investigate "Functional Biometrics". Functional biometrics treat the human body as a function that transforms a stimulus from an underlying authentication system and outputs a resulting body reflection metric, which is received as a characteristic response by the authentication system. An excellent authentication example that made use of the concept of functional biometrics is the work by Schneegass et al. [448] in which they developed and tested a research prototype that utilises the bone conduction of sound through the user's skull for user authentication. Their prototype implementation achieved a user identification accuracy of up to 97.0% when comparing the data from ten participants [448].

Whilst functional biometrics can overcome some of the existing challenges of biometric authentication, for example, the lack of changeability of biometrics, at the time of writing this

thesis, it is unclear how such systems can be integrated into existing systems and how they are perceived by users. This is not to say that biometric authentication systems are not promising, as evidenced by some widespread adoptions in commercial systems (for example, Apple's Touch ID); however, many challenges must be addressed before biometric authentication can deliver reliable and secure authentication. Biometric schemes often require a non-biometric fallback method, such as entering a 4-digit PIN when fingerprint authentication authentication does not function. This is another reason why knowledge-based authentication systems have a reason to exist and are likely to be part of people's lives in the (near) future. Therefore, whilst this thesis put forward a review of both the knowledge-based authentication field and the biometric authentication field, the validation of using VR studies for real-world USEC research focuses on knowledge-based authentication in the remainder of this thesis. Investigating the use of VR studies for biometric authentication prototypes is an interesting direction, and although beyond the scope of this work, it surely deserves some attention in the near future (for example, to build upon the Wizard of Oz study by Mecke et al. [323] and simulate realistic behavioural biometric authentication scenarios in VR).

Now that the review has discussed the broader human-centred authentication field and the corresponding prototypes, the next section concludes the thesis' review by discussing common empirical research methods when evaluating authentication prototypes and highlighting the associated research challenges.

2.5.3 Authentication Prototypes and Empirical Research

USEC researchers often deal with the usability and security of their authentication prototypes. Considering both is important because prototypes that are not usable will be misunderstood by users, likely putting their security at risk. Kainda et al. [231] argued that there is a "blurred line dividing usability and security factors – some of the usability factors cause users to behave insecurely, and some of the security factors obviously impair performance". Sasse and Fléchain [437] emphasised that systems are often too complex for many users and that the research community needs to do better in designing and implementing USEC artefacts.

Evaluations of authentication prototypes can be roughly classified into usability and security, but usability and security investigations vary a lot [56, 544]. Many factors affect the users' experience in knowledge-based user authentication, including human factors (for example, cognitive disabilities, age, cognitive styles) and technological factors (for example, the device used for interaction) [236]. Contextual factors such as the age of the users can further impact a system's usability [355] and the users' security preferences [396]. Over time, the USEC community has established some "best practices" when designing and evaluating authentication prototypes. For example, *task efficiency*, the time needed to authenticate, *task effectiveness*, the number of authentication attempts, and *the users' preferences* are prominent

usability metrics when proposing and investigating novel authentication prototypes [236]. For security evaluations, researchers often opt for trained attackers who observe authentications [43, 101, 308] or perform a theoretical security analysis [4, 99, 180].

In contrast to the broader USEC field, where online surveys are widely applied to forge forward USEC research, the evaluation of authentication prototypes often requires lab studies that allow participants to experience the prototypes in action and in a controlled (lab) environment (cf., Table 2.1). There are some evaluations of novel authentication methods based on online studies [132, 141], but those are often the exception and only feasible if the prototypes do not rely on specialised or custom-built hardware. The design, implementation, and evaluation of USEC prototypes introduce unique challenges, which this thesis reviews in the next section.

2.5.4 Research Challenges in the Authentication Research Field

Garfinkel and Lipford [155] put forward key challenges in the broader USEC field, including the inherently interdisciplinary nature and the user evaluation challenge where users are not in a position to evaluate a system's security or their security behaviour. The following sections put a variety of the more general USEC challenges in the context of authentication research and discuss them based on existing works in the literature.

2.5.4.1 The Interdisciplinary Nature of Authentication Research

Authentication research is interdisciplinary in its nature. A famous example that showcases the interdisciplinary nature of authentication research and its challenges is the USEC prototype by De Luca et al. [102]. Their prototype consisted of two smartphones attached to each other and presented users with a front and back screen for input. The form factor of their initial research artefact, including the size and the weight, was a burden for one-handed interaction [102]. As a result, the usability results were not generalisable. A year later, De Luca et al. [98] built a thinner and lighter hardware prototype using 3D printed cases and investigated its impact on the usability findings. They found that a more professional research artefact, together with an enhanced algorithm, positively impacted the user experience, highlighting the importance of prototype fidelity when conducting usability evaluations. In a similar vein, several of Mecke et al.'s prototypes required expertise in building hardware prototypes despite not being fully implemented [323]. Others such as Islam et al. [219] developed a touch sensor for augmented reality glasses for usable and secure authentication. They created an authentication prototype using two printed circuit boards, a touch sensor, and an Arduino (a single-board microcontroller). Similarly to De Luca et al. [102], they discussed an improved version of their prototype to more practically integrate it into eyewear and to investigate the impact of the prototype's size on usability [219].

The discussed examples touch on the interdisciplinary nature of USEC research: researchers are often required to have expertise in building hardware prototypes, conducting empirical research, and considering the impact of the prototype's fidelity and form factor on usability and security evaluations. Designing, building, and evaluating authentication prototypes often requires knowledge in HCI, engineering, USEC, and the expertise from many more neighbouring research areas to fulfil the overarching goal: eventually providing end-users with usable and secure systems.

2.5.4.2 Authentication Research and User Evaluations

Compared to traditional HCI research, where a lot of emphasis is on the usability of systems, USEC research is additionally concerned with the user behaviour and how contextual factors impact a system's security. Section 2.3 has already touched on user evaluations in the broader USEC space. This section discusses the more specific evaluation challenges when exploring the strengths and weaknesses of authentication prototypes.

Empirical evaluations in authentication research are particularly challenging due to the authentication's secondary role in reality [437], the ethical and legal constraints when going into the wild [100, 529], and the ecological validity concerns when conducting and simulating authentication research in the lab. For example, instructing participants to pay attention to privacy and security in a lab study biases them and does not represent a realistic scenario someone would experience in real life [482]. Whilst in situ studies are possible, they are often accompanied by ethical and legal constraints [100, 529], restricting the USEC research that is possible in the wild. An excellent example is the field study of real-world ATM use by De Luca et al. [100]. Their observations led to forward-looking insights into the people's behaviours when interacting with and authenticating on ATMs; however, ethical and legal constraints prohibited detailed video recordings of the keypad and the users' inputs. The challenges around ecological validity, i.e., the extent to which user study participants behave the way they would in real life, are commonly known in the USEC domain [133, 259]. However, solving the lack of high ecological validity completely remains challenging due to the authentication's secondary role [437]. To contribute towards solutions to these challenges, researchers introduced deception to simulate, for example, an adversary's presence [85] or attacks during user studies [110, 121]. Dunphy et al. [118] recreated the sights, sounds, and experiences that are typical of ATM interaction in public spaces to represent a realistic authentication scenario. However, their simulation was still restricted to what is possible to simulate in a physical laboratory environment. Some other attempts to contribute towards realistic USEC research include filtering participants using self-reported data [133], using online platforms such as Android's Play Store to receive feedback by users who authenticate on their own device [22], and utilising crowd-sourcing platforms, such as MTurk, to study

larger and more diverse user samples [58].

The work presented in this thesis will complement and advance the existing USEC research arsenal. Chapter 5 and chapter 6 will show how the use of VR studies facilitates more realistic USEC research through user evaluations in reasonable approximations of real-world scenarios without being required to access public and security-sensitive environments in the wild.

2.5.4.3 Many Roads Lead To Usability and Security Evaluations

Finding a “perfect” authentication method is challenging and often a balance between various usability and security factors [53]. The USEC community makes use of many widely-applied usability and security metrics (cf., [236, Table 1], [236, Table 2], and [53, Table 1]). However, these metrics may be more relevant for a given context, and vice versa. For example, a highly shoulder-surfing resistant authentication method that uses non-visual interaction for discreet PIN entry [108] is more suitable for low-frequency usage than the traditional PIN entry method due to its additional time costs. Yet, such an authentication mechanism might not be suitable for contexts that require high-frequency usage (for example, when frequently unlocking a smartphone). Other examples include *The Memory Palace* by Das et al. [91], an authentication system that encodes authentication secrets as paths along a 3D virtual labyrinth, which is significantly more memorable but slower than Android’s 9-dot pattern. Das et al. [91] highlighted the system’s usefulness for infrequent authentication scenarios where speed plays a secondary role, but acknowledged the shortcomings in scenarios where authentication may be more frequently required.

As for security, researchers often define a threat vector, a potential risk to the system’s security. *Threat modelling* is defined as the formal process of identifying, documenting, and mitigating security threats to a system [368]. For example, an authentication method that uses smart glasses for smartphone unlocking [553] protects the users from smudge attacks, shoulder surfers, and camera attacks. However, the technique has not been designed against the many other threats that might break the system’s security, such as man-in-the-middle attacks that give attackers access to the AR user’s private near-eye display.

There are many threats to authentication systems, including having access to a list of hashed passwords due to data breaches [237] or shoulder surfing other people during their authentication [125, 243]. Accounting for the many threat vectors is often impossible and requires researchers to prioritise threats based on their likelihood and security severity.

“There are of course no set of rules, principles or formalisms that, when followed, are guaranteed to produce usable computer systems. If such rules existed, we would almost certainly all be using them, and the usability problem would be solved.” – Garfinkel [154]

In line with Garfinkel’s comments on designing usable systems [154, p. 43], the key message here is that there are, of course, no rules or research methods that guarantee usable and secure authentication prototypes and allow for strict comparisons across different works. However, as this thesis will show in chapter 3, there is a lack of well-evaluated research methods that allow for realistic usability and security evaluations of USEC prototypes and highlight the potential differences to the results from lab evaluations.

2.5.4.4 The Continuously Changing Technological Landscape

The technological landscape is changing rapidly, but designers and developers often miss the opportunity to consider the impact of novel devices and input methods on security and privacy. A famous example of a rapid technological change is the replication study by Sotirakopoulos et al. [482] who attempted to replicate a study on SSL warnings by Sunshine et al. [496]. Although the work by Sotirakopoulos et al. [482] was only conducted two years later, the replication was close to impossible due to the change in the underlying web infrastructure of browsers. A similar trend can be seen in the vast number of devices for which authentication methods need to be designed. Table 2.1 highlights the range of authentication prototypes and their form factor, moving from more traditional desktop settings to smartphones and head-worn devices, such as VR/AR headsets. Alt and von Zezschwitz [17] argued to “fundamentally rethink the way in which we design new technologies” [17] as the proliferation of new technologies forms several major security and privacy challenges.

It is often unclear what the new technologies are capable of collecting and how data collection can facilitate, but also negatively impact, people’s security and privacy. In an area where various technologies have found widespread adoption, researchers must think outside the box when defining threat models for authentication prototypes. The need for novel authentication methods that consider the changing technological landscape can be further demonstrated with an example from the VR/AR domain: whilst established authentication methods such as the Android’s unlock pattern can be adopted for new technologies (for example, for authentication in VR [158]), individual characteristics of new technologies, such as the third dimension of VR/AR [150, 308], open up opportunities to contribute towards more usable and secure authentication methods. For instance, research has shown that using the third dimension of virtual environments in combination with eye-gaze input results in highly usable and secure authentications [157, 308]. However, due to the fast pace of the technological landscape and the primary applications of these technologies, which are often not about security and privacy, technology has often found its way into society before detailed USEC research has been conducted and looped back to the systems’ designs. One such example is Meta’s VR parental supervision tool¹¹ that employs pattern-based authentication. Pattern-based authentication

¹¹<https://www.oculus.com/blog/introducing-future-vr-parental-supervision-tools-to-help-support-families>,

provides some level of added security, but the existing literature shows that patterns are prone to observations and researchers have already proposed various more secure and usable authentication methods in the last few years [157, 229, 308, 367].

2.5.5 Summary

Section 2.5 of the literature review contextualised the term “authentication” and reviewed the existing authentication research landscape, including various prototypes, how research is being conducted in this space, and some of the existing research challenges.

The literature is crowded with authentication prototypes, but there are only little efforts to move research artefacts, which are often evaluated in laboratory settings (cf., Table 2.1), out of the lab and contribute towards generating real-world impact. The overarching challenges when conducting human-centred security research have been discussed by others, most notably by Garfinkel and Lipford [155], who highlighted its interdisciplinary nature and the challenges around ecological validity.

There needs to be a change in how authentication research is being conducted to solve the existing and upcoming challenges in an era where technology becomes (or even already is) ubiquitous and changes rapidly [17]. Addressing the challenges USEC researchers experience when designing, building, and testing USEC prototypes, while facilitating and effectively accelerating research in this space, is a first step towards supporting the transition of USEC research into practice and keeping up with the fast-paced technological landscape.

The final section of this literature review summarises the main lessons learned from reviewing the broader HCI, USEC, and VR research fields. It concludes with a discussion of how this thesis puts forward a complementary research method that aims at supporting and facilitating USEC research involving real-world prototypes.

2.6 Summary and Conclusions of Literature Review

The literature review discussed empirical research in HCI, USEC, and VR. It then reviewed the authentication research landscape, including prototypes, research methods, and the current research challenges. As discussed in section 2.2.5, the use of VR studies for simulating real-world research has shown some promising applications in the broader HCI field. However, it is unclear how these findings transfer to the USEC field and what the challenges are when applying VR studies as a complementary research method in USEC. Although individual research projects discuss some obstacles USEC researchers experience during the

implementation and evaluation of their USEC prototype, as reviewed in Section 2.5.4, many of the more large-scale and overarching challenges are hard to map out in individual research papers. Understanding the existing hindrances in USEC research on prototypes is essential to map out where VR studies can best support and facilitate research in this space. Therefore, the lack of a longitudinal and rich set of experiences and challenges USEC researchers face in their work has motivated the first research question:

RQ₁ What are the challenges that USEC experts experience when designing, implementing, and evaluating security and privacy-enhancing prototypes?

Section 2.2.5 has discussed how HCI and neighbouring research fields have found application for VR studies to, for example, simulate field studies on public displays [291] or conduct detailed studies of the use of research artefacts in situ and in VR [528, 539]. However, there is still a research gap in the use of VR studies in USEC research and it remains unclear where VR studies can facilitate and forge forward research in this space. The experts' challenges and experiences when conducting USEC research involving prototypes in chapter 3 and the lack of an understanding of the applicability of VR studies for simulating real-world USEC research have motivated the following two research questions:

RQ₂ Which findings of VR-based usability evaluations on USEC prototypes match the findings from corresponding evaluations in traditional physical lab settings?

RQ₃ Which findings of VR-based security evaluations on USEC prototypes match the findings from corresponding evaluations in traditional physical lab settings?

As this thesis discussed in Section 2.2, there is no “one-size-fits-all” research method, and the extent to which VR studies are suitable for USEC research has not been explored until now. Building on fundamental research that provides the first evidence of the use of VR studies for USEC research in chapter 4, the overarching aim must be to facilitate and advance research in this space and work towards generating real-world impact. These needs have motivated two additional research questions that promote USEC research by a) using VR studies to contribute towards building a bridge over the methodological gap between lab studies and field studies (chapter 5), and b) moving traditional USEC research on real-world prototypes out of physical laboratory environments (chapter 6). Whilst chapter 5 contributes novel research methods for shoulder surfing research and VR-simulated in situ authentication research, it simultaneously contributes to the validation of VR studies for USEC research:

RQ₄ Can substitutional in situ studies using VR provide a bridge over the methodological gap between lab and field studies?

RQ₅ Can the use of VR studies move traditional USEC research on real-world prototypes out of physical labs?

The remaining chapters will address the research questions in the order they have appeared in this thesis, starting with RQ₁ in chapter 3 that presents nine key challenges impeding artefact contributions in USEC, including challenges that have not seen in-depth discussion in prior literature. Research in chapter 4 and chapter 5 will lay out the foundation of establishing VR studies as a complementary research method through several replication and comparison studies between VR and state-of-the-art research methods in the lab (RQ₂, RQ₃, and RQ₄). Extensive comparisons between the results collected in VR and lab studies will establish the use of VR studies for prototype-focused USEC research and provide the basis for moving traditional USEC research out of physical labs, which will be showcased in chapter 6 (RQ₅).

The final research question is answered based on the breadth and depth of the research conducted in this thesis. The findings and lessons learned from the research in chapter 4, chapter 5, and chapter 6 highlight the advantages and disadvantages of using VR studies for usability and security evaluations of USEC prototypes. Chapter 7 then answers an overarching research question (RQ₆) to provide a complete picture of this work and conclude the thesis:

RQ₆ What are the advantages and disadvantages of using VR studies for USEC research involving prototypes compared to traditional USEC research in physical laboratories and in the field?

III

SCOPING THE PROBLEM: KEY CHALLENGES IN
USABLE SECURITY RESEARCH

Chapter 3

Scoping the Problem: Key Challenges in Usable Security Research

This chapter is based on the following publication:

[Publication 3] Mathis, F., Vaniea, K., & Khamis, M. (2022). Prototyping Usable Privacy and Security Systems: Insights from Experts. In *International Journal of Human–Computer Interaction*. Taylor & Francis, DOI: [10.1080/10447318.2021.1949134](https://doi.org/10.1080/10447318.2021.1949134)

3.1 Introduction

Prototyping is an integral part of human-centred research and design [134, 365, 556]. Wobbrock and Kientz [556] emphasised that artefact contributions are one of the main research contributions in HCI: where researchers design innovative prototypes, tools, and techniques that demonstrate novel forward-looking possibilities or generate new insights through implementing and evaluating the prototypes. However, as discussed in the Introduction and the Literature Review, conducting USEC research that involves prototypes comes with unique challenges, including hardware deployments in ecologically valid contexts and evaluations with adequate sample sizes.

This chapter provides the first interview-based insights into the USEC experts' challenges when designing, implementing, evaluating, and publicising prototype-focused USEC research to discover where this thesis can support, facilitate, and advance USEC research and contribute towards the transition of research artefacts into practice. Interviews with twelve expert

and nascent USEC researchers from academia and industry who have made significant contributions to USEC research and whose work involved prototyping novel systems were conducted to unveil and better understand their research challenges. The interviews contribute answers to the first research question of this thesis:

RQ₁ What are the challenges that USEC experts experience when designing, implementing, and evaluating security and privacy-enhancing prototypes?

The chapter presents nine key challenges impeding artefact contributions in USEC, including challenges that have not seen in-depth discussion in prior literature. For example, the implementation challenges due to the scarcity of appropriate hardware; the difficulties in conducting ecologically valid studies, especially when evaluating hardware USEC solutions; and the lack of publication venues where novel and well-evaluated USEC systems are encouraged. The insights from the USEC experts coupled with the in-depth discussions of the results are valuable to the USEC community as well as to neighbouring research communities and inform the empirical research in chapter 4, chapter 5, and chapter 6 of this thesis.

3.1.1 Chapter Structure

Section 3.2 describes the methodology applied to elicit insights into the USEC experts' research challenges. Section 3.3 outlines the outcomes of the interviews, which were analysed as described in Section 3.2. The results are discussed in Section 3.4. The closing section, Section 3.5, will revisit the first research question of this thesis and outlines the next steps.

3.2 Methodology

This section describes the recruiting process of the experts, the structure of the interviews, the research approach and analysis, and some potential limitations of the methodology.

3.2.1 Recruiting USEC Experts

The interview study went through an ethical review by the University of Glasgow College of Science & Engineering ethics committee (ref: #300190041). Potential interviewees were selected to obtain a mix of researchers and practitioners who work at the intersection of HCI and USEC research. Experts who published works at USEC venues, for example, ACM CHI and USENIX SOUPS, and had hands-on experience in designing, implementing, and evaluating USEC prototypes were recruited. A rough literature review was conducted to compose an initial list of suitable authors and potential interviewees. Additional people were

added based on their HCI and USEC expertise to fill out the list. The ACM Digital Library, IEEE Xplore, and Google Scholar was used to find scholars with published USEC work at highly ranked HCI and security venues, for example, ACM CHI, USENIX SOUPS, and IEEE S&P. Broad search terms like “*usable security*”, “*usable privacy*” formed the basis of the search, which were followed-up with more specific search terms that were relevant for the research scope: “*security prototype*”, “*privacy prototype*”. The papers were then reviewed to identify those that included building security and/or privacy-protection solutions and at least one user-centred evaluation. To further improve the coverage, a snowball approach was applied: the references in the papers were reviewed for relevant titles and added to the list of reviewed publications. Google Scholar and the dblp computer science bibliography were used to determine the publication profile and the experience of the identified authors in the area. The relevant identified publications were recorded for later use in the interviews. A pool of 56 potential interviewees with significant expertise in USEC and prototyping was identified. Two researchers sorted the list with an eye towards multiple variables: selecting people with a range of seniority, university, industry, country, research domain, and experience publishing prototype papers in USEC venues. Researchers who were more senior and had recent USEC prototype publications were ranked higher.

Invitations were sent out to twenty potential interviewees as a first step (cf., Appendix B). The invitation asked if the person was willing to be interviewed about their research. Although recruiting senior people is time-consuming and challenging, fourteen responses from twenty invitations (70%) were secured. Two declined due to unavailability, the remaining twelve agreed to participate. Eleven interviews took place via Skype and were audio and video recorded with consent. One preferred an email interview, which is a viable alternative [324]. While progressing through the interviews, a few novel insights emerged after the tenth interview. Two more interviews were conducted, and nothing new was observed in the twelfth interview (theoretical saturation) [178]. Therefore, no additional requests were sent out.

3.2.1.1 Demographics of the Experts and Interview Material

The final interview sample consisted of 12 USEC experts (4 females, 8 males). Interviewees were from North America, Europe, and Asia, and work in academia (6), industry (2), or in both academia and industry (4). At the time of the interviews (November and December 2019), 10 interviewees held a PhD (1 full professor / 4 associate professors / 1 assistant professor / 1 adjunct professor & security research scientist / 1 user experience researcher / 1 USEC research engineer / 1 research fellow). The sample included two senior PhD candidates who had published USEC research in top-tier venues and received best paper awards. Their inclusion widened the covered spectrum as they had more recent hands-on experience in implementing prototypes and conducting user studies.

Table 3.1: The interviewees have published a significant number of work ($\bar{x}_{pub} = 123.42$) that is highly cited ($\bar{x}_{cite} = 3740.75$). The data reported is from early 2020.

Anonymised Participants*	Publications	Citations	h-index	Job title	Academia	Industry
P1	[0,50]	[0,100]	[0,5]	PhD candidate	✓	✗
P2	[50,100]	[2.500,5.000]	>30	User Experience Researcher	✗	✓
P3	[100,250]	[2.500,5.000]	>30	Associate Professor	✓	✗
P4	[0,50]	[100,250]	>5	PhD candidate & UX Researcher	✓	✓
P5	[0,50]	[100,250]	>5	USEC Research Engineer	✗	✓
P6	[50,100]	[500,1.000]	>10	Assoc. Prof. & UX Researcher	✓	✓
P7	[0,50]	[0,100]	>5	Research Fellow in USEC	✓	✗
P8	[250,500]	[10.000,20.000]	>50	Full Professor	✓	✓
P9	[100,250]	[2.500,5.000]	>25	Associate Professor	✓	✗
P10	[50,100]	[1.000,2.500]	>15	Associate Professor	✓	✗
P11	[0,50]	[1.000,2.500]	>10	Assistant Professor	✓	✗
P12	[250,500]	[10.000,20.000]	>50	Adj. Prof. & Security Research Scientist	✓	✓
\bar{x}	123.42	3740.75	22.5	-	$\Sigma 10$	$\Sigma 6$

*To protect the experts' identities, intervals for the number of publications, citations, and h-indices are reported.

All interviewees worked in the broader field of USEC, including user authentication, anti-phishing efforts, mobile security and privacy, and web privacy. The interviewed experts had on average 123.42 publications ($max = 386$, $min = 18$, $SD = 129.81$), 3740.75 citations ($max = 14627$, $min = 25$, $SD = 4857.28$) and an h-index of 22.5 ($max = 56$, $min = 3$, $SD = 16.69$). All reported numbers, i.e., publications, citations, and h-indices, entail all kinds of publications, including USEC works. The overall number of publications is reported because all publications eventually contribute to a researcher's h-index; precisely extracting the number of USEC-specific papers is challenging (and arguably infeasible).

The final set of publications ($N = 27$) used to set up context during interviews ranged from 2010 to 2019 ($Md = 2018$). Of the 27 publications used in the interviews, 14 papers comprised software-based prototypes and nine comprised hardware components. Four additional USEC papers from the interviewed experts were used, three of which are considered highly influential in the USEC field. The fourth paper reports on research on an in-the-wild deployed security prototype. One of these additional publications discussed, for example, the last decade of USEC prototypes and outlined learned lessons when developing and evaluating USEC prototypes. Table 3.1 shows an anonymised overview of the participants. To keep the interviewed experts anonymised no additional information about the experts and their scientific works can be provided.

3.2.2 Interview Structure

Semi-structured interviews were informed by the content of the interviewees' publications, which the interviewer familiarised himself with. All publications were drawn from the initial literature review used for the sampling procedure, outlined in Section 3.2.1. The corresponding publications were used as example papers attached to the initial email request.

This allowed the lead researcher to efficiently use the interviewees' time and add context to their opinions. It also facilitated detailed discussions, allowing the interviewees to explore examples and the interviewer to ask informed follow-up questions. The semi-structured interview questionnaire is available in Appendix B. Follow-up questions were prepared and used if needed. All interviews covered the following topics:

- **Typical Research Journey from Idea to Publication:** This question aimed to understand how the interviewee normally progresses from a research idea to publication(s) and how that progression occurs within their broader research community. Journeys typically included topics such as idea generation, resources, prototype development, idea refinement, evaluation, and publication.
- **Research Challenges and Limitations:** Interviewees were asked about their challenges when conducting research that involves prototypes and their opinion about the more general challenges and limitations of USEC research.
- **The Ecological Validity of Current Evaluations:** Insights on different study types employed by USEC experts were collected, allowing the thesis to understand the obstacles to conducting ecologically valid evaluations. Inspired by the literature that argued for developing novel methodologies to understand and design for emerging technologies and mitigate new threats [17, 155], the experts were asked whether they see the USEC's current evaluation approaches as the way to go in the future, or if they would prefer to see a change.

Finally, the interviewees were debriefed and asked if they have any final questions or thoughts. The interview closed with an informal chat. Interviews lasted 48.5 minutes on average. The experts were offered an £8 online shopping voucher for their time. Some of them waived the compensation for different reasons, such as donating it or keeping it for future research.

3.2.3 Research Approach and Data Analysis

Open coding followed by a thematic analysis [57] was applied on the interview data to identify patterns in meaning across the interviews and derive themes. This allows the first chapter to a) build the insights and key challenges directly from the raw data of the expert interviews and b) uncover the main concerns and challenges of USEC experts when prototyping USEC research artefacts. The initial literature review in advance of the interviews allowed the research team to better understand the research area and line up potential interviewees, who were then contacted by email, as previously described in Section 3.2.1. Doing this enabled the

lead researcher to familiarise himself with the experts' works and access a promising USEC sample to contribute answers to the first research question (RQ₁).

For the data collection, semi-structured interviews with open-ended questions were conducted. While interviews were ongoing, two researchers regularly met to discuss the notes taken by the interviewer about interesting observations and thoughts that emerged during the interviews and the publications associated with the upcoming interviews. These meetings allowed the researchers to reflect on the findings regularly and keep those points in mind in further interviews. Once all the interviews were completed, the lead researcher transcribed all audio recordings and open-coded the transcriptions. The initial open coding scope was drawn from the regular discussion meetings. The lead researcher took additional memos [433], i.e., brief notes about the thoughts, ideas, and questions that come to the researcher's mind during data gathering or data analysis. A second researcher went through the raw interview data and added additional memos. This process generated 325 open codes and 93 memos. The lead researcher then organised all codes and printed those out to have a paper-based piece for each code. Two researchers then conducted a paper-based affinity diagram of the open codes. The transcript, memos, and audio was revisited when additional information about a code's context was needed. The researchers organised the codes into groups which were then further refined into themes.

In summary, this chapter will present, for the first time, the USEC researchers and practitioners' experiences and challenges when conducting USEC prototyping research. The themes from the data analysis of the semi-structured interviews, the experts' verbal comments, and the synthesised key challenges when prototyping USEC research artefacts are presented in Section 3.3, along with a discussion of the findings in combination with the previous literature in Section 3.4. Before presenting the results of the interviews, some methodological limitations of this research chapter are discussed in the next section.

3.2.4 Methodological Limitations

Whilst the methodology of conducting expert interviews is common in HCI and USEC (for example, [484, 503]), some specific decisions have limitations to keep in mind. First, experienced researchers who have been successful in publishing works involving USEC prototypes were selected. Their experience is valuable, but it is also biased towards those who ultimately succeeded in publishing. The challenges faced by those who tried and failed to conduct this type of research due to issues such as lack of mentorship, or choosing too challenging problems, are not well represented in this chapter.

Second, the interviews with the USEC experts on a research and community level might not have captured all sides of the conversation. For example, the views of entities such as research institutions and funding agencies are not covered, which this thesis leaves to future work.

Third, participants were likely biased by the publications selected and sent to them before the interview. Pre-selecting publications helped both the interviewer and the interviewees to scope the interviews in a time-productive manner, but the scoping also likely impacted the topics the interviewees chose to discuss. Four of the participants, two pairs, had co-authored papers in the reviewed paper set. Given some participants' seniority and the field's size, such a situation is expected. However, no publication was used in more than one interview session to ensure that the experts' verbal comments do not revolve around the same publications.

Finally, the interviews were retrospective in nature, focusing on the experts' past experiences. Whilst retrospective interviews are effective for learning about rare events or those that take place over a long time period, they also suffer from a bias towards memorable events. The interviewees described projects where the initial idea generation was sometimes years in the past, likely resulting in some issues of memory bias.

3.3 Results

This section presents the key findings of the interviews: 1) threat modelling, 2) prototyping USEC systems, 3) sample size and selection, 4) evaluations, 5) USEC's research culture, and 6) USEC's real-world impact. Participant numbers (P1 – P12) protect the experts' anonymity. Table 3.1 shows some metadata of the interviewed sample. Introductory preambles introduce the topic and set the frame of the challenge before presenting each key challenge.

3.3.1 Threat Modelling is not Straightforward

Threat modelling is commonly used in USEC research to describe an attacker's assumed skills and capabilities. Many input and feedback methods can be observed by bystanders, which led to a lot of emphasis on shoulder surfing. Shoulder surfing is a social engineering technique where a bystander looks over a victim's shoulder to obtain personal information such as PINs, passwords, or other confidential data [56, 125]. Many publications are concerned with shoulder surfing, which is likely further raised given the technological change of mobile devices (for example, VR/AR devices) and their vulnerability to observations [158]. Shoulder surfing evaluations impact the design and evaluation of USEC prototypes because researchers need to consider the threat of observations in the design and development process. Whilst there are many different threats, including, for example, social engineering attacks [268], online/offline guessing attacks [164], smudge attacks [28, 532], and thermal attacks [3], shoulder surfing [61, 95, 157, 307] was frequently brought up by several experts who exhibited a range of opinions about what constitutes a "realistic" threat model.

Many interviewed experts focused on authentication research in the past, which is not surprising given the authentication's dominant role in USEC research [17, 110, 155]. Consequently, the example of shoulder surfing was brought up several times in reference to how threat models, prototypes, and study designs interact with each other. The expert interviews revealed two opposing opinions regarding valid threat models that address security. P5 argued that the relevance of shoulder surfing attacks depends heavily on the context. They explained that the threat has different implications in different countries, implying that cultural differences in perceived personal spaces impact susceptibility to shoulder surfing [411] and that these attacks scare them:

“in the U.S. as well as in Europe you may not really feel [that] shoulder surfing attacks are something that you should really care about [...]. In over-populated countries like India you have a lot of people [...] when you go to an ATM machine or to places like coffee shops [...] there are like three people standing right behind you. [...] shoulder surfing is a really big problem” - P5

P8 emphasised that such attacks happen in the real world and that researchers have to consider the end-users' concerns as well as the USEC experts' views to accurately assess the value and validity of certain threats:

“keep asking the users about what they are worried about; often they are less worried about the NSA and more worried about their parents/their partner” - P8

However, even if a threat model is appropriate for a given context and important to end-users, the experts had different opinions on the value of specific threat models. Some experts mentioned that *“shoulder surfing is a problem, but it's hugely overblown” (P9)* and that shoulder surfing evaluations are not interesting from a security perspective:

“fundamentally for me the problem with observation attacks is, [they] are not that interesting from a security point of view, it's a real niche attack [...] [researchers] report performance against observation attacks with a very narrow threat model: “can you see it”; which is incredibly, it's very very narrow” - P3

P2 emphasised the problem that there is no common agreement among the USEC experts regarding the validity of specific threat models and that *“[researchers] think they use the worst case scenario, but actually they did not” (P2)*. P7 further described the threat modelling challenges using shoulder surfing as an example. They emphasised the mismatch between the researchers' assumptions and the reality and that it is crucial to consider social norms when studying threat models because *“people move closer than [researchers] actually thought they ever would, or they stay further away because they respect people's social norms” (P7)*.

KEY CHALLENGE 1

The experts' opinions regarding the value of specific threat models vary widely. A good threat model needs to match the contextual realities of end-users, but those realities are not always known or may only impact a specific subset of people, making threat modelling a non-trivial part of USEC research.

3.3.2 Prototyping USEC Research Artefacts

Prototyping is an integral part of human-centred research and design [134, 556], one of the main types of research contributions in HCI [556], as well as a major theme in USEC [155]. Themes about the hardware challenges when building USEC prototypes and the deployment and evaluation challenges are reported in Section 3.3.2.1 and Section 3.3.2.2.

3.3.2.1 Development and Hardware Challenges in USEC

The experts voiced that developing USEC prototypes is challenging and costly due to limited access to appropriate hardware and their limited prototyping expertise:

“I think we actually really need more collaborations between the usable security people and the people who are fairly close to building [hardware prototypes].” - P8

P1 voiced that they faced issues with one of their prototypes with an eye tracker due to inappropriate lighting conditions. The interplay between multiple hardware components caused problems and resulted in significantly more effort, additional pilot tests, and excluding data from the actual user study:

“I combined [the hardware] all together [...] and then [faced] issues [...] because they are all working with infrared and [operate] on the same wave length” - P1

“if you recruit 50 participants [...] you have to discard five to ten participants because the eye tracker is not working” - P1

The experts voiced that hardware limitations led to many prototypes *“[that] were made very quickly [and] are not well made” (P3)* or that hardware is used inappropriately, threatening the ecological validity of the research findings:

“we slapped the phone on [a user’s] wrist and put a little active part in a corner, so it was sort of a like big wrist watch but it was not usable [...] the validity of using a phone on users’ wrist is relatively low” - P3

The experts attested that the lack of appropriate hardware, partially due to a lack of funding, is a fundamental problem in their research:

“usually we do not have funding to buy new equipment [...] then we have to come up with ideas of how we can build that hardware” - P4

P2 mentioned that such hardware and funding bottlenecks have a noticeable impact on USEC research. For example, P2 voiced that their prototypes were significantly heavier than traditional mobile devices at the time of their research:

“a lot of negative feedback in those evaluations was around the weight of the prototype [...] [the weight] made it more difficult to use [the prototype]” - P2

Some experts mentioned they had to adjust their research projects due to the lack of appropriate research equipment:

“we try to have as fast as we can the first prototype and see what are the challenges from the development side because often we need to alter the project to fit to the equipment we have” - P4

Others mentioned that setting up different hardware components at their intended place can be challenging and that these physical restrictions often forced them to devise alternative, imperfect solutions:

“I didn’t really manage to put [the front camera] exactly in the middle because the eye tracker was [already] there” - P1

3.3.2.2 Deployment and Evaluation Challenges in USEC

Regarding evaluating USEC prototypes and conducting research beyond evaluations in the lab, the experts explained that they have a hard time in assessing their prototypes. There are a lot of issues around deployability, especially when using new or custom-built hardware. Although USEC experts strive for real-world deployments to increase ecological validity, P4 still sees the transferability of findings to people’s lives as one of the major problems:

“the major problem with evaluating privacy and security systems is that how can you visualize that the users are acting the way they would act if they would [use] it in their everyday life” - P4

P2 further voiced that deploying one of their prototypes to a large sample was impossible and explained the situation of having access to only one device:

“there’s a lot of issues around deployability, specifically when it comes to using new hardware [...] the deployment was impossible [...] we had one device and that device we could hand out to one person at a time” - P2

When using new hardware, the experts highlighted that the cost of failure might be high and that it is crucial to invest only in equipment that is likely to become publicly available and provides promising future use cases. P2 further emphasised the noticeable impact of limited deployable hardware on research and that they could not run a memorability study as part of their evaluation due to the lack of appropriate hardware:

“we did not run a memorability study [for our authentication scheme] mainly due to hardware issues [...] the magical formula would be having an infrastructure that allows to [build hardware-based prototypes] in a very quick way” - P2

In summary, the experts reported that research involving hardware prototypes often introduces additional research hurdles. Besides the hardware challenges, many of the challenges voiced by the experts eventuate from limited access to appropriate resources and the lack of funding, which is further discussed in the context of USEC’s research culture in Section 3.3.5.3.

KEY CHALLENGE 2

The experts emphasised that evaluations of USEC research artefacts are expensive and often infeasible to do in an ecologically valid way, especially when they are large-scale and require special equipment or hardware-prototyping experience.

3.3.3 Sample Size and Selection Process in USEC

Concerns about the user study samples, especially discussions around the appropriate size of a sample and its characteristics, are highly dominant when conducting experiments and are frequently discussed in HCI [66] and USEC [406]. Similarly, the interviewed experts raised concerns about the sample sizes in USEC evaluations (Section 3.3.3.1) and how participants for user studies are recruited within the broader research field (Section 3.3.3.2).

3.3.3.1 Small Sample Sizes in USEC

The experts highlighted the importance of collecting large datasets, especially for security evaluations. P3 described the problem with the pool of real-world passwords that is significantly larger than a small subset of passwords collected from user studies: *“there’s 70 million from a cracked database, you got six and a half thousand – that’s like a drop in the ocean” (P3)*. P3 further voiced that evaluations with small datasets cannot be used to assess security:

“[we] have got 12-20 users [...] the security data is of no value and the conclusion is that there is no value inside the small sample size” - P3

Across all the experts, there was a consensus that the sample size and selection is a fundamental and ongoing challenge that goes beyond USEC. P11 repeatedly emphasised the challenge of

achieving large sample sizes: “*finding a large sample size is really hard*” (P11). Whilst small sample sizes are problematic, as pointed out by P3 and others, external factors such as having access to different research environments have a notable impact on the sampling process. For example, P11 voiced that they face significant issues when recruiting participants and that their resources are limited:

“I recently moved to another country and I was really happy to get 25 [participants] [...] I was really happy to get them but well ...” - P11

Similarly, P1 expressed that the lack of participants is one of their primary research problems and that it is often challenging to convince potential participants to come to the lab, which, according to P1, is not on the university’s campus and would require some additional travel for participation. Throughout the interview, P11 further voiced that relatively small samples are too small for some researchers and immediately invalidate the research. However, according to P11, many researchers overlook the still valuable research and its contributions to the USEC research field despite small user study samples.

3.3.3.2 Biased Participant Recruitment in USEC

Additionally to the experts’ sample size challenges, there were discussions and concerns about the participant recruitment process – how participants are recruited for user studies.

“[sample size/selection] is one of the largest outstanding problems with all HCI systems work which is that we evaluate [our systems] by knocking on the doors of friends and colleagues and be like ‘hey, come do my user study and I’ll give you \$10’” - P11

P7 echoed the problem of evaluations using the experimenters’ social circles and that it is often unclear what happens if research artefacts are evaluated with a more diverse sample and with people unfamiliar with the technology:

“we run [studies] within our social circles, what happens if we get someone who’s elderly, who’s not familiar with technology [...]” - P7

P2 further highlighted that although they have access to a gigantic user pool, which is not comparable to the often limited user pools in academic environments, their user pool still runs out. According to P2, their company still relies on vendors to access an even more extensive set of participants for user testing.

KEY CHALLENGE 3

The experts voiced that the sample size and the recruitment of participants for studies are problems across multiple disciplines and major concerns in USEC. Small sample sizes and biased participant selection reduce the value and validity of USEC research.

3.3.4 Evaluation Methodologies in USEC

Building upon the various research methods in USEC as reviewed in the Literature Review, the experts commonly discussed lab studies and field studies and held strong opinions. There were themes around the importance of both of them (Section 3.3.4.1), the value/cost trade-off (Section 3.3.4.2), and the perceived value of field studies in USEC (Section 3.3.4.3).

3.3.4.1 The Importance of Lab Studies and Field Studies

The experts emphasised the necessity of different evaluation approaches and that starting with lab studies is often a prerequisite for evaluating USEC prototypes:

“there’s a place for both [...] I don’t think it makes any sense to go directly into the field to evaluate new systems when we haven’t done any lab studies at all” - P9

The experts agreed that lab studies must be conducted before going into the field and that the value of lab studies should not be underestimated. However, there was an agreement across all the experts that field studies have the potential to lead to high ecological validity. P11 voiced that different study types have different pros and cons and that imperfect evaluations of USEC prototypes can still be valid contributions to the research community:

“we can have ideas - that’s the strength of academia – ideas that are totally radical and new and not going to be evaluated perfectly in the context of a lab study [...] but that doesn’t mean they don’t have value [or] can’t inspire the direction of the usable security and privacy future” - P11

P7 further emphasised the importance of taking prototypes out of the lab and placing them into realistic environments and scenarios. Other experts mentioned that real-world investigations are of particular significance as participants manifest “demand characteristics” [354]: they subconsciously change their behaviour to fit the experimenter’s purpose. For example, P11 highlighted the uncertainty of the effects of lab studies on results and that they “cannot be sure whether [participants] are acting as [they] would act in the wild or if they’ve changed their behaviour because they know they are being part of a study” (P11).

3.3.4.2 The Trade-off Between Added Value and Costs in USEC

The trade-off between effort in applying a methodology, for example, a lab study vs a field study, and the value and the ecological validity of the corresponding findings was highly discussed by the experts and is considered to be a domain challenge in USEC research [155]. The experts stated that running lab studies is considered simpler than running studies in the

field and that this is one of the main reasons why the USEC research community conducts a plethora of studies in laboratory environments and significantly less studies in the wild:

“[the] uncharitable view would be that [running lab studies rather than field studies] is just easier to do” - P8

“my take is that it’s a mix of convenience, not knowing better, and impossibility as in certain [situations] you can’t do [experiments] that are difficult to do that it’s not worth the additional effort” - P2

P9 voiced that *“it’s easier to do a lab study; the odds of something going wrong are way too high” (P9)*. P3 added that before conducting field studies it is important to compare the value versus the effort of going out to the field:

“there is a place for [field studies] but is there enough added value in field studies generally that this is important?” - P3

Overall, the experts voiced that field studies can be powerful and it is not unlikely that the corresponding results deviate from research in laboratory environments. However, they argued that researchers need to be clear about what they are seeking rather than being exploratory when conducting field studies. For example, P3 elaborated that *“field studies can be valuable but there needs to be a clear value [...] the data will differ from a lab” (P3)*. P4 highlighted the strength of field studies as they provide insights into how the research artefacts are really going to be used and how people are going to accept them. However, P5 emphasised that it is hard to pinpoint causes of effects in field studies and that achieving accurate results through field studies only is challenging.

KEY CHALLENGE 4

The experts mentioned that the choice of evaluation methodology is highly context-dependent and that it is important to have a clear vision and expectation of the evaluation scale. There is a clear value in field studies, but there is a need for preceding lab studies as pinpointing the sources of problems using field studies is challenging.

3.3.4.3 The USEC Experts’ Views on Field Studies

Some experts reported believing that *“field studies are sort of a gold standard” (P11)* and that they *“would like to see more about how security fits into real life as opposed to specific little corner cases that are easy to run” (P9)*. P10 highlighted that the suitability of field studies heavily depends on the required investigation and the legal/ethical considerations and that this differs a lot between different countries. The experts described some unsuccessful attempts when studying research prototypes in real-world settings:

“we looked at investigating [our security system] within a real setting but there were just too many legal and ethical constraints around that” - P2

P6 and P7 added that field studies are expensive and that they often rely on findings from lab studies only due to budget constraints and technological issues.

KEY CHALLENGE 5

The experts voiced that legal, ethical, and budget constraints play a significant role in deciding whether to conduct field studies in USEC.

3.3.5 USEC’s Research Culture

Different research fields and individual researchers manifest different behaviours, values, expectations, attitudes, and norms, forming a unique research environment and culture. Open science and reproducibility, for example, are recognised as vital features of science across research fields and considered as a disciplinary norm and value [321]. However, in practice, there are significant differences across research communities. Wacharamanatham et al. [533] showed that the process of sharing artefacts is an uncommon practice in the HCI community. Cockburn et al. [78] showed that preregistration of scientific research has received little to no published attention in HCI. When it comes to USEC and the researchers’ behaviours, values, expectations, attitudes, and norms, the experts mentioned challenges about the expected, often hard to reach, high ecological validity of the USEC prototype evaluations (Section 3.3.5.1), the USEC researchers’ reserved enthusiasm about novel, well-evaluated prototypes (Section 3.3.5.2), and the lack of access to research resources (Section 3.3.5.3).

3.3.5.1 Towards (high) Ecological Validity in USEC

An important objective in USEC is to achieve high ecological validity – the extent to which a study adequately reflects real-world conditions. A password study by Fahl et al. [133] showed that participants in lab studies behave differently than how they act in the real world. When used with care, many insights from self-report security data can translate to the real world. However, self-reported data can vary from data collected in the field, and alternative research methodologies should be considered for studying complex constructs [409].

Some interviewed experts mentioned that USEC researchers often expect high ecological validity and generalisability of the study findings, which matches the research approaches in the literature. For example, to increase the sample sizes, target more representative samples, and achieve the expected high ecological validity, USEC researchers often aim to role-play real-world situations in the lab [133], conduct field studies [189, 294, 296], or utilise online studies [74, 190, 297]. However, P12 stated that a real-world evaluation of all systems’ usability

and security aspects is almost impossible: *“the difficulty in evaluating system security is that the lack of security can have many different sources.”* (P12). P12 further emphasised the complexity of security evaluations:

“All secure systems are alike. But there are many different ways for a system to be not secure. It is not possible to enumerate them all.” - P12

A concern by P1 was about the lack of common evaluation approaches and that their evaluation metrics, such as interaction time and input accuracy, often have to evolve from literature reviews because of the lack of any standards. The researchers' various evaluation approaches exacerbate the problem of determining which metrics to investigate and which evaluation method to employ for assessing USEC prototypes. P2 voiced that the variety of evaluation approaches often leads to various prototype evaluations and conclusions:

“if you look at five different usable security papers you can't compare them because they have used slightly different approaches of evaluating the different parts of their systems [...] you can't really say which one was better or worse” - P2

P2 particularly emphasised the subjectivity of privacy and security and that many researchers have strong opinions regarding the evaluation of USEC research artefacts. The lack of standardised sets and metrics to evaluate security schemes makes it even harder to address the ecological validity and perform comparisons between multiple works:

“I am not aware of any standard scenarios that can say ‘okay, here now we can compare it if we're running a lab study’” - P2

In line with Key Challenge 4, P11 underlined the need for a clear vision of what is expected of evaluations that are either conducted in laboratory settings and are likely less ecologically valid, or are conducted as organised field studies that are still limited to an extent due to research participation effects [354,372]:

“we just need to be a little bit more open to what sort of solutions/evaluations we are expecting out of [something] that has not actually been deployed in the real world” - P11

KEY CHALLENGE 6

The experts emphasised that aiming for evaluations with high ecological validity is crucial in USEC. However, they also voiced that current USEC prototype evaluations are often incapable of achieving the expected high ecological validity.

3.3.5.2 Creating Space for Novel USEC Prototypes

P11 expressed that whilst identifying usability and security issues in existing systems is important (i.e., problem-scoping research), it is equally important to conduct problem-solving research where USEC prototypes are designed, implemented, and scientifically assessed to contribute towards usable, secure, and privacy-preserving systems and experiences. Some experts raised the concern that the USEC community is very focused on the evaluation part when a significant contributing element is building a prototype functional enough to demonstrate effectiveness in terms of deployment and usability:

“the [USEC community] wants to evaluate everything when like a big part of your contribution is just the fact that you could build this [system]” - P11

P11 further emphasised that without recognising the value of functional solutions/prototypes, which may come with limitations imposed by the real world, the USEC community might struggle to engage with the realities of solving problems:

“I feel like we as a community refuse to accept that kind of contribution – then you know, we’re shooting ourselves in the foot, we’re never going to be part of the broader conversation.” - P11

Other experts criticised the community’s focus on realistic use cases, which results in limited enthusiasm for building speculative future-oriented USEC prototypes and experiences. P8 mentioned that they had seen some shifts recently, but problem-scoping and problem-solving USEC research are still not balanced:

“I like some of the shift we’ve seen recently [...] to actually really look at finding ways of supporting [users]” - P8

Considering the implementation of novel USEC prototypes, P11 argued that there is still a lack of future-oriented USEC research where use cases are more speculative or avant-garde:

“in general the usable security and privacy security community is not very imaginative [...] they don’t really like thinking too far in the future” - P11

P10 agreed to some extent and voiced that the USEC community does not appreciate research where they have to imagine worlds that do not exist.

KEY CHALLENGE 7

The experts emphasised that problem-solving USEC research is relatively scarce. Whilst problem-scoping USEC research lays the foundation for further investigations, research that designs, implements, and evaluates USEC prototype solutions is needed.

3.3.5.3 The Availability of Resources in USEC

The experts highlighted the lack of open-source material within the USEC community that negatively affects their research. According to P6, there is a significant lack of open-source implementations of USEC prototypes. P4 voiced that the lack of resources makes it time-consuming and challenging to build specific features. P11 suggested collectively building a platform that supports researchers in their research:

“how can we create a platform that will make it super easy for other researchers to build upon the foundation that you’ve created?” - P11

The experts voiced that the available hardware often defines their research. For example, P2 faced challenges when aiming to evaluate a USEC prototype’s usability while users walked:

“we had the idea of putting people on a treadmill for the evaluation [...] but then didn’t have a treadmill” - P2

Building upon the sample size discussions in Section 3.3.3, the experts asserted that finding a broad user base is even more critical in academic research settings and that this is where most academic studies suffer because the resources for recruiting are limited.

KEY CHALLENGE 8

The experts voiced that the current USEC research community does not consistently support sharing research resources, such as access to hardware prototypes, software implementations, and platforms for conducting user studies.

3.3.6 Academia and Industry in USEC

The experts voiced that having access to security systems used by companies for research purposes is challenging. The lack of access often results in lower ecological validity as well as barriers to transitioning research results into practice. For example, one issue in USEC research is that potential industry partners are concerned about harmful findings and do not allow any “vulnerability research” [151], including prototype-building work. P2 related such an incident in their research:

“we did have some connections with [companies] but they are like: ‘you can’t touch our machines’” - P2

The experts voiced that this type of research can be of great value, but that there are concerns over the legal challenges. Building upon the discussions around USEC’s research culture, Section 3.3.6.1 presents the experts’ comments on the lack of collaborations between academia and industry. Section 3.3.6.2 reports on the limited real-world impact of USEC research.

3.3.6.1 Status Quo of Academia and Industry in USEC

The experts mentioned that although there are collaborations between academia and industry, there is still room for improvement when it comes to exchanging knowledge, sharing research resources, and accelerating impact. One of the resulting problems is the lack of hardware accessibility, similar to Key Challenge 8, which leads to limited research contributions and decreases the ecological validity of USEC research:

“if they just lent us [an ATM] for a period of time it would have been really good to do our studies” - P8

P7 brought up the challenges with financial institutions in usability and security research:

“which bank would allow [to install] some random prototypical hardware; probably no bank” - P7

According to P8, the lack of access to hardware in highly realistic settings is only one problem. They voiced that some companies' fears of security leaks significantly impacts USEC research:

“if you are working in the security in an environment where there is real-world security, they often won't let you do any observations and I think that's really bad [...] they are afraid that you're going to find something that means the security isn't working” - P8

Besides the companies' fears of security leaks, one of the interviewed experts voiced that many researchers are restricted from publishing findings based on observations within companies:

“[I have] been lucky to have done observational studies a couple of times [but I was not] allowed to publish them” - P8

3.3.6.2 USEC Research and its “Real-World” Impact

When asking the experts whether they see controlled lab studies as the “way to go” to evaluate USEC research artefacts and what progress they would like to see within the USEC community, discussions around the impact of USEC research on real-world applications came up and that this transition, moving USEC research into practice, is still lacking. Some experts voiced that the problem is not that the USEC research community lacks ideas for usable and secure artefacts, but that they would like to see how solutions fit into real life. For example, P9 voiced that many publications end in a heap of USEC prototypes that never find their way into people's daily lives:

“there's a lot of proposed authentication schemes out there and a lot of them aren't gonna move forward like a lot of them are ideas, they didn't really work

out, they're not really showing any promise and so you know, we discarded them”
- P9

Although some usable and secure technologies are widely deployed nowadays, for example, anti-phishing technologies or two-factor authentication, a large part of USEC research has not found adoption in the real world. Examples include enhancing authentication on mobile devices [41, 98, 243, 530] or protecting the users' privacy when interacting with public displays [103, 402, 530]. The experts voiced that a major reason for the limited impact is the huge gap between prototype evaluations and being able to use these prototypes in the real world: *“there's a huge gap between possibility and building the system and commercialisation”* (P5). Complementing the lack of real-world impact, P8 emphasised that many researchers lack interest of changing their existing theories or skillset, pointing towards some resistance to change within the USEC community. The experts highlighted that the interests of the USEC researchers vary widely and that they are concerned about some other researchers' mindsets:

“I've seen this in rebuttals [...] when I write a review about something [...] and they are like oh well so many other people have published lab studies, why should I have to go out and do something differently [...] it's a lot harder, it's a lot more work and as long as I can get this stuff published why should I bother?” - P8

The experts highlighted that USEC research should go beyond publications and not be entirely driven by the “publish or perish” mindset [315]. P11 encourages the USEC community to think big and collaboratively aim for more than “little projects”:

“How can we make that little project the next like D3.js¹ for usable security?” - P11

P12 further criticised the opinionated mindset of many researchers and that someone's academic career is often considered more important than having real-world impact:

“most researchers' goal is to produce papers and get their degree or tenure; few researchers are [actually] building and deploying working systems” - P12

KEY CHALLENGE 9

The experts voiced that there is a lack of strong collaborations between academia and industry. They further argued that some resistance to changes exists within the USEC community, resulting in limited real-world impact.

¹Bostock et al. [54] presented data-driven documents (D3) as a novel approach for visualisations at IEEE Transactions on Visualization and Computer Graphics in 2011. Originating from research conducted at Stanford University, D3.js found its way into web development and is nowadays a library for data-driven visualisations.

3.4 Discussion

This thesis has identified nine key challenges that restrict and often guide USEC research involving prototypes. Each key challenge contributes answers to RQ₁: the research challenges in USEC are manifold and it is hard to pinpoint a single source (Key Challenge 1 – 9). To better support the USEC community and facilitate the process of conducting usability and security research on prototypes, this thesis chapter first discusses similar challenges in neighbouring HCI disciplines and then puts forward the use of VR studies as a novel research method in chapter 4, chapter 5, and chapter 6.

The following sections discuss the implications of the interviews and provide ways forward for both individual researchers and the USEC community as a whole.

3.4.1 There is No One Best Way for Doing USEC Research

The experts noted that it is impossible to enumerate all system security aspects, but imperfectly prototyped and evaluated USEC research artefacts can still have value and inspire the direction of the USEC's future. The opinions brought up by the experts regarding the design, development, and evaluation of prototypes are not far away from the HCI literature. Greenberg and Buxton [173] and Shneiderman et al. [462] emphasised that the choice of the evaluation methodology should evolve from the actual problem (for example, what are the people's needs?) and from appropriate research questions. In USEC, it is essential to note that a prototype's usability and security often highly depends on the specific context as external factors have been shown to impact a system's state and a user's behaviour [231].

The value, benefits, and drawbacks of different evaluation methods were echoed by the experts together with the non-trivial part of threat modelling (Key Challenge 1 and 4). In the next section, the experts' comments are discussed and tied back to the broader research field.

3.4.1.1 Adjusting Expectations of Prototype Implementations and Evaluations

According to some of the experts, for example, P12's statement in Section 3.3.5.1, or P11's statements in Section 3.3.5.1 and Section 3.3.5.2, the USEC's challenges are exacerbated by some researchers' expectations of exhaustive evaluations that assess every single aspect of a prototype's characteristics in an ecologically valid setting. Many reasons make this often infeasible when evaluating prototypes, including 1) the need to run lab studies first to evaluate the new elements in the prototype and pinpoint causes of problems, 2) not having the hardware resources available to produce multiple prototypes for in-the-wild testing, and 3) being restricted in field research due to ethical and legal constraints that are beyond individual researchers' capabilities (for example, [100,529] and Key Challenge 5).

The hardware prototyping and ecological validity challenges voiced by the experts (Key Challenge 2 and 6) can be found in neighbouring research communities, such as Ubicomp. Prototyping novel ubiquitous systems is challenging [111, 174] and often requires additional expertise and specific tools (for example, knowledge about different electronic components and access to soldering irons). Greenberg and Fitchett [174] described developing and combining physical devices and interfacing them within the application software as one of the biggest obstacles. In a similar vein to the lack of sharing research resources and expertise in building hardware (Key Challenge 2 and 8), they observed that researchers who develop systems based on physical devices are often required to start from scratch and face many difficulties [174]. In their project, building a reactive media space environment, one of their colleagues (an electrical engineer) joined the team and supported the hardware-building process [174]. More than ten years later, there are similar interdisciplinary collaborations in USEC research. One example that highlights the interdisciplinary nature of USEC is the Back-of-Device prototype by De Luca et al. [102] and their follow-up work, XSide [98]. Their research highlights the impact of a USEC prototype's form factor on the usability and the evaluation results' generalisability. Whilst Greenberg and Fitchett [174] benefitted greatly from an electrical engineer, the authentication prototype by De Luca et al. [98] greatly benefitted from the 3D printing expertise of one of their colleagues.

The message here is that collaborations and novel technologies greatly facilitate and improve USEC prototypes and corresponding evaluations. The USEC experts' challenges, for example, Key Challenge 2 and Key Challenge 6, suggest that expectations of prototype implementations and evaluations must be adjusted in situations where building "perfect" prototypes and conducting highly realistic evaluations is too challenging or even infeasible.

3.4.1.2 Bridging the Gap Between Lab and Field Studies

The USEC community has been debating the respective value of lab studies and field studies for some time, with the interviewed experts similarly mentioning the need to be open to alternative evaluation approaches. Discussions around lab studies and field studies, especially when and how field studies are "*worth the hassle*", are discussed in neighbouring communities as well, as discussed in Section 2.2.2 in the Literature Review. A corresponding critical evaluation and comparison of a lab and field study impacted the Mobile HCI research field as a whole in the subsequent years [253, 254]. Kjeldskov et al. [254] discovered more usability issues in the lab than in a similar field study for roughly half the cost. Consequently, they concluded that the added value of field studies is very little and neglectable, which resulted in a heated debate as their evaluation did not cover long-term use and adoption [215].

In USEC research, the long-term use and evaluation of prototypes is a critical component. Previous works have shown that habituation can impact the users' perceptions and security

behaviours [482, 496]. Other works emphasised the importance of habituation and its key role in USEC research in the classification of genuine login attempts [502], in research on authentication prototypes [99], and in research on security alert dialogues [312]. Greenberg and Buxton [173] argued that there is a need to recognise many other appropriate ways to evaluate and validate work and that usability evaluations can be ineffective if naively done “by rule” rather than “by thought” and that “a combination of methods – from empirical to non-empirical to reflective – will likely help triangulate and enrich the discussion of a system’s validity.” [173].

There are several researchers who have outlined the need to fundamentally rethink current study paradigms [17] and frameworks for understanding privacy risks and solutions in personalisation-based systems [515]. For example, the uptake of smart speakers or VR/AR headsets that collect sensitive user (and bystander) data requires a change in the USEC prototype designs and evaluations [17]. There have been suggestions to improve the ecological validity of usability and security evaluations in the lab, as discussed in Section 2.3.1 in the Literature Review. For example, role-playing real-world situations [133] to mimic scenarios where security is a secondary task and to better reflect real-world scenarios [435]. However, simulating real-world scenarios in the lab cannot compete with the ecological validity of more realistic field studies. A potential direction to address the challenges around ecological validity is to leverage novel technologies for prototype development, deployment, and evaluation. As one of the interviewed experts brought up, 3D printing can facilitate the prototyping process of USEC research artefacts. There are already published works that highlight the strengths of 3D printing for USEC research, for example, the work by De Luca et al. [98] and Marky et al. [299], which can inspire and inform future USEC research.

Similar to the use of 3D printing technology to support human-centred research and contribute towards the implementation of research prototypes, there has been a movement in using virtual and augmented reality to conduct user-centred evaluations of IoT devices and public displays [291, 528, 539], as discussed in Section 2.2.5 in the Literature Review. Following P11’s emphasis on aiming for something beyond little projects, contributing a research method capable of location-independent evaluations of USEC research artefacts could be a powerful addition to the more conventional lab studies and field studies.

The key message of the interviews here is that the USEC research community must be mindful of the researchers’ challenges when evaluating USEC prototypes. A potential solution to the existing challenges between research in the lab and in the field could be the use (and the community’s acceptance) of novel technologies and well-evaluated research methods for USEC prototyping research, including the use of 3D printing [98, 299] and VR/AR to simulate and evaluate real-world research artefacts – similar to the broader HCI field where the use of VR/AR for simulating real-world environments and research artefacts has already found application (for example, [291, 528, 539]).

3.4.2 Selecting Sample Sizes in the Presence of Constraints

A major discussion point in the interviews was about the sample sizes and the participant recruitment process (Key Challenge 3). The experts' concerns about the sample sizes are not surprising – sample sizes and selections form a major domain challenge in USEC research [155,406]. Looking at the content of neighbouring HCI communities, there is a wide range of sample sizes and compositions used. Caine highlighted that twelve participants is the median sample size across papers published in CHI 2014 [66]. Focusing on usability only, Turner et al. [517] found that five users allow discovering 80% of a system's usability problems. In line with the experts' concerns on how the participant recruitment happens in USEC research (for example, “*we evaluate [our systems] by knocking on the doors of friends and colleagues and be like ‘hey, come do my user study’*” (P11)), Lazar et al. [276] argued that there are many HCI studies that manifest small and non-diverse samples; therefore, often do not allow generalising results. The following sections discuss the realities of the sample size selections and the participant recruitment in USEC in more detail.

3.4.2.1 The “Realities” of Sample Size Selections

The interviewed experts voiced that many different opinions exist within the community for USEC evaluations and corresponding sample sizes and selections. For example, P3 argued that a sample size of 12-20 users is too small to have any value. There are indeed published works that come with large sample sizes. For example, Ur et al. [519] conducted an online study with 4509 participants to detail the security and usability impact of a password meter's design dimensions. Cheon et al. [74] assessed and evaluated a security framework in large crowd-sourced online studies (N = 2619 and N = 4000). Markert et al. [297] conducted an online study to analyse the security of smartphone unlock PINs with 1220 participants. However, samples of that scale are rare and often challenging to achieve, especially when the research is conducted in the lab. There is a plethora of published works at top USEC venues such as ACM CHI (for example, [90, 98, 102, 239, 248, 530]) or USENIX SOUPS (for example, [95, 269, 506]) that studied noticeable smaller samples. Some security evaluations are even based on a single expert attacker (for example, [43, 95, 102, 269]) or a small sample of trained participants (for example, [6, 44, 245]). The reviewed works, along with Table 2.1 in the Literature Review, highlight the wide range of acceptable participant numbers within USEC and how much that acceptance varies across subdomains, resulting in no single rule about how many participants and what type of participants are required for USEC research.

The key message of the interviews here is that working collectively towards a research standard or a set of roughly defined guidelines could be beneficial for both individual researchers and the USEC research community as a whole. This could help the USEC community to

a) support early career researchers in their USEC research decisions when defining, for example, which research method? which sampling method? how many participants?, and b) facilitate comparisons between works. It is important to stress that the question should not be how many user study participants are required, but to ask how many participants *and* how participants can (and should) be recruited respecting affordability, infrastructure, time, and research question as the availability of the resources can vary a lot, as voiced by the experts.

3.4.2.2 The Quest to Find (many) Participants in USEC

As shown in several USEC works, for example, by Markert et al. [297], Cheon et al. [74], and Felt et al. [136], online crowdsourcing platforms or university-industry collaborations facilitate achieving large sample sizes and allow the investigation of USEC research artefacts at large scales. Yet, the deployment and corresponding evaluations of USEC prototypes remains a challenge. Online studies are often not suitable to, for example, evaluate hardware-based prototypes or research artefacts for platforms that participants do not own. For instance, conducting research on phones with a touch-sensitive rear [102], on smart glasses [553], or on AR glasses [150] often requires participants to physically attend user studies. The USEC prototypes are often not robust enough for in-the-wild deployments, or the required hardware and infrastructure are not yet available for detailed user testing (for example, when investigating novel AR authentication methods [150, 282]).

In line with the suggestions in Section 3.4.1.2, the key message of the interviews here is to investigate alternative study platforms and research methods for conducting USEC research on prototypes that balance a) delivering realistic experiences when interacting with USEC prototypes with b) reaching out to a large and diverse participant sample.

3.4.3 Problem-Scoping and Problem-Solving USEC Research

In the following, this chapter refers to Figure 3.1 to make the results more tangible and highlight the USEC's current bottlenecks. The figure is based on the experts' statements and the interpretation of the conducted interviews (Key Challenge 1 to 9). It distinguishes between: ❶ the real world, ❷ problem-scoping research, and ❸ problem-solving research. USEC research puts a strong emphasis on problem-scoping research where user behaviour is observed to identify usability, privacy, and security issues (for example, [33, 100, 189, 191, 216, 279, 296, 297, 351, 405]). The generated knowledge is then used to inform, teach, and protect people (for example, [18, 68, 250]). However, some of the experts communicated that there is relatively less progress in leveraging these findings to develop novel, well-evaluated USEC prototypes and facilitate their transition into practice. Balancing problem-scoping with problem-solving USEC research has the potential to result in noticeable real-world

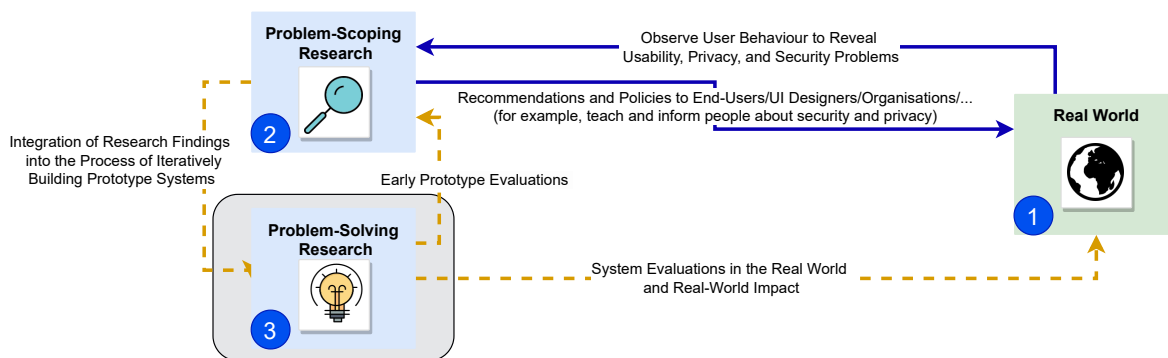


Figure 3.1: The schematic figure represents a substantial part of conducted work in USEC with a focus on USEC research artefacts. Dotted orange lines indicate underdeveloped links and solid blue lines strong links.

impact. P11 emphasised the importance of investing in problem-solving USEC research because otherwise, as they put it: *“we’re shooting ourselves in the foot and never going to be part of the broader conversation”* (P11) (Key Challenge 7). The message of the interviews here is that it is crucial to collectively understand what end-users, the people who are going to use the systems, need before building a large number of different research prototypes that end in publications but do not contribute to the bigger picture: transition research into practice and provide end-users with usable, secure, and privacy-preserving systems. Whilst spotting privacy and security issues is essential, it is equally important to integrate the findings from user studies into an iterative research process and build solutions that solve some of the existing security and privacy concerns. As voiced by the experts, contributing towards problem-scoping and problem-solving USEC research can foster collaborations between researchers and research groups, potentially contributing to closing the gap and strengthening the research loop depicted in Figure 3.1.

3.5 Chapter Conclusion

Although some of the reported key challenges are more relevant to USEC, many issues the USEC experts face when designing, prototyping, and evaluating research artefacts can be found in the neighbouring HCI communities. The history of USEC research, including how the community has been established, the experts’ voiced challenges, and the Literature Review in chapter 2, shows that USEC does not exist in a vacuum. USEC has adopted many research methods from the neighbouring HCI communities [155]. Still, as the expert interviews have shown, the inherently interdisciplinary nature of USEC and the challenges around security evaluations, threat modelling, ecological validity, and the lack of solid links between problem-scoping and problem-solving research make USEC unique, complex, and often impede the transition of USEC research artefacts into practice.

By synthesising, for the first time, opinions from USEC experts that have not seen in-depth discussions in prior literature, and raising awareness of the challenges when prototyping and evaluating USEC prototypes, this thesis provides common ground for ongoing discussions within the USEC community and outlines nine key challenges, some of which this thesis aims to address by introducing and validating the suitability of using VR studies for simulating real-world USEC research involving prototypes. To conclude this chapter, the first research question of this thesis, RQ₁, is answered. Then, the next steps of this thesis are outlined.

3.5.1 Research Question 1 (RQ₁) and Ways Forward

This chapter contributes answers to the first research question:

RQ₁: What are the challenges that USEC experts experience when designing, implementing, and evaluating security and privacy-enhancing prototypes?

USEC research is often concerned with usability and security evaluations of prototypes. For prototypes that involve hardware elements (for example, eye trackers [243] or custom-built handheld devices [102]), the interviews have shown that researchers face significant challenges in building the prototype, recruiting participants for the evaluation, and conducting realistic evaluations that allow generalising the results. The findings of this chapter are summarised in nine key challenges, which are discussed in Section 3.3 and Section 3.4 and cover the broader USEC field and how USEC research is currently being conducted. The discussed key challenges inform the remaining research of this thesis in chapter 4 - chapter 6 as follows:

In line with Alt and von Zezschwitz [17], this thesis has recognised from the expert interviews that there is a need to support and facilitate USEC research that involves designing, building, and evaluating novel research artefacts. The interviewed USEC experts emphasised that their resources to build, evaluate, and adequately distribute USEC prototypes are often limited (Key Challenge 2 and 8). They raised concerns about achieving high ecological validity, recruiting large and diverse samples, and conducting research in the field (Key Challenge 3 and 6). Furthermore, the experts brought up the trade-off between added value and effort when conducting field studies. They voiced that field studies are often resource-intensive and introduce ethical and legal constraints (Key Challenge 4 and 5).

To contribute answers and work towards potential solutions to these challenges and close the existing gap between problem-scoping, problem-solving, and real-world impact (Key Challenge 7), this thesis proposes and investigates the use of VR studies for the implementation and evaluation of USEC prototypes. As prior research in other domains has shown (cf., Section 2.2.5 in the Literature Review), VR studies facilitate field studies and can provide a methodological bridge over the controlled lab studies and the more exploratory field studies.

Despite the first evidence of the use of VR studies for simulating real-world research and its success in other domains (for example, in HCI [291, 528]), it remains unclear how this relatively novel research methodology can (and should) be applied for USEC research and can forge forward the USEC research field as a whole. This thesis aims to fill this gap in chapter 4 and chapter 5 through detailed replications and comparisons of the findings of VR studies and equivalent real-world studies. Additionally, it showcases in chapter 6 a potential future of USEC prototyping research through the help of remote VR studies for simulating real-world research artefacts and study environments to allow for future large-scale and cross-country user studies. Whilst the threat modelling concerns (Key Challenge 1) and the lack of collaborations between academia and industry are important to address (Key Challenge 9), contributing towards solutions requires additional input from academic institutions and industry partners, which is beyond the scope of this thesis.

3.5.2 Contributions

In summary, the research in this chapter makes the following contributions:

- It synthesises, for the first time, the challenges faced by USEC experts when designing, developing, distributing, and evaluating USEC prototypes. Whilst other works have touched on what makes USEC research challenging in general, there has not yet been a structured attempt to elicit challenges experienced by researchers who deal with USEC prototypes. This chapter puts forward such a compilation of experienced challenges, contributing an *empirical contribution* [556] that provides new knowledge through findings based on expert interviews.
- It lays out nine key USEC challenges that resulted from twelve USEC expert interviews, discusses them in the scope of neighbouring communities, and proposes ways forward that inform the remaining thesis research in chapter 4, chapter 5, and chapter 6.

IV

VALIDATING THE USE OF VR STUDIES FOR
EVALUATING USEC PROTOTYPES

Chapter 4

Validating the Use of VR Studies for Evaluating USEC Prototypes

This chapter is based on the following two publications:

[Publication 4] **Mathis, F., Vaniea, K., & Khamis, M. (2021).** Replicueauth: Validating the use of a lab-based virtual reality setup for evaluating authentication systems. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI 2021). ACM, DOI: [10.1145/3411764.3445478](https://doi.org/10.1145/3411764.3445478)

[Publication 5] **Mathis, F., Vaniea, K., & Khamis, M. (2021).** Observing Virtual Avatars: The Impact of Avatars' Fidelity on Identifying Interactions. In Academic Mindtrek 2021 (Mindtrek 2021). ACM, DOI: [10.1145/3464327.3464329](https://doi.org/10.1145/3464327.3464329)

4.1 Introduction

As a first step, and in advance of contributing towards potential solutions to the key challenges presented in chapter 4, research is needed to provide the first evidence of the use of VR studies for usability and security evaluations of real-world USEC prototypes. If VR studies, in which virtual replicas of real-world prototypes are tested, achieve similar results as their counterpart evaluations in physical laboratory environments, then they can be applied in follow-up studies to advance USEC research involving prototypes. The Literature Review in Section 2.2.5 outlined promising HCI applications for VR studies, such as for studying user behaviour in front of public displays in the field [291] or investigating simulated real-world research artefacts [528, 539]. In USEC, such VR studies can a) reduce the costs of studies evaluating real-world USEC prototypes as researchers do not need to build them physically; b) allow for the recruitment of many and diverse participants through a remote

VR research approach, thereby increasing the ecological validity of the research findings; and c) reduce the need for face-to-face studies, which is advantageous at times where in-person studies or field studies are challenging or even prohibited due to, for example, COVID-19 restrictions [149] or not having access to security-sensitive systems such as ATMs.

However, the USEC community is far away from being able to apply VR studies for USEC research: initial validation work is required to provide the first evidence of the use of such a novel research method for usability and security evaluations of real-world prototypes. Therefore, for the first time, this thesis chapter investigates how VR studies can complement usability and security evaluations of real-world USEC prototypes. Can researchers learn about the user's performance on a real-world prototype by measuring their performance on a VR replica? Does a USEC prototype's vulnerability to observations when used in VR map to similar weaknesses if used in the real world? Consequently, this chapter asks and contributes answers to the following two research questions:

RQ₂ Which findings of VR-based usability evaluations on USEC prototypes match the findings from corresponding evaluations in traditional physical lab settings?

RQ₃ Which findings of VR-based security evaluations on USEC prototypes match the findings from corresponding evaluations in traditional physical lab settings?

Determining which results from VR usability and security studies match those obtained from the real world is valuable in multiple ways. If USEC researchers can quickly iterate and evaluate their USEC prototypes in VR before deploying them in the real world, they save costs, time, and effort. This chapter describes the evaluation and presents the results of an alternative methods replication study of CueAuth [245], a previously published authentication prototype (cf., Figure 4.1), to investigate if a real-world study of a USEC prototype can be

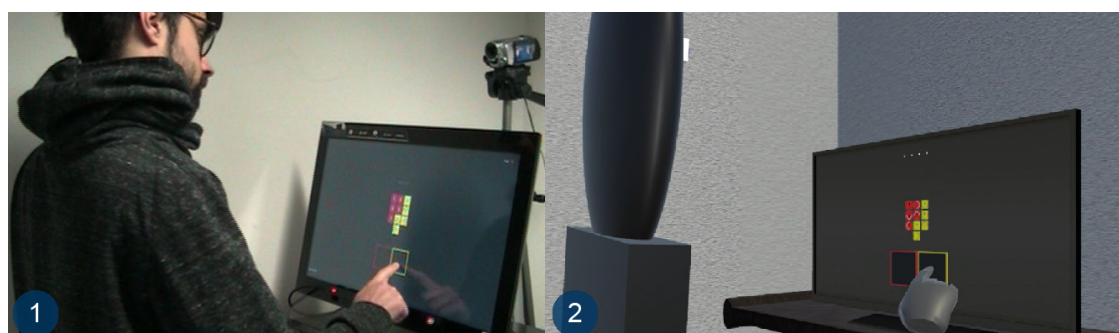


Figure 4.1: For the first time, this chapter evaluates the suitability of using VR studies for human-centred usability and security evaluations of real-world USEC prototypes. The work in this chapter replicated a recently introduced authentication scheme called CueAuth [245] (①) in VR (②). It then evaluated the usability and the security of the VR replica to compare the results to the original real-world evaluation [245]. The left part of the figure is a copy of a figure in the original CueAuth paper by Khamis et al. [245].

similarly run in VR. CueAuth [245] was replicated in a virtual environment and evaluated using two in-depth user studies: 1) a VR *usability study* (N = 20) using VR replicas of the authentication methods used in the CueAuth paper as well as similar metrics, and 2) an online *security study* (N = 22) to study the observation resistance of the VR replica under two threat models, as done in CueAuth [245]. In advance of the *security study*, a pre-security study (N = 28) was conducted to define the user representation required for follow-up security evaluations based on VR authentication recordings. The results of the VR usability and security studies are compared with the real-world evaluation findings reported in the CueAuth paper [245].

The studies and the comparisons to the earlier work allow this chapter to draw novel insights into the advantages and disadvantages of VR studies for usability and security evaluations of prototypes, contributing partial answers to the final research question of this thesis:

RQ₆ What are the advantages and disadvantages of using VR studies for USEC research involving prototypes compared to traditional USEC research in physical laboratories and in the field?

The overarching research question, RQ₆, is fully addressed in chapter 7, subsequent to the remaining empirical research chapters. This chapter concludes by discussing the study results of the usability and security studies in Section 4.6 and Section 4.8, and how VR studies can be applied to facilitate and advance research in the USEC prototyping space. The thesis further contributes to the validation of using VR studies for simulating real-world USEC research in chapter 5 and chapter 6, and demonstrates how VR studies can advance USEC research.

4.1.1 Chapter Structure

The fourth chapter of this thesis contains three user studies that contribute to the first validation of using VR studies for real-world USEC research. It first outlines the ethics and the participants' compensations in Section 4.2. It then describes the studied USEC prototype, CueAuth [245], and its input methods in Section 4.3. Section 4.4 provides an overview of the studies. The chapter then discusses the usability evaluation findings of the VR replica of CueAuth [245] in Section 4.6. Finally, this chapter presents the security evaluation findings of CueAuth's virtual replica in Section 4.8 and concludes with a discussion of the findings in Section 4.10, contributing answers to RQ₂ and RQ₃.

4.2 Ethics and Compensation

The studies in this chapter were approved by the University of Glasgow College of Science & Engineering ethics committee (*ref: #300190050* and *ref: #300190215*). The participants

were compensated with €15.00 for the usability study and £7.50 each for the pre-security study and the main security study. The participants in the security study took part in a draw to receive an additional £7.50 based on their observation performance. This compensation method was used in the original real-world CueAuth study [245] and is frequently applied in security studies to motivate participants [158, 307, 308]. The participants could optionally share photos of notes they took during the security study for an additional compensation of £0.5. A new set of participants was recruited for each study to avoid any learning effects. The usability study in Section 4.6 was conducted in Austria due to COVID-19. The pre-security study in Section 4.7 and the main security study in Section 4.8 were conducted online (cf., Section 4.7.2 and Section 4.8.1).

4.3 The Selection of the USEC Prototype

USEC research covers a wide range of areas, many of which are not necessarily easy to create in-lab experiments for. As the Introduction and the Literature Review have shown, there is a lot of unexplored potential of the use of VR studies for facilitating and advancing real-world USEC research. To provide, for the first time, evidence of the feasibility of VR studies for USEC research, this thesis has to define a suitable prototype that allows for detailed comparisons between a usability and security evaluation in VR and the real world. As a result, this chapter reports on a holistic user evaluation of CueAuth [245], a fast and highly secure authentication system for public displays, such as vending machines and ATMs. CueAuth covers a range of input methods that are used in security research, including touch input [102, 540], mid-air input [6, 25], and eye-gaze input [104, 235, 272]. Furthermore, it provides a holistic usability and security evaluation, and its underlying concept has already been studied in different contexts [245, 530, 545], making *RepliCueAuth*, a virtual replica of CueAuth [245], an ideal candidate for the first validation of the use of VR studies for USEC research.

4.3.1 CueAuth and RepliCueAuth: An Overview

The virtual replica, *RepliCueAuth*, manifests the same characteristics and input methods as the original real-world CueAuth prototype [245]. To enter a PIN in CueAuth, users either perform touch gestures, mid-air gestures, or smooth pursuit eye movements [525]. The underlying concept of CueAuth [245] is based on cues on the screen (cf., Figure 4.2-2a, 3a, 4a). For touch and mid-air, the arrows on the respective digits show the users which gestures they have to input. The absence of an arrow indicates that the users have to tap (in touch) or perform a gesture towards the front (in mid-air). In eye gaze, CueAuth [245] employs smooth pursuit

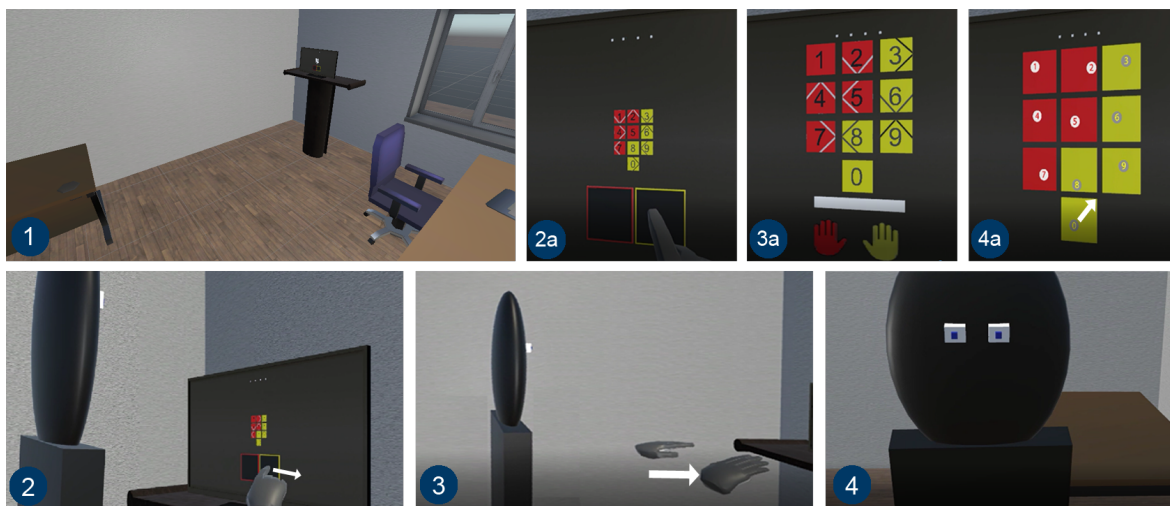


Figure 4.2: CueAuth’s authentication setup was transferred into a VR environment with a situated display that shows PIN-pads (cf., ②-a, ③-a, ④-a) featuring the cues [245]. To select a digit, the user responds to the cue displayed on its button. After selection, the cues are randomly reshuffled. To enter “0” via touch in the shown example, the user performs a touch gesture to the right in the yellow box to correspond to the yellow “0” button with a right-arrow (②). To enter “1” in mid-air, the user performs the gesture to the front with their left (red) hand as the “1” button is red and has no arrows (③). In eye gaze, each digit moves along a distinct trajectory, allowing the users to follow the movement with their eyes to make selections. To enter “0”, the user has to follow the diagonal movement of the digit with their eyes (④).

movements [525], a calibration-free gaze interaction method. Smooth pursuit eye movements are compared to the trajectories of animated targets (i.e., digits 0 - 9) to determine which digit users gaze at. The cues are randomly reshuffled after every user input due to security. The concepts of the input methods are described in more detail below:

- **Touch Input:** In touch, the users are required to observe which cue is shown on the digit and then perform the corresponding touch gesture in the respective box (cf., Figure 4.2-②a). For example, digits on the left are entered in the red box (i.e., 1,2,4,5,7), whereas digits on the right (i.e., 3,6,8,9,0) are entered in the yellow box. An arrow means a touch gesture to the displayed direction. For example, to enter the digit “3” in Figure 4.2-②a, a downwards touch gesture in the yellow box is required.
- **Mid-Air Input:** In mid-air, the users raise their hands and select digits via mid-air gestures in the direction of the corresponding arrow (cf., Figure 4.2-③a). The gestures are performed with the left hand if the digit is coloured red and with the right hand if the digit is coloured yellow. For example, to enter the digit “3” in Figure 4.2-③a, the users are required to perform a gesture to the right with their right hand.
- **Eye-Gaze Input** In eye gaze, the users provide input by following moving targets with their eyes, referred to as “pursuits” [525]. Pursuits’ advantages over location-based gaze

gestures are manifold. Smooth pursuit input does not require accurate gaze estimation; thus, it does not require eye tracker calibration and minimises pre-interaction times. Furthermore, Pfeuffer et al. [382] emphasised that calibration negatively impacts the usability and user experience when interacting with public displays as such interactions are often rather of shorter duration [343]. To select the digit “0” in Figure 4.2-4a, the users are required to follow the diagonally moving target with their eyes.

4.4 Overview of Studies

A repeated-measures design was applied to contribute a VR usability and security evaluation of a USEC prototype and shed the first light on the feasibility of using VR studies to evaluate real-world USEC prototypes. The study design is in line with the original real-world study of CueAuth [245] to avoid introducing any confounding variables. Conditions were counterbalanced using a Latin Square. The usability and main security study are designed as conceptual replications using “alternative methods” [551, 552], which means the VR studies in this thesis follow the original real-world study design [245], but they use a different research method to measure the dependent variables (for example, in a virtual lab environment instead of a physical laboratory). In the following sections, the usability and security evaluations are described. The results of each study are reported in Section 4.6 and in Section 4.8. Section 4.7 outlines the pre-security study, which informed the VR avatar used in CueAuth’s security evaluation in Section 4.8.

4.4.1 Overview: Usability Evaluation of RepliCueAuth

The usability study, which this thesis will discuss in Section 4.6, was conducted in person to ensure consistency of the study environment and the protocol between the participants. The participants answered various questionnaires directly in VR. Providing them with in-VR questionnaires [395, 432] ensures a consistent VR experience and does not break their focus [395]. Doing so also makes the methodology more applicable for future remote VR studies, as this thesis will show in chapter 6. Participants with similar user profiles as reported in the original real-world study [245] were recruited, i.e., normal/corrected-to-normal vision and no prior experience with cue-based authentication. The usability study collected the participants’ entry time, entry accuracy, and perceived workload when interacting with a virtual replica of CueAuth. The study concluded with semi-structured interviews to compare the participants’ opinions and preferences to the qualitative data reported in the CueAuth’s original real-world study [245].

4.4.2 Overview: Security Evaluation of RepliCueAuth

For the security evaluation, which this thesis will discuss in Section 4.8, Prolific [394] and pre-recorded videos recorded in the VR environment were used. Prolific is an established platform for online subject recruitment for scientific purposes and regularly used for advertising academic studies in HCI and USEC (for example, [14, 300]). The recordings of authentications in VR (cf., Figure 4.2-3/3a) were embedded in Qualtrics (licensed by the University of Edinburgh), an online survey tool that study participants can access via web browsers. The overall study procedure is in line with the original real-world study [245], where the participants were shown video recordings of authentications in the real world. However, to provide insights into the feasibility of using VR studies for USEC research, the recordings for this study were based on VR material. In other words, instead of a human in the real world who performed an authentication, the videos showed an avatar in the VR environment performing an authentication. The use of authentication recordings to assess a prototype's resistance to observations, i.e., its security against shoulder surfing, is a common research approach when evaluating the security of a USEC prototype (for example, [29, 98, 102, 239]). The security study collected the participants' successful attack rate when observing the authentications in VR, their attack duration, and their confidence in the guesses. As done in the original real-world study, the Levenshtein distances [280] between the correct PINs and the participants' closest guesses were calculated to conduct a more nuanced security analysis of participants' observation performance. The study concluded with semi-structured interviews on the participants' attacking strategies and their security and usability perceptions of the different input methods. Section 4.7 outlines the pre-security study to inform the VR user representation for the main security study in Section 4.8.

4.5 Data Analysis of the Usability and Security Study

The statistical analysis entails 1) an analysis of the repeated-measures usability, pre-security, and main security VR studies, and 2) a comparison between the results of the VR studies and those obtained in the original real-world studies [245]. The latter feeds the discussion of the validity of using VR studies for USEC research by comparing the VR-based usability and security results to the results reported as part of the real-world evaluation [245]. To visualise the data, bar charts and violin plots [204] were used. All data figures in this chapter, as well as in chapter 5 and chapter 6, were created using <https://ggplot2.tidyverse.org/> for [R].

4.5.1 VR Usability and Security Study Data Analysis

In the statistical analysis of the usability and security study, the same statistical tests as used in the original work [245] were applied. A repeated-measures ANOVA (IV: input method) was applied in the usability study, and a two-way repeated-measures ANOVA (IV: input method, threat model) in the security study. ANOVAs were applied on parametric and normal/near-normal distributed data as they are fairly robust to deviations from normality [49, 161, 444]. Post-hoc pairwise comparisons were Bonferroni corrected for controlling familywise errors. Greenhouse–Geisser adjustment was used to correct for violations of sphericity. For the qualitative analysis, a code book [433] based on the findings of CueAuth [245] was used to support the comparison. A VR code was added to capture VR-related comments that were not present in the original real-world usability study. An eye was kept out for potential new codes, but nothing new was observed. The content of the semi-structured interviews was fairly simplistic, so the lead researcher did all the qualitative data analysis.

4.5.2 Validation Data Analysis

The data set from the original real-world study of CueAuth was obtained through the original paper’s first author [245] to allow the thesis to comment on the findings’ differences regarding the study paradigms, i.e., real-world study vs VR study. A two-way mixed ANOVA with one between-subjects factor (study type: VR vs real world) and one within-subjects factor (input method) was applied in the usability study. A three-way mixed ANOVA with one between-subjects factor (study type) and two within-subjects factors (input method and threat model) was applied in the security study. Whilst this approach allows the thesis to reveal significant differences between the two study types, a non-significant outcome (i.e., p -value > 0.05) does not indicate that the values are equal or there is no effect of study type on the measures [233]. The sample size in the conceptual replication VR studies is determined by the original CueAuth study [245]. For statistical tests, such sample sizes increase the likelihood of type-2 errors, i.e., false negatives. Therefore, additionally to reporting non-statistically significant pairs between the VR studies and the real-world studies of CueAuth [245], the similar patterns found across the two study types are reported. In the following pages, this thesis uses a tag (i.e., —Validation—) to highlight the paragraphs that contribute to the comparison between the two study types (cf., page 100).

4.6 Usability Evaluation of RepliCueAuth

This section presents the usability evaluation of the virtual replica of CueAuth and ties its results to the original real-world study. Before reporting the results, details about the apparatus,

the implementation, and the methodology are provided.

4.6.1 Apparatus and Implementation

The VR prototype of CueAuth was implemented using Unity3D (C#), the Leap Motion SDK [338] for the hand tracking, and the Tobii XR SDK [514] for eye tracking. As headset the HTC VIVE Tobii DEV KIT (with an integrated 120 Hz Tobii eye tracker) [510] was used and connected to a VR-ready laptop (*Razer Blade 15, NVIDIA GeForce RTX 2080*) [403]. The implementation followed the implementation details reported in the CueAuth's original study [245] as best as possible. Due to the nature of VR some implementations differ, but the overall concepts and implementations remained the same:

- **Touch Input:** Instead of calculating the distance between on-screen touch points to detect touch gestures, as done in CueAuth [245], colliders and the `OnCollisionEnter`, `OnCollisionStay`, and `OnCollisionExit` event listeners [509] around the user's touch point were used. One collider was positioned at the user's initial touch point, and the others (left/right/top/bottom) ≈ 3.5 cm away; this value is based on pilot tests. A touch gesture is registered depending on which collider the user's finger collides with. If none of the colliders is touched, but the touch exits at the user's initial touch point, the system recognises a tap. Similarly to work by Kim et al. [249], who used an acrylic panel to provide VR users with haptic feedback when touching a virtual plane in VR, a physical surface was location-mapped to the VR touchscreen to provide passive haptic feedback when performing touch gestures on the virtual screen (cf., Figure 4.3).
- **Mid-Air Input:** Instead of tracking mid-air gestures through an external device, as done in the real-world study with a Microsoft Kinect One, two HTC VIVE trackers were



Figure 4.3: A physical surface in the real world (❶) was location-mapped to the virtual screen (❷) in VR. Following this approach provides the participants with haptic feedback when touching a virtual screen without using an actual touchscreen (highlighted in red).

attached to the participants' wrists. The default position is where the participants' hands are raised and parallel to the elbows (cf., the virtual avatar in Figure 4.2-3). A small threshold area (≈ 10 cm, determined through pilot tests) around the default position was defined as "no input area". Gestures were detected using colliders the same way as done for touch. After each gesture, the participants' hands had to return to the default position before the next input, which is in line with the real-world implementation [245].

- **Eye-Gaze Input:** As done in the original CueAuth paper [245], the implementation by Vidal et al. [525] was used to detect smooth pursuit movements. A moving digit, used as a stimulus for pursuits [525], is selected if the correlation between its trajectory and the user's eye movements exceeds a Pearson correlation coefficient threshold. The Pearson correlation coefficient has been calculated as done in the original real-world study of CueAuth [245]. The stimulus with the highest correlation above a predefined threshold (> 0.8 in [245]) to a user's eye movements is defined as the stimulus at which a user gazes at. The threshold, as well as the trajectories of the stimuli (circular, linear diagonal, and zigzag), are based on the original study [245]. Different configurations could lead to different entry accuracies and entry speeds [525].

4.6.2 Methodology and Study Design

The within-subjects usability evaluation of CueAuth, summarised in Section 4.4.1, follows the real-world study design [245] with the input method as the independent variable with three levels: touch input, mid-air input, and eye-gaze input. The entry accuracies, the entry times, and the participants' perceived workload using the NASA-TLX questionnaire [193] were collected. Additionally, the participants answered 5-point Likert scale questions, as done in the real-world study (for example, "*Input using touch is easy.*", from *Strongly Disagree* to *Strongly Agree* [245]). The participants then went through a semi-structured interview to allow the thesis to collect qualitative feedback and learn more about their experiences and opinions when using touch, mid-air, and eye-gaze input in VR (cf., Appendix C).

4.6.2.1 Procedure and Task

Each participant went through three blocks in total, one per condition. The order of the blocks, including a training session, an authentication session, and the questionnaires, was counterbalanced using a Latin square. The participants first filled in the demographics. The input methods were then explained before each authentication session, where participants authenticated using one of the input methods. The participants then performed training runs to become acquainted with the corresponding input method. These training runs were excluded from the analysis. The study prototype verbally announced each predefined 4-digit PIN.

Each participant had one chance to enter each PIN. After entering 16 PINs per authentication session, the participants filled in the questionnaires in VR. The same was repeated for the other conditions. The study concluded with semi-structured interviews.

4.6.3 Demographics

Twenty participants (8 female, 12 male) were recruited through social media, word-of-mouth, and local societies. Participants were on average 27.25 years (range: 18 - 57, SD = 8.31) and their demographics (gender, age) correspond roughly to the demographics of the participants in the real-world study: 13 female participants, ages ranging from 18 to 33 years (M = 24.1, SD = 3.9) [245]. The data of five participants (vs three in the real-world study [245]) had to be removed due to technical issues with the VR tracking system. Three (P3, P4, P16) in the touch input condition, and two (P14, P18) in the mid-air input condition.

4.6.4 Results

The reporting of the results follows the original real-world CueAuth paper [245]. First, the entry accuracies, the entry times, and the participants' perceived workloads are reported. Then, the qualitative feedback collected during the semi-structured interviews is reported. The statistical analysis follows the approach this thesis described in Section 4.5.

4.6.4.1 Entry Accuracy

There was a significant main effect of the input method on entry accuracy, $F_{(1,451,20,311)} = 5.791$, $p < 0.05$, $\eta_p^2 = 0.293$. Post-hoc pairwise comparisons revealed significant differences ($p < 0.05$) in entry accuracy between touch input (M = 89.31%, SD = 8.48%) and mid-air input (M = 80.21%, SD = 5.77%). No significant differences were found between the other pairs. Entry accuracies are high for all methods with M = 89.31% (SD = 8.48%) for touch, 81.25% (SD = 13.30%) for eye gaze, and M = 80.21% (SD = 5.77%) for mid-air input.

Validation

When comparing the VR study results to the real-world study results [245], no statistically significant interaction (input method \times and study type) for entry accuracy was found, $F_{(2,62)} = 0.401$, $p = 0.671$, $\eta_p^2 = 0.013$. There was no main effect of study type on entry accuracy, $F_{(1,31)} = 0.058$, $p = 0.812$, $\eta_p^2 = 0.002$. However, touch input was significantly more accurate than eye gaze in the real-world study [245], which was not the case in VR. The entry accuracies in the original real-world study were M = 93.38% (SD = 26.05%) for touch input, M = 82.72% (SD = 38.53%) for eye-gaze input, and M = 84.19% (SD = 39.1%) for mid-air input [245].

Usability Observation 1

There was no evidence of significant differences of the entry accuracies for touch input, mid-air input, and eye-gaze input between the VR study and the real-world study.

4.6.4.2 Entry Time

There was a significant effect of input method on entry time, $F_{(1,229,15.972)} = 69.778$, $p < 0.05$, $\eta_p^2 = 0.843$. Significant differences were found between eye-gaze input ($M = 16.75$ s, $SD = 4.36$ s) and touch input ($M = 6.06$ s, $SD = 1.87$ s), and between eye-gaze input and mid-air input ($M = 5.54$ s, $SD = 1.16$ s) ($p < 0.05$). Input using touch was significantly faster than eye-gaze input in VR, matching the results from the real-world study [245]. Touch input was significantly faster than mid-air input in the real-world study, which was not the case in the VR study. Figure 4.4 shows the distributions.

Validation

When comparing the entry times collected in VR to the real-world study, there was evidence of a significant interaction effect (study type \times input method), $F_{(1,054,31.614)} = 13.908$, $p < 0.05$, $\eta_p^2 = 0.317$. Follow-up analysis revealed a statistically significant difference in entry time when using touch input between the VR study and the real-world study, $F_{(1,33)} = 24.617$, $p < 0.05$, $\eta_p^2 = 0.427$. Touch input was significantly faster in the real world ($M = 3.73$ s, $SD = 0.67$ s) than in VR ($M = 6.06$ s, $SD = 1.87$ s). The other was found for eye gaze, $F_{(1,35)} =$

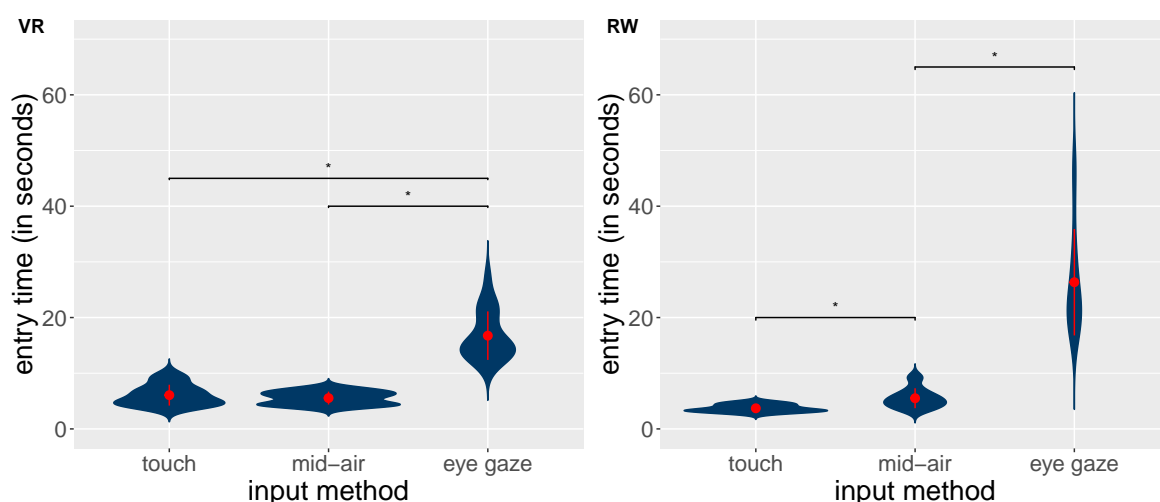


Figure 4.4: In the VR study, the participants authenticated significantly faster when using touch input and mid-air input than eye-gaze input. There was no evidence that touch input in VR is faster than mid-air input in VR, and vice versa. This is different from the real-world study (RW), where touch input was significantly faster than mid-air input and eye-gaze input [245]. Red point range denotes mean \pm standard deviation. * denotes statistical significance, $p < 0.05$

15.728, $p < 0.05$, $\eta_p^2 = 0.310$. Input with eye gaze was significantly faster in VR than in the real world ($M = 16.75$ s, $SD = 4.36$ s vs $M = 26.35$ s, $SD = 9.56$ s). No significant difference between the study types was found when providing mid-air input ($M = 5.54$ s, $SD = 1.16$ s for VR vs $M = 5.51$ s, $SD = 1.79$ s for the real world [245]).

Usability Observation 2

Touch input was significantly faster in the real world than in VR, whereas eye-gaze input was significantly faster in VR than in the real world. Entry times using mid-air input remained the same across VR and reality.

4.6.4.3 Perceived Workload (NASA-TLX)

There was no evidence of significant differences of the mean NASA-TLX values between the input methods, $F_{(2,38)} = 0.389$, $p = 0.681$, $\eta_p^2 = 0.020$. The overall task load indices are $M = 34.83$ ($SD = 23.61$), $M = 34.0$ ($SD = 19.68$), and $M = 30.33$ ($SD = 18.33$) for touch, mid-air, and eye gaze. Despite a non-significant main effect, repeated-measures ANOVAs on the level of each NASA-TLX dimension were run to investigate if there is an effect on the level of the individual subdimensions. A significant main effect was found for input method on performance, $F_{(2,38)} = 7.615$, $p < 0.05$, $\eta_p^2 = 0.286$. Post-hoc pairwise comparisons revealed a significant difference between eye-gaze input and touch input ($p < 0.05$), and between eye-gaze input and mid-air input ($p < 0.05$). Figure 4.5-VR shows the mean scores.

Validation

When comparing the participants' perceived workload in the VR study to the participants' perceived workload in the real-world study, a two-way mixed ANOVA revealed a significant interaction effect (study type \times input method), $F_{(2,76)} = 7.233$, $p < 0.05$, $\eta_p^2 = 0.160$. Follow-up ANOVAs on the level of each input method revealed a statistically significant difference in the participants' perceived workload when using eye gaze between the two study types, $F_{(1,38)} = 7.803$, $p < 0.05$, $\eta_p^2 = 0.170$. Follow-up analysis on the level of each NASA TLX's sub-dimension revealed a significant difference ($p < 0.05$) between eye gaze in VR and the real world regarding physical workload, frustration, and effort. For the VR study, the mean raw values for physical workload, frustration, and effort were $M = 18.50$ ($SD = 21.83$), $M = 29.25$ ($SD = 29.79$), and $M = 40.0$ ($SD = 27.96$). For the real-world study, the values were $M = 46.5$ ($SD = 26.71$), $M = 39.25$ ($SD = 29.44$), and $M = 57.75$ ($SD = 24.73$). Figure 4.5 shows the values for each input method and study environment.

Usability Observation 3

There was no evidence that the participants' perceived workload differed regarding touch input and mid-air input in both the real world and in VR. However, the participants reported significantly lower physical workload, less frustration, and less effort when providing gaze input in VR than in the real world.

4.6.4.4 Qualitative Feedback

Qualitative feedback was collected through the semi-structured interviews (cf., Appendix C). Although a strict comparison of qualitative data is challenging [160], there are many similarities between the VR study and the original real-world study [245]:

Theme 1: Exposure to the Input Methods. Similar to the participants' comments in the real-world study, the VR study participants reported being previously exposed to touch input and mid-air gestures (for example, through the Xbox video game console). However, they were less exposed to gaze-based interaction, which was experienced by only two participants before. Interestingly, nine participants in the original real-world study had prior experience with eye gaze (vs two in the VR study) because of previous user studies [245]. This difference in the participants' experiences between the study types makes it likely that the

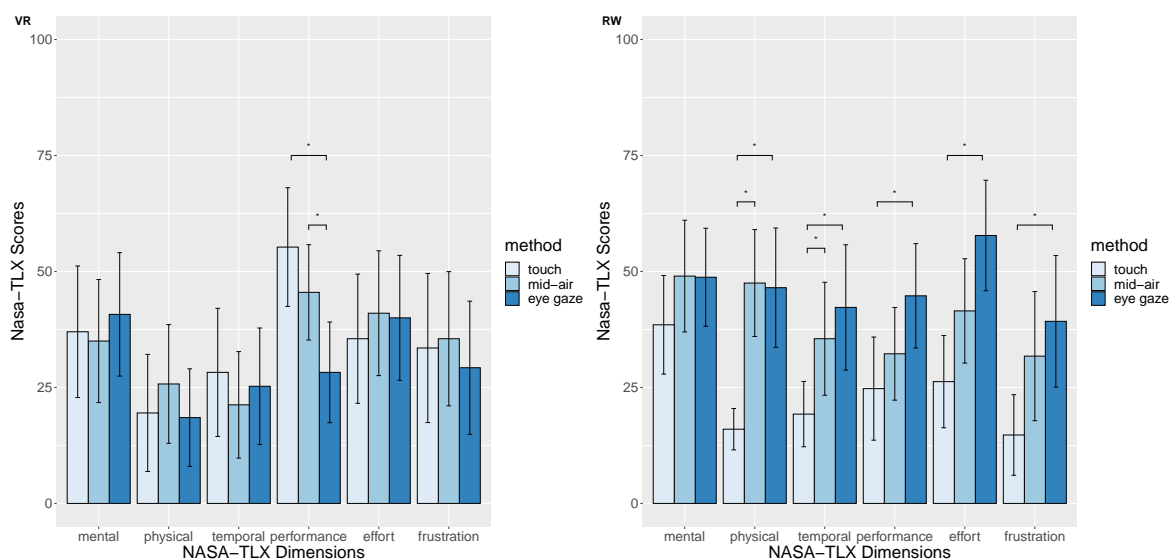


Figure 4.5: There was no evidence of significant differences in the VR study between the mean NASA-TLX values on the level of each input method, indicating that the participants perceived all three methods as equally demanding. However, analysing the NASA-TLX dimensions revealed a significant perceived performance difference ($p < 0.05$) between eye gaze and touch, and between eye gaze and mid-air. Eye gaze in the real-world study (RW) was perceived as more demanding than in VR. Black lines denote the 95% confidence interval.

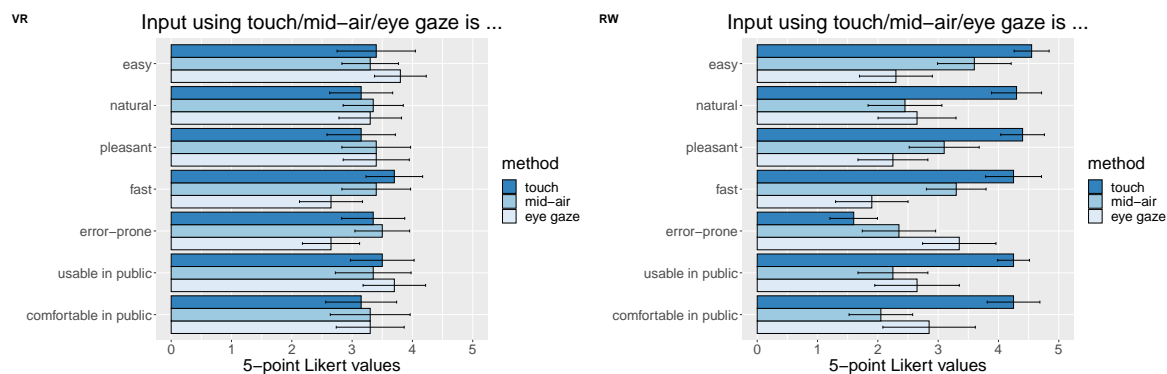


Figure 4.6: The participants in the VR study rated the three input methods on 5-point Likert scales, from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*). RW shows the results of the original real-world study [245]. Black lines denote the 95% confidence interval (CI).

original real-world study of CueAuth [245] recruited participants within another environment (for example, in an academic setting where user studies are quite common). Although this finding does not contribute to the validation of using VR studies for real-world USEC research, it shows that participants' exposure to the input methods depends on the participant recruitment and not necessarily on the environment (VR vs real world). There was no evidence of notable differences in the participants' perceptions of the input methods when experiencing them in VR and the real world.

Theme 2: Perception of the Input Methods. The participants' perceptions of the methods in VR matched those from the real-world study. Whilst the ranking of the input methods in Section 4.6.4.5 suggests that they preferred eye-gaze input over mid-air input and touch input, they associated touch input with more positive attributes than eye-gaze input. Examples include intuitive, realistic, and effortless. Although mid-air gestures were similarly positively perceived, there were more negative attributes associated with mid-air gestures than touch input. The participants voiced that mid-air input requires more explicit movements than touch input. Furthermore, it has been stated that mid-air gestures look weird. For example, P13 expressed that “[mid-air input is] neither fish nor fowl” (P13). Eye gaze was perceived as long-winded and exhausting, but safer than touch input and mid-air input. Both mid-air input and eye-gaze input were described as hygienic.

Theme 3: Usability of the Input Methods. There were mixed comments on the usability of the input methods. Touch was found simple and familiar; however, the participants mentioned that providing touch gestures with virtual hands feels strange. Whilst some participants perceived mid-air gestures as comfortable, others noted that it feels weird in public: “looks like a jumping jack” (P17). Eye gaze was perceived as long-winded but secure. The interviews suggested that touch input was perceived as slightly more usable than

mid-air input and eye-gaze input. The additional technological layer, i.e., the virtual hand, was taken negatively, which deemphasises the original advantages of touch input over mid-air and eye-gaze input in the VR study (cf., Figure 4.6-VR/RW).

Theme 4: Enhancements of the Input Methods. The participants voiced the lack of proper feedback when providing input and suggested extending the interaction space of mid-air gestures by more subtle finger gestures or facial gestures. Similar enhancements were mentioned by the participants of the real-world study [245]. Others noted technological limitations in the hand tracking and the gesture detection. For eye gaze, the participants criticised the pixelated targets they had to follow with their eyes.

Theme 5: Perception of the VR Environment and Its Impact. Most participants voiced that the virtual environment did not impact how they interacted with the prototype. However, a few mentioned that they felt isolated in VR, which allowed them to pay more attention to the task in VR than in the real world. P13 voiced that they were unaware of the researcher’s presence during the study. P19 mentioned that they were careful when performing specific movements because they feared bumping into real-world obstacles. P20’s opinion was different, voicing that they treated the VR environment as a safe space and felt “freer” to perform gestures. The participants described people in the real world as “additional noise”, but there was no evidence that this was similarly perceived in the VR study. P6 mentioned the virtual environment was “*too clean to be realistic*” (P6).

Overall, the interviews represent a similar picture to the participants’ answers to the Likert scale questions (cf., Figure 4.6). Compared to the participants’ qualitative feedback in CueAuth’s real-world study [245], touch input was perceived as, for example, more challenging and more error-prone in VR. These results can be attributed to the hand tracking used in the VR study, which this thesis discusses further in Section 4.6.5.

Usability Observation 4

Most qualitative feedback was similar across the VR study and the original real-world study of CueAuth. However, the different sensing capabilities impacted the participants’ perceptions and preferences of some input methods. The VR study negatively impacted the advantages of touch input over mid-air input and eye-gaze input compared to the CueAuth’s real-world study.

4.6.4.5 Usability Ranking: The Participants' Preferred Input Method

The participants were asked to rank their preference of the three input methods. Raw scores were multiplied by their weight factor: $\times 3$ for rank 1, $\times 2$ for rank 2, $\times 1$ for rank 3, and then summed up to compute weighted scores. Eye-gaze input was the most preferred input method (45), followed by mid-air input (41), and touch input (37), which is in line with the qualitative feedback received in the interviews, but differs from the real-world study where participants preferred touch input over mid-air input and eye-gaze input [245].

Usability Observation 5

In contrast to the real-world study, where touch was the most preferred input method and eye gaze the least preferred input method, the VR study participants preferred eye-gaze input over mid-air input and touch input.

4.6.5 Discussion of RepliCueAuth's VR Usability Evaluation

The conceptual replication study of CueAuth resulted in similar entry accuracies in all three input methods (Usability Observation 1). There was no evidence that the participants' perceived workload was significantly different in the VR study compared to the original real-world study when providing touch input and mid-air input (Usability Observation 3). Despite the promising results, the usability study suggests that tracking accuracy, i.e., hand and eye-gaze tracking, affects a prototype's usability evaluation findings when evaluated in VR. The impact of the technology on the usability evaluation findings of virtual replicas of real-world prototypes was apparent in the VR study as follows:

Compared to the real-world study of CueAuth [245], touch input was significantly slower in VR than in reality, and eye-gaze input was substantially faster in VR than in the real world (Usability Observation 2). Eye-gaze input was perceived as significantly less physically demanding, less frustrating, and required less effort in VR than in the real world (Usability Observation 3). Eye-gaze input was the most preferred input method in the VR study, whereas it was the least preferred input method in the real-world study (Usability Observation 5). The differences between the VR study and the real-world study show that the use of VR does not necessarily provide the often desired "all-in-one solution" [556] and that USEC researchers need to have a clear vision what they can expect from usability evaluations on VR replicas of real-world USEC prototypes. However, as chapter 5 and chapter 6 will show, there are many opportunities for the use of VR studies to provide USEC researchers with insights into the usability and security of prototypes and advance research in this space. The differences in the participants' touch and eye-gaze input between the VR study and the real-world study are discussed in more detail in the next two sections. The first research question of this chapter,

RQ₂, will be answered in the conclusion of this chapter in Section 4.10, together with the findings of CueAuth's VR-based security evaluation (cf., Section 4.8).

4.6.5.1 Touch Input in VR: The Same as in Reality, but Still Different!

Contrary to mid-air input and eye-gaze input, where no additional virtual artefacts are required when providing input in VR, touch requires visualising the participants' hands because humans leverage the visual feedback of their hands to provide precise input in the real world. In the VR study, the virtual hand is located between the user and the interaction, which negatively affects the input performance, as shown in this chapter in Section 4.6.4.2. The participants in the VR study voiced the virtual hand did not always perfectly align with their real hand. Although the participants associated touch input with positive attributes, their entries were not as fast as those logged in the real world. This finding is in line with Knierim et al. [255], who showed that the participants' typing performance is affected by virtual avatar hands.

The research presented in this chapter (and also in the remaining thesis) intentionally abstained from using expensive hardware such as an OptiTrack system or other high-end sensors to track the participants' hand and finger movements. One of the goals of VR studies is to cut down prototyping costs and make studies of this type more accessible to the broader USEC community, which is only possible if the additional VR hardware required for such studies can be kept low. That being said, this thesis does not put forward an empirical (and/or systematic) evaluation of the costs of VR studies for simulating real-world research. However, creating several virtual replicas using the same commodity hardware is more affordable than building individual hardware prototypes or using specialised, expensive tracking equipment each time. This implies that using VR studies can contribute towards more accessible and cost-efficient USEC research in the long run. However, although the technological limitations potentially disappear in the future due to improved VR technology (for example, improved hand tracking) and acquaintance of humans with VR, the current limitations suggest that the use of existing "low-budget" VR technology is sometimes not able to provide the participants with exactly the same experiences they would face in the real world, particularly when introducing additional virtual layers for input (for example, virtual hands that are mapped to users' real hands).

4.6.5.2 The Consequences of Different Eye Tracking Approaches

Contrary to the real-world study, where the participants preferred touch input the most, the participants in the VR study preferred eye-gaze input over touch input and mid-air input (Usability Observation 5). Additionally, they provided faster eye-gaze input in VR than in the real-world study (Usability Observation 2). This difference in performance and perception of eye gaze is due to the different eye tracking systems used. In the real-world study, Khamis

et al. [245] used a stationary eye tracker mounted at the bottom of the display. This likely resulted in the eye tracking quality being influenced by their participants' heights, distances to the display, and the ambient and lighting conditions in their study room. These are known problems in studies that involve stationary eye trackers [92, 244, 360]. In fact, the authors of CueAuth [245] admit that "error rates and entry times are influenced by [their] setup and implementation" [245].

In the VR study in this thesis, an eye tracker integrated into the head-mounted display was used. This configuration is the most common setup of VR headsets with eye trackers due to the form factor of current headsets. This means that many real-world artefacts that negatively impact eye tracking quality (for example, bad lighting conditions) were absent in the VR study. Eye-gaze input in VR was independent of the participants' heights, positions, and the surrounding lighting conditions. This explains why participants were faster and preferred using eye gaze in the VR study compared to the participants in CueAuth's real-world study. However, the findings of both the VR study and the original real-world study of CueAuth [245] depend on the technology used. A more advanced stationary eye tracker than the one used in the original real-world study [245] would likely achieve results similar to those found in the VR replication study.

4.6.5.3 Summary and Next Steps

In summary, the findings of the usability study show that the use of VR studies for USEC research can help to mitigate the limitations of hardware used in the real world. However, results from VR studies can also be misleading: if researchers want to assess the usability of gaze-based interaction assuming ideal tracking conditions, then VR would help them to achieve that. On the downside, if researchers want to determine the usability of the same prototype with noise and other external factors, then they would need to account for these factors in VR. Otherwise, they risk being misled into thinking that the (virtual) prototype works better than it does (in the real world). As the Literature Review in chapter 2 has shown, HCI and USEC researchers have traditionally considered similar trade-offs when comparing lab studies with field studies, with one optimising for control and the other for ecological validity. The transferability of the quantitative results from VR to the real world highly depends on how well reality and its limitations are emulated. It is important to recognise that using VR studies is not an alternative to more traditional research methods such as laboratory studies and field studies, but it complements existing research methods by simulating real-world USEC research artefacts and study environments that may be challenging (and sometimes even infeasible) to study in the real world.

So far, this thesis chapter has put forward the use of VR studies for the usability evaluation of USEC prototypes, with CueAuth [245] as an example prototype. As a next step, Section

4.8 will describe and validate the use of VR studies for human-centred security evaluations to provide a full picture of the VR's feasibility for simulating real-world USEC research and conducting user testing. Validating the use of VR studies for both usability and security evaluations is important to design both usable and secure USEC prototypes [9, 568]. The importance of usability and security evaluations is evidenced by the large number of USEC works on prototypes that cover both a usability and security assessment of their research artefacts (cf., Table 2.1 in the Literature Review). To allow this thesis to comment on the suitability of VR studies for security evaluations, the next sections will first present the pre-security study (Section 4.7) and then the main security study (Section 4.8), which evaluates the CueAuth's security using recordings based on a VR implementation.

4.7 Pre-Security Study: Defining the VR Avatar

The pre-security study is used to define the avatar fidelity, i.e., the extent to which different avatars convey the look of a human in the real world (similar to [526]), that is required to represent a human's interactions in VR. As shown in Section 2.5 in the Literature Review and voiced by the expert interviews in chapter 3, the security evaluations of USEC prototypes are often concerned with shoulder surfing, which involves humans and how they behave when interacting with USEC prototypes. The Literature Review in chapter 2 has shown that avatars play an essential role in virtual environments [527], but it is often not clear which avatar best imitates humans and their interactions. A common approach is to embody avatars of various appearances, from abstract to more realistic avatars [364]; however, to date, there has not been an empirical evaluation of the VR avatar required for subsequent security evaluations on VR replicas of real-world USEC prototypes. In order to validate the use of VR studies for security evaluations on USEC research prototypes, it is important to ensure that the user representation used in virtual environments matches a human in the real world.

The pre-security study aims to fill this gap by the first investigation of the impact of different avatar fidelities on the observers' interaction identification performance to then allow for a follow-up security evaluation of CueAuth [245]. In other words, it compares how well different avatars can be observed when performing touch, mid-air, and eye-gaze input. CueAuth's original real-world study conducted the security evaluation on a user in the real world [245]. However, the security study in this thesis will use a virtual avatar to represent a human. Researching and defining the avatar fidelity before the security study of CueAuth is vital to reduce the threat of the user representation as a potential confounding variable.

The pre-security study in this section covers three avatar fidelities ranging from an abstract avatar to a highly realistic avatar by Microsoft Research [169]. It synchronises interactions performed by a human in the real world with user interactions in VR. The different avatars are



Figure 4.7: The pre-security study investigates the impact of an avatar’s fidelity on a bystander’s performance when identifying the avatar’s touch, mid-air, and eye-gaze gestures. There was an abstract avatar (②) and two more realistic avatars (③, ④), which are provided by Microsoft Research [169] and modified based on the research purpose. Touch, mid-air, and eye-gaze gestures performed by a human in the real world (①) were used as a baseline.

evaluated in front of a public display that allows for touch input, mid-air input, and eye-gaze input (cf., Figure 4.7). The gestures performed by the user in the real world (baseline) and the VR avatars are summarised in Table 4.1. Touch, mid-air, and eye-gaze gestures were chosen to cover a wide range of different input methods and gestures when interacting in virtual environments (for touch: [307, 497]; for mid-air: [481, 498]; for eye gaze: [307, 384]). Furthermore, studying these three input methods helps the thesis to tie the results to CueAuth’s security evaluation, which this thesis presents in Section 4.8.

4.7.1 Apparatus and Implementation

Unity 3D (C#) was used to implement a virtual environment and the different avatars. A human in the real world performed all gestures outlined in Table 4.1 in the real world and VR. The interface of CueAuth was replicated to support performing the gestures for touch, mid-air, and eye gaze (cf., Figure 4.7). For eye-gaze gestures, moving targets were used as stimuli that move along the trajectories as outlined in Table 4.1. This approach is equivalent to prior work, which used moving targets to enable gaze-based interaction [245, 525]. All avatars used

in the study are tracked through the HTC VIVE VR headset’s position and rotation in the space. A Leap Motion Controller was used for the hand and finger tracking. The integrated Tobii eye tracker (and the Tobii XR SDK [514]) was used for the eye-gaze movement of the avatars. For the more realistic avatars, an avatar provided by Microsoft Research [169] was used and slightly modified to fit the purpose of the research. The joint angles of the arms were calculated using Unity’s Animation Rigging package and Inverse Kinematics (IK) [213]. This allowed to increase the realism of the avatars and couple the avatar’s hands to its body.

No additional hardware or high-end tracking systems were required, except for two additional HTC VIVE trackers. Implementing the highly realistic VR avatar requires significantly more expertise and effort compared to the implementation of the abstract avatar, which does not require any inverse kinematics calculations nor additional HTC VIVE trackers and is based on simple 3D shapes, i.e., a cube for the head and three cuboids that are merged together for the body. The same low-polygon hand asset was used for all avatars to avoid introducing potential confounding variables. For the VR headset and to record the interactions in the virtual environment, the Tobii HTC VIVE [510], a VR-ready laptop (*Razer Blade 15, NVIDIA GeForce RTX 2080*), and OBS [480] were used. The interactions in the real world were recorded with a NIKON D5300 single-lens reflex camera on a tripod.

4.7.2 Methodology and Study Design

The pre-security study was designed as a within-subjects experiment with two factors. The first factor was the input method with three levels: touch input, mid-air input, and eye-gaze input. The second factor was the avatar that represents a person with four levels: 1) human in the real world (baseline); 2) an avatar that shows eyes and hands only, similar to avatars used in social virtual rooms like Mozilla Hub [341]; 3) an avatar that included a realistic virtual body and head as previous work found that eye gaze is closely associated with head movements under natural conditions [45, 463] and could thus affect humans when observing

Input Method	Gesture	Number of different Gestures	Description
touch	left/right	2	Touch gesture to the left/right side.
	up/down	2	Up- and downwards touch gesture.
	tap	1	Single tap gesture on the screen surface.
mid-air	left/right	2	Mid-air gesture to the left/right.
	up/down	2	Up- and downwards mid-air gesture.
	front	1	Mid-air gesture to the front.
eye gaze	linear diagonal	4	Diagonal eye movements (all four directions).
	circular CW/CCW	2	Clockwise and counter-clockwise circular eye movements.
	zigzag	2	Vertical/horizontal zigzag eye movements.

Table 4.1: Gesture set based on CueAuth [245], covering touch, mid-air, and eye gaze.

interactions; and 4) a full-body avatar, where virtual eyes and hands are attached to a highly realistic avatar [169]. Figure 4.7 shows the avatars. Table 4.1 summarises the gestures.

Qualtrics [397], an online survey tool accessible via the standard web browsers, and Prolific [394], an established crowdsourcing platform for participant recruitment, were used to deploy the study online. Compared to the usability evaluation presented in Section 4.6, both the pre-security study and the security study in Section 4.8 are non-immersive VR studies. VR setups were used to simulate and record authentications, but the participants experienced those recordings on a traditional desktop screen. Assessing a USEC prototype's security against observations through video recordings on a traditional desktop screen is a common research method in USEC (for example, [102, 245, 308, 530]). Furthermore, the gestures by the human in the real world and the avatars in VR were pre-recorded as this is a common approach in the literature when assessing a prototype's resistance against observations [27, 102], studying human action perception [273], or examining the users' willingness to perform gestures within a specific context for a particular task [547]. Recording the interactions and embedding video prototypes [547] into a survey allowed conducting the study online and providing all participants with the same set of material, eventually contributing to internal validity. The video material for the study is publicly available¹.

In summary, the pre-security study represented an interactive online survey with (VR) video prototypes that portrayed different human gestural movements. The same research method was used for the main security study of CueAuth in Section 4.8.

4.7.2.1 Procedure and Task

After obtaining informed consent, demographics were collected. The study then proceeded with explainer videos of all three input methods and the gestures in Table 4.1 to introduce the different input methods and the participants' tasks. Figure 4.7 shows excerpts of the explainer videos for all three conditions. The same camera position and angle was used throughout the study. Attention check questions were added to ensure the participants understood the input methods and their study tasks. The participants were then shown pre-recorded interactions performed by a human in the real world or one of the three avatars in the virtual environment. The order of the conditions was randomised and balanced using the Qualtrics' Randomizer [398]. The participants watched the gestures, i.e., left/right/up/down/tap in *touch*, once each before providing a guess. After each avatar, the participants filled in the NASA-TLX questionnaire [194] to indicate their perceived workload when observing the interactions. Reporting the participants' perceived workload when observing human interactions is a common method in human-centred security research [245, 281]. The study concluded with

¹All VR videos are stored on https://youtube.com/playlist?list=PLs1tzNuOyzfwdvIpth_3NjDC2jb.T6LpI, last accessed 22/01/2023

5-point Likert scale questions asking about the participants' perceived realism of the avatars and a ranking of the avatars in terms of preference when observing them.

4.7.3 Demographics

Twenty-eight participants ($N = 28$) were recruited. The data of eight participants were removed due to several reasons. Some participants mentioned that they faced some issues with the video playback, whilst others provided low-quality feedback throughout the study, indicating that they did not meaningfully participate. The importance of cleaning data and removing low-quality responses to increase the ecological validity has been discussed in previous works (for example, [133, 339, 409]). Therefore, the analysis is based on 20 participants (11 male, 9 female) aged between 18 and 54 ($M = 31.79$, $SD = 10.51$). Out of the 20 participants, 19 (95%) mentioned that they had heard about the term "Virtual Reality" before, and nine participants (45.0%) voiced that they have experienced VR before.

4.7.4 Results

The following sections report the results of the participants' a) interaction identification performance, the number of correctly identified gestures for touch input, mid-air input, and eye-gaze input; b) perceived workload when observing the interactions using the NASA-TLX questionnaire [194] as a common approach for improving shoulder surfing resistance is to overwhelm the observer's short-term memory [119, 240]; c) avatar preference; and d) perceived realism of the avatars.

4.7.4.1 Successful Interaction Identifications

There was no evidence of a statistically significant two-way interaction between avatar and input method on the number of successful interaction identifications, $F_{(3.529, 67.043)} = 2.068$, $p = 0.103$, $\eta_p^2 = 0.098$, no main effect of avatar, $F_{(3, 57)} = 0.285$, $p = 0.836$, $\eta_p^2 = 0.015$, and no main effect of input method, $F_{(1.156, 21.973)} = 1.255$, $p = 0.297$, $\eta_p^2 = 0.062$. The overall average values of successful interaction identifications are $M = 14.5$ ($SD = 2.56$; overall: 80.56%) for observations on the human in reality, $M = 14.6$ ($SD = 2.80$; overall: 81.1%) for the abstract avatar, and $M = 14.25$ ($SD = 2.97$; overall: 79.17%) and $M = 14.45$ ($SD = 2.65$; overall: 80.28%) for the two more realistic avatars. The individual values on the level of each input method are as follows, ordered by the user representation: human in the real world, abstract avatar, more realistic avatar with a body and a head, full-body avatar: Touch: $M = 5.0$ ($SD = 0.0$), $M = 4.75$ ($SD = 0.55$), $M = 4.45$ ($SD = 0.51$), $M = 4.65$ ($SD = 0.59$); Mid-Air: $M = 4.65$

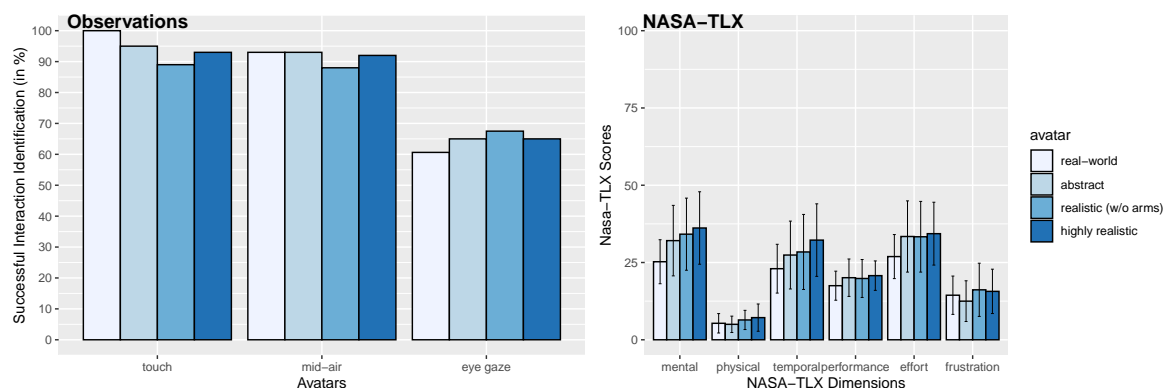


Figure 4.8: The participants' interaction identification performance and the perceived workload when observing the VR avatars and the human in the real world performing touch, mid-air, and eye-gaze input as shown in Figure 4.7. There was no evidence that interactions performed by an abstract avatar were easier/more difficult to observe than more realistic avatars and a human in the real world. The same was found for participants' perceived workload when observing the VR avatars.

(SD = 0.75), M = 4.65 (SD = 0.67), M = 4.40 (SD = 0.75), M = 4.60 (SD = 0.68); Eye Gaze: M = 4.85 (SD = 2.34), M = 5.20 (SD = 2.38), M = 5.40 (SD = 2.39), M = 5.20 (SD = 2.35).

All in all, there was no evidence that observations on one avatar were more accurate than on one of the others and that observations on a human in the real world were easier/more challenging than on the VR avatars. Figure 4.8 summarises the number of successful interaction identifications for each avatar and input method.

4.7.4.2 Perceived Workload (NASA-TLX)

A one-way repeated-measures ANOVA on the participants' mean perceived workload when observing the different avatars and the human in the real world was conducted. There was no evidence of a significant effect of avatar on the participants' perceived workload when observing them, $F_{(2.138, 40.626)} = 2.922$, $p = 0.062$, $\eta_p^2 = 0.133$. The mean NASA-TLX values were low for all four avatars: M = 18.74 (SD = 8.97) for observing the human in the real world, M = 21.75 (SD = 13.73) for the abstract avatar, and M = 23.06 (SD = 14.64) and M = 24.39 (SD = 14.12) for the more realistic avatars. The results show that the participants' perceived workload was slightly lower when observing the person in the real world. Figure 4.8 shows the mean values of each NASA-TLX dimension and avatar.

4.7.4.3 Perceived VR Avatar Realism

For eye gaze, a Friedman test revealed a statistically significant difference in perceived avatar realism depending on the type of the user representation, $\chi^2(2) = 9.435$, $p < 0.05$. The more realistic avatars were indeed perceived as more realistic than the abstract avatar ($p < 0.05$).

The median values were 2.5 for the abstract avatar and 4.0 for the more realistic avatars. For mid-air input, there was no statistically significant difference in perceived avatar realism depending on the type of the user representation, $\chi^2(2) = 1.660$, $p = 0.436$. The median values were 4.0 for all three avatars. The same was found for touch, $\chi^2(2) = 1.714$, $p = 0.424$, with median values of 4.0 for all three avatars.

For the overall perceived avatar realism, there was no evidence of a significant difference between the abstract avatar and the more realistic avatars, $\chi^2(2) = 2.324$, $p = 0.313$. The median realism values were 4.0 on a 5-point Likert scale for all three VR avatars.

4.7.4.4 Preference of User Representation

The participants were asked to rank the human in the real world and the VR avatars based on their preference of observing them, i.e., “Which user representation made it easier to observe the different interactions (1 = best, 4 = worst)?”. Raw scores were multiplied by their weight factor: $\times 4$ for rank 1, $\times 3$ for rank 2, $\times 2$ for rank 3, $\times 1$ for rank 4, and summed up to compute weighted scores. For mid-air input, the human in the real world achieved the highest average ranking (74), next to the full-body avatar (55), the abstract avatar (39), and the more realistic avatar (36). A similar pattern was found for touch, with the person in the real world being preferred (76) compared to the full-body avatar (51), the more realistic avatar (40), and the abstract avatar (39). The average ranking for eye gaze was the highest for the person in the real world (55), followed by the abstract avatar (52) and the full-body avatar (50). The more realistic avatar achieved the lowest average ranking in eye gaze (46).

Average ranking scores for each avatar were calculated to comment on the participants’ overall avatar preference. The results suggest that the human in the real world achieved the highest score (avg ranking score = 68.33) next to the highly realistic full-body avatar with an average ranking score of 52. The other more realistic avatar (w/o arms) and the abstract avatar were ranked similarly, with average scores of 40.67 and 43.33, respectively. The participants’ avatar preference is reflected in the qualitative feedback reported in Section 4.7.4.5.

4.7.4.5 Qualitative Feedback

A qualitative data analysis was applied to spot the main patterns in the participants’ feedback. The data set was reasonably simplistic and small, with 1-2 sentences for each question. Therefore, the lead researcher did all the qualitative data analysis. Participant numbers (*P1 to P20*) are used to ensure anonymity whilst presenting rich raw data.

When observing the mid-air interactions, there was a general agreement that the person in the real world was the easiest to observe. Some participants mentioned that they needed arms for clarity. For example, “the real world was easiest for me followed by the [most

realistic one] which is closest to real world. I think this is because of the fully developed arm extension” (P20). Interestingly, the need for arms was mentioned in combination with the more realistic avatar, but not with the abstract avatar, which suggests some kind of uncanny valley effect [337, 456, 477] as the participants did not necessarily expect to see arms in the abstract avatar condition. The same was mentioned for touch. This is an interesting finding and aligns with the comments by McMahan et al. [320] on interaction fidelity. They speculated that the reason why low degrees of interaction fidelity have reached comparable user experience as high degrees of interaction fidelity may be an effect of familiarity where users link high fidelity VR to the real world, whereas their perception of low fidelity VR builds on associations that are rooted in their familiarity with existing HCI interfaces [320].

Some participants noted that all avatars were perceived as easy to observe: *“they were all easy to identify so any would be fine”* (P1). However, one participant mentioned that *“the lack of arms in the mid-fidelity avatar was off putting”* (P3). Contrary to P3’s comment, P6 voiced that the *“low-fidelity avatar was more obvious and less distracting as it’s so basic”* (P6). The general feedback was that the attached arms helped the participants to identify the different gestures. One participant mentioned that the movements in the real world were smoother and that the virtual hands of the avatars did not come with visual and physical feedback: *“the physical feedback of the human hand touching the screen helped me see exactly what was happening. if the 3D finger bent back a little as it touched the screen (even if it was in an unrealistic uncanny way) then maybe that would help”* (P9). For eye gaze, the participants mentioned that the human in the real world slightly moved their head when performing the eye-gaze gestures, whereas this seemed to be less present when observing the VR avatars. Other participants mentioned that observing the person’s real-world eye movements was more straightforward than the avatar’s eye movements. One participant explained this around the fact that they are used to seeing eye movements from real humans rather than avatars: *“the real world was the clearest to me. I think because I am most used to seeing [eye] movements from real humans.”* (P20). Contrary to P20’s comment, P15 voiced that the human in the real world blinked with their eyes, making it harder to observe eye movements in the real world.

In summary, the qualitative feedback suggests that the comparisons mostly happened between the person in the real world and the virtual avatars. The participants mentioned surprisingly few differences between the VR avatars. However, one point many participants addressed was the lack of arms in one of the more realistic avatars (cf., Figure 4.7-3 on page 110).

4.7.5 Discussion of the Pre-Security Study and Its Implications

The findings of the pre-security study suggest that an abstract avatar’s interactions are distinguishable to the same extent as interactions performed by a highly realistic avatar and a human in the real world (cf., Section 4.7.4.1). There was no evidence that observing one of

the VR avatars led to a higher perceived workload than the others. However, the participants' perceived avatar realism, their avatar preference when observing the interactions, and the qualitative feedback suggest that the more realistic avatars were preferred over the abstract avatar. All in all, this thesis section argues that it is important to consider the research questions and the aim of the research when deciding on the avatar design. A large body of work showed that avatar fidelity significantly impacts social interactions [152] and that more realistic avatars evoke stronger acceptance in virtual body ownership [274]. The gained benefits of using abstract avatars in applied VR research (for example, being cost-effective and more accessible to the broader USEC field) might not outweigh the advantages of more realistic avatars in settings where social interactions and virtual body ownership are of importance (for example, in social VR [431, 546]). However, abstract avatars already provide researchers with valuable insights into the observers' interaction identification performances and enable them to distinguish between different human gestural movements. The key message of the pre-security study is that there is excellent potential for abstract avatar designs in research where user observations are at the core of the exploration. For example, when aiming at investigating the participants' identification performance during observations of gestures performed by VR avatars, abstract user representations already provide valuable answers.

As a result of the promising results of abstract VR avatars in the pre-security study and to further contribute towards making the use of VR studies for USEC research more accessible (cf., the comments on using low-budget equipment in Section 4.6.5.1), the main security study this thesis will discuss in Section 4.8 will use an abstract avatar to represent a human and their interactions in a VR environment. The combination of both the pre-security study and the main security study in Section 4.8 will allow the thesis to comment on the use of VR studies for security evaluations of real-world prototypes.

4.8 Security Evaluation of RepliCueAuth

To provide, for the first time, insights into the use of VR studies for the security evaluation of real-world USEC prototypes, it is essential to change the environment variable, i.e., from traditional real-world video recordings as used in CueAuth [243] to video recordings based on a virtual environment, without introducing any potential confounding variables. Therefore, the main security study is, in line with the pre-security study and the original real-world study [245], based on two-dimensional (VR) recordings. VR setups were used to simulate and record authentications and, in line with CueAuth's real-world study [245], the participants experienced the recorded VR authentications on a traditional desktop screen. Chapter 5 will investigate the differences between non-immersive VR [147] observations, i.e., on a standard desktop screen, and immersive VR observations, i.e., in VR, in a follow-up investigation.

4.8.1 Methodology and Study Design

The security evaluation of CueAuth, summarised in Section 4.4.2, follows the real-world study design [245] with the input method as independent variable with three levels: touch input, mid-air input, and eye-gaze input, and the threat model with two levels: single attack and repeated-video attack. The difference between the VR investigation and the original real-world study is that the authentications are recorded from within VR instead of in the real world. This means that the authentications were performed by an abstract avatar instead of a human in the real world. As done in the real-world study [245], the participants' successful attack rates, the Levenshtein distance between the correct PIN and the participants' closest guess; the participants' attack durations; and the participants' confidences in their guesses were collected. Unlike data collected through the semi-structured interviews in the real-world study [245], the VR security study relied on open questions in written form. As done in the pre-security study, Qualtrics [397] and Prolific [394] were used to deploy the study online. The video material for the study is publicly available². The statistical analysis in Section 4.8.3 follows the approach this thesis described in Section 4.5. The order of the input methods was counterbalanced using the Qualtrics' Randomizer [398] and its embedded data function.

4.8.1.1 Threat Models

Two frequently used threat models were used when evaluating the security of CueAuth using VR video recordings: single attacks and repeated-video attacks. Both threat models are in line with the original real-world security study [245]. The threat models represent the scenario where a bystander observes another person's authentications. In both threat models, the participants know how the authentication prototype works and have an optimal view of the input. The type of the attack was alternating, as also done in the original real-world study [245], resulting in four attacks for each threat model and input method.

- **Single Attack:** The participants have only one chance to observe the authentication, a common threat model applied in authentication research when evaluating the security of USEC prototypes (for example, [102, 180, 530]).
- **Repeated-Video Attack:** The participants can watch a video recording of the authentication more than once if required, in line with prior works (for example, [102, 157, 308]). The participants can pause, rewind, slow down, and speed up the video.

²All VR videos are stored on <https://youtube.com/playlist?list=PLs1tzNuOyzfyDx7Y99JatsN42a5MEIxea>, last accessed 22/01/2023

4.8.1.2 Study Procedure and Task

The participants were introduced to the threat models and RepliCueAuth's input methods through explainer videos. Control questions after each introduction were added where the participants had to guess a single-digit entry based on a given picture of the authentication scheme and the VR avatar's interaction. This ensured that the participants understood how the input methods worked. The participants were navigated back to the explainer videos if they did not pass the control questions. Each block then displayed eight PINs, four for each threat model. No participant attacked the same PIN more than once. After the observations, the participants could provide up to three guesses and rate their confidence in their guess using a 5-point Likert scale from *Strongly Disagree* to *Strongly Agree*. The study concluded with a questionnaire about the participants' perceptions of the input methods regarding their security and usability. Finally, the participants were provided with a file upload where they could upload pictures or screenshots of any notes they took.

4.8.2 Demographics

Twenty-two (22) participants were recruited (10 female, 12 male). Six participants were excluded (vs two in the real-world study [245]) as it was clear they did not put reasonable effort into the attacks (0 out of 24 attacks were successful). This led to an average participant age of 24 years (range: 19 to 35, SD = 4.86). The sample is slightly more diverse regarding gender (50% female participants) compared to the sample in the real-world study (18.18% female participants) [245]. The participants' age in both studies is almost identical, with an average age of 24 years in the VR-based replication study and 26.9 years in the original real-world study [245].

4.8.3 Results

First, the participants' successful attack rates and the Levenshtein distances [280] between the correct PIN and their closest guesses are reported to provide a detailed security evaluation of CueAuth when replicated in VR. Then, the participants' attack duration and confidence, along with qualitative feedback, are reported.

4.8.3.1 Successful Attack Rate

In the single attack threat model, only a few attacks ($M = 10.95\%$, $SD = 18.19\%$) on touch input were successful. Not a single attack was successful against mid-air input and eye-gaze input. When attackers could rewind the videos, slightly more than half of the attacks on touch

input ($M = 59.38\%$, $SD = 22.12\%$) and mid-air input ($M = 59.38\%$, $SD = 27.20\%$) were successful, but not a single attack on eye-gaze input.

There was a significant effect of the threat model, $F_{(1,15)} = 125.952$, $p < 0.05$, $\eta_p^2 = 0.894$, and the input method, $F_{(1.340,20.097)} = 37.426$, $p < 0.05$, $\eta_p^2 = 0.714$, on the participants' attack rates. There was also a significant interaction effect (threat model \times input method), $F_{(2,30)} = 30.829$, $p < 0.05$, $\eta_p^2 = 0.673$. Whilst follow-up analysis revealed a significant difference of the attack rates between the input methods in the single threat model, $F_{(2,30)} = 5.787$, $p < 0.05$, $\eta_p^2 = 0.278$, post-hoc pairwise comparisons did not confirm these differences. In the repeated-video attack threat model, there was evidence of a significant difference of the attack rates between the input methods, $F_{(1.383,20.744)} = 41.336$, $p < 0.05$, $\eta_p^2 = 0.734$. Attacks on touch input ($M = 59.38\%$, $SD = 22.12\%$) and mid-air input ($M = 59.38\%$, $SD = 27.20$) were significantly more successful ($p < 0.05$) than attacks on eye-gaze input ($M = 0\%$, $SD = 0.0\%$). The same pairs were significantly different in the real-world study [245].

Validation

When comparing the VR study findings to the original real-world study, there was a statistically significant three-way interaction effect (study type \times input method \times threat model), $F_{(2,68)} = 3.226$, $p = 0.046$, $\eta_p^2 = 0.706$, but no statistically significant two-way interaction effect (study type \times input method), $F_{(2,68)} = 0.219$, $p = 0.804$, $\eta_p^2 = 0.006$; (study type \times threat model), $F_{(1,34)} = 2.542$, $p = 0.120$, $\eta_p^2 = 0.070$. In line with CueAuth's real-world evaluation [245], repeated-video attacks on touch input and mid-air input were more successful than on eye-gaze input. There was no evidence of significant differences between the input methods for single attacks in both the VR replication study and the real-world study. However, single attacks on touch were 10.95% successful in the VR replication study, whilst they were not successful at all in the real-world study [245]. The other was found in repeated-video attacks on touch input, with 59.38% successful attacks in the VR replication study and 74% in the original real-world study [245]. Attack rates on mid-air and eye-gaze input in the VR replication study match more accurately with the original study. Results are summarised in Table 4.2.

Security Observation 1

The successful attack rates against VR avatars are largely similar to attacks against a human in the real world, as done in CueAuth's original real-world study [245]. Similar patterns can be found in the VR study and the original real-world study, summarised and put in comparison between the two study types in Table 4.3 on page 130.

Table 4.2: Single attacks against mid-air and eye gaze resulted in no successful attacks. Only 10.95% of the observations were successful on touch input. Repeated-video attacks were equally successful for touch and mid-air input (59.38%), with the Levenshtein distance showing that attacks on touch input were closer to the correct PIN. The participants' confidence (1 = not confident at all, 5 = very confident) remained the same for eye gaze across the threat models. Their confidence increased in repeated-video attacks on touch and mid-air input.

	Single Attack				Repeated-Video Attack			
	Success	Distance	Confidence	Duration	Success	Distance	Confidence	Duration
<i>Touch</i>	10.95%	2.03	1.74	103.00 s	59.38%	0.59	4.34	150.98 s
<i>Mid-air</i>	0.00%	2.94	1.33	79.68 s	59.38%	0.83	4.53	138.80 s
<i>Eye Gaze</i>	0.00%	3.55	1.06	63.48 s	0.00%	3.45	1.09	150.54 s

4.8.3.2 Levenshtein Distance

There was a significant effect of the threat model, $F_{(1,15)} = 88.679$, $p < 0.05$, $\eta_p^2 = 0.855$ and input method, $F_{(2,30)} = 170.284$, $p < 0.05$, $\eta_p^2 = 0.919$ on the Levenshtein distance. There was also evidence of a significant interaction effect (threat model \times input method), $F_{(2,30)} = 46.126$, $p < 0.05$, $\eta_p^2 = 0.755$. Follow-up analysis on the level of each threat model revealed a significant effect of input method on the Levenshtein distance in case of single attacks, $F_{(2,30)} = 30.551$, $p < 0.05$, $\eta_p^2 = 0.671$. There were significant differences between all three input methods ($p < 0.05$). Between touch input ($M = 2.03$, $SD = 0.77$) and eye-gaze input ($M = 3.55$, $SD = 0.32$), touch input and mid-air input ($M = 2.94$, $SD = 0.60$), and eye-gaze input and mid-air input. There were also significant effects of the input method on the Levenshtein distances in case of repeated-video attacks, $F_{(2,30)} = 335.889$, $p < 0.05$, $\eta_p^2 = 0.957$. There was a significant difference between touch input ($M = 0.59$, $SD = 0.29$) and eye-gaze input ($M = 3.45$, $SD = 0.31$), and mid-air input ($M = 0.83$, $SD = 0.46$) and eye-gaze input. Repeated-video attacks did not improve attacks on eye-gaze input, but when attacking touch input and mid-air input ($p < 0.05$). There was no evidence of a significant difference between the participants' observation performance in single and repeated-video attacks when observing eye gaze $F_{(1,15)} = 0.652$, $p = 0.432$, $\eta_p^2 = 0.042$. For touch input, the participants' performance differed significantly between the threat models, $F_{(1,15)} = 58.778$, $p < 0.05$, $\eta_p^2 = 0.797$. Repeated-video attacks were significantly more accurate ($M = 0.06$, $SD = 0.29$) than single attacks ($M = 2.03$, $SD = 0.77$). The same was found for mid-air input, $F_{(1,15)} = 35.596$, $p < 0.05$, $\eta_p^2 = 0.864$, with repeated-video attacks being significantly more accurate ($M = 0.83$, $SD = 0.46$) than single attacks ($M = 2.94$, $SD = 0.60$).

Validation

When comparing the Levenshtein distances in the VR study to the real-world study, there was evidence of a significant three-way interaction effect (study type \times input method \times threat model), $F_{(2,68)} = 5.319$, $p < 0.05$, $\eta_p^2 = 0.135$, and a significant two-way interaction (method \times study type), $F_{2,68} = 15.959$, $p < 0.05$, $\eta_p^2 = 0.319$. There was no evidence of a significant

two-way interaction (threat model \times study type), $F_{(1,34)} = 3.823$, $p = 0.059$, $\eta_p^2 = 0.101$. A more nuanced analysis on the level of each threat model and input method revealed that single attacks on eye-gaze input in the real-world study were statistically significant closer ($p < 0.05$) to the correct PINs ($M = 2.81$, $SD = 0.76$) than in the VR replication study ($M = 3.55$, $SD = 0.32$), $F_{(1,34)} = 12.898$, $p < 0.05$, $\eta_p^2 = 0.275$. The other was found for touch input, with single attacks resulting in significantly more accurate observations to the correct PINs in VR ($M = 2.03$, $SD = 0.77$) than in the real world ($M = 2.83$, $SD = 0.67$), $F_{(1,34)} = 10.883$, $p < 0.05$, $\eta_p^2 = 0.242$. The difference between the observations on mid-air input in VR and in reality was not significant, $F_{(1,34)} = 0.573$, $p = 0.454$, $\eta_p^2 = 0.017$. The values for single attacks on mid-air input were $M = 2.94$ ($SD = 0.60$) for VR and $M = 2.78$ ($SD = 0.67$) for the real world.

In repeated-video attacks, the guesses on eye-gaze input were more accurate in VR than in the real world, $F_{(1,34)} = 18.277$, $p < 0.05$, $\eta_p^2 = 0.350$. The value for the VR study is $M = 2.63$ ($SD = 0.72$) and for the original real-world study $M = 3.45$ ($SD = 0.31$). There were no significant differences for touch input, $F_{(1,34)} = 0.361$, $p = 0.552$, $\eta_p^2 = 0.011$, and for mid-air input, $F_{(1,34)} = 0.996$, $p = 0.325$, $\eta_p^2 = 0.028$, between the study types. For touch input, the value in the VR study was $M = 0.594$ ($SD = 0.287$) and $M = 0.500$ ($SD = 0.568$) for the real-world study. For mid-air input, the values were $M = 0.828$ ($SD = 0.569$) and $M = 0.638$ ($SD = 0.641$).

The reported differences in the VR replication study in the single-attack threat model were not in line with the original real-world study. In contrast to the real-world study [245], the VR study revealed significant differences in the input methods on the Levenshtein distances in the single-attack threat model. However, in repeated-video attacks, attacks on touch input and mid-air input were significantly closer to the correct PINs than on eye-gaze input, which aligns with the real-world study. Both the VR and the real-world study [245] suggest that repeated-video attacks improved the participants' performance in all three input methods, although only slightly when observing eye-gaze input.

Security Observation 2

The attacks on mid-air input were equally close in both study types, whereas attacks on eye-gaze input were more accurate in the real-world study. In the VR replication study, the participants performed significantly closer single attacks on touch input than in the real-world study. In both the VR study and the real-world study, the repeated-video attacks were more successful on touch input and mid-air input than on eye-gaze input.

4.8.3.3 Attack Duration

As done in the original real-world study, attack durations were analysed for repeated-video attacks as participants could play, pause, and rewind the authentications as often as they wished.

There was no effect of the input method on the participants' attack duration, $F_{(1.241,16.130)} = 0.115$, $p = 0.792$, $\eta_p^2 = 0.009$. The attack durations on touch, mid-air, and eye-gaze input were $M = 150.98$ ($SD = 65.83$), $M = 138.80$ ($SD = 44.68$), and $M = 150.54$ ($SD = 33.19$).

Validation

When comparing the attack durations in the VR study to the original real-world study, there was no evidence of a significant interaction effect (study type \times input method), $F_{(1.296,44.049)} = 1.938$, $p = 0.168$, $\eta_p^2 = 0.054$. In both studies, attacks on mid-air input were the fastest. The duration for observing mid-air input in the real world was $M = 91.09$ ($SD = 44.68$), $M = 163.41$ ($SD = 114.17$) for eye-gaze input, and $M = 103.91$ ($SD = 42.27$) for touch input [245]. Attacks on mid-air input were significantly faster than on eye-gaze input in the real-world study, which was not the case in the VR study. All other pairs match – there were no significant differences between touch input and mid-air input, and between touch input and eye-gaze input in the real-world study of CueAuth [245] and in the VR study.

Security Observation 3

There was no evidence that the participants in the VR study spent more or less time on their attacks than in the original real-world study. In both studies, there was no significant difference in the attack duration between touch and mid-air input, and between touch and eye-gaze input. However, attacks on touch and mid-air input took slightly longer in the VR replication study than in the real-world study.

4.8.3.4 Participants' Confidence in Their Attacks

There was a significant effect of the threat model, $F_{1,15} = 252.842$, $p < 0.05$, $\eta_p^2 = 0.944$, and the input method, $F_{2,30} = 284.938$, $p < 0.05$, $\eta_p^2 = 0.950$, on participants' level of confidence. There was also evidence of a significant interaction effect (threat model \times input method), $F_{(2,30)} = 147.413$, $p < 0.05$, $\eta_p^2 = 0.908$. Follow-up analysis revealed a significant main effect of input method on the participants' confidence when performing single attacks, $F_{(2,30)} = 15.838$, $p < 0.05$, $\eta_p^2 = 0.514$, and repeated-video attacks, $F_{(2,30)} = 341.548$, $p < 0.05$, $\eta_p^2 = 0.958$. In terms of single attacks, there were significant differences ($p < 0.05$) between touch input ($M = 1.74$, $SD = 0.48$) and mid-air input ($M = 1.33$, $SD = 0.35$), and between touch input and eye-gaze input ($M = 1.06$, $SD = 0.14$). For repeated-video attacks, attackers were significantly less confident about their guesses when PINs were entered with eye-gaze input ($M = 1.09$, $SD = 0.27$) compared to touch input ($M = 4.34$, $SD = 0.74$) and mid-air input ($M = 4.53$, $SD = 0.54$). There was a significant difference in the participants' confidence on touch input, $F_{(1,15)} = 158.627$, $p < 0.05$, $\eta_p^2 = 0.914$, and on mid-air input, $F_{(1,15)} = 284.594$, $p < 0.05$, $\eta_p^2 = 0.950$, between the threat models. There was no evidence of a significant difference

between the threat models when observing eye gaze, $F_{(1,15)} = 0.211$, $p = 0.652$, $\eta_p^2 = 0.914$. The participants' confidence was significantly higher ($p < 0.05$) in repeated-video attacks in case of touch input ($M = 4.34$, $SD = 0.74$) and mid-air input ($M = 4.53$, $SD = 0.54$) than in single attacks on touch input ($M = 1.74$, $SD = 0.48$) and on mid-air input ($M = 1.33$, $SD = 0.35$). Results are summarised in Table 4.2.

Validation

When comparing the participants' confidence between the two study types, there was evidence of a significant three-way interaction effect (input method \times threat model \times study type), $F_{(1,399,47,582)} = 10.485$, $p < 0.05$, $\eta_p^2 = 0.236$. There was also a statistically significant interaction effect (input method \times study type), $F_{(1,620,55,076)} = 15.403$, $p < 0.05$, $\eta_p^2 = 0.31$, but no evidence of a significant interaction effect (threat model \times study type), $F_{(1,34)} = 3.727$, $p = 0.062$, $\eta_p^2 = 0.099$. Follow-up analysis revealed no significant main effect of input method on the participants' confidence when observing touch in the single attack threat model, $F_{(1,34)} = 0.237$, $p = 0.630$, $\eta_p^2 = 0.007$. For single attacks on touch input, the participants' confidence was $M = 1.89$ ($SD = 1.19$) for the real-world study and $M = 1.73$ ($SD = 0.47$) for the VR study. Participants were significantly more confident in their single attacks on mid-air input in the real-world study than in the VR study, $F_{(1,34)} = 6.963$, $p < 0.05$, $\eta_p^2 = 0.170$. The same was found for eye gaze, $F_{(1,34)} = 10.732$, $p < 0.05$, $\eta_p^2 = 0.240$. The values for mid-air input were $M = 2.03$ ($SD = 1.01$) for the real world and $M = 1.33$ ($SD = 0.35$) for VR. For eye gaze, the values were $M = 1.93$ ($SD = 1.04$) and $M = 1.06$ ($SD = 0.144$), respectively.

Repeated-video attacks on mid-air input and on eye-gaze input resulted in significant different levels of confidence between the study types (mid-air: $F_{(1,34)} = 6.72$, $p < 0.05$, $\eta_p^2 = 0.165$, eye gaze: $F_{(1,34)} = 16.159$, $p < 0.05$, $\eta_p^2 = 0.322$). Participants were significantly less confident when observing mid-air input using real-world recordings ($M = 3.71$, $SD = 1.16$) than observing mid-air input based on VR recordings ($M = 4.53$, $SD = 0.54$). The opposite was found for eye gaze, with participants having been more confident in the real-world study $M = 2.28$ ($SD = 1.15$) than in the VR study $M = 1.09$ ($SD = 0.27$). There was no evidence of a significant difference for touch input in the repeated-video attack threat model, $F_{(1,34)} = 3.727$, $p = 0.062$, $\eta_p^2 = 0.099$. Participants were slightly more confident in the VR study ($M = 4.34$, $SD = 0.74$) than in the real-world study ($M = 3.75$, $SD = 1.09$). Results are summarised in Table 4.3.

In line with the real-world study, the participants were more confident in their repeated-video attacks on touch input and mid-air input than in their single attacks on touch input and mid-air input. Additionally, they were significantly more confident in their repeated-video attacks on touch input and mid-air input than in their repeated-video attacks on eye gaze, which was the case for the VR replication study and CueAuth's original real-world study [245].

Security Observation 4

There was no evidence that participants were more confident in attacking touch input in either one of the study types. However, they were more confident in their single attacks on mid-air input and eye-gaze input in the real-world study than in VR. Participants were more confident when performing repeated-video attacks on eye gaze in the real world compared to the VR study but less confident when attacking mid-air input.

4.8.3.5 Qualitative Feedback

Unlike the data collected through semi-structured interviews in the real-world study [245], this thesis relied on open questions at the end of the online study. Three main areas of interest were defined to learn more about the participants' (1) observation strategies, (2) perception of the usability of the input methods, and (3) perception of the security of the input methods.

Theme 1: Participants' Observation Strategies. The participants' attacking strategies in the VR replication study aligned with those in the original real-world study [245] to a great extent. Whilst the majority of the participants mostly noted down the PIN numbers on a piece of paper, some others sketched the authentication system or used software tools such as Excel spreadsheets to support their observations (cf., Figure 4.9). They raised that further training could help them in running successful attacks. Single-view observations on touch input and on mid-air input were perceived as too fast. Participants found it challenging to switch between the hand movements and the digits on the screen. Observations on eye-gaze input were perceived as too challenging. The participants mentioned they could hardly see the eyes move. Instead, they guessed the direction of the avatar's eyes to indicate on which

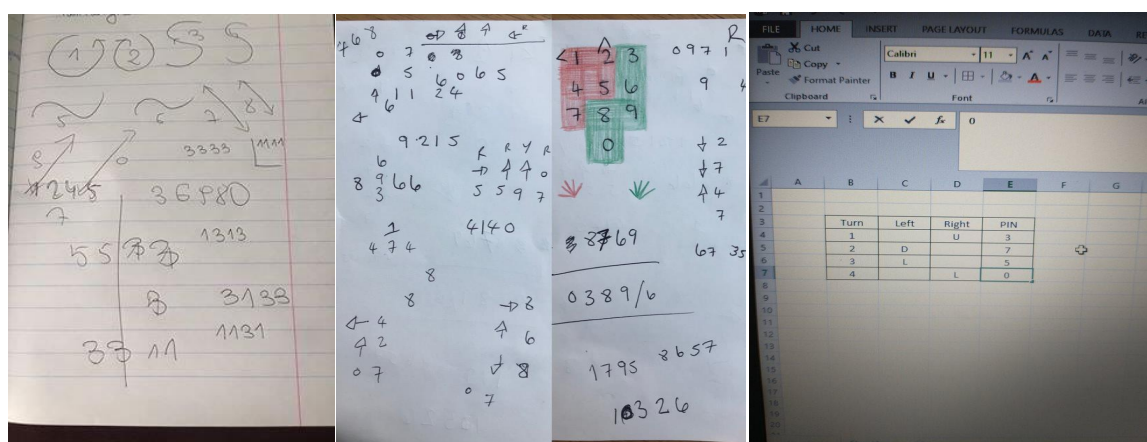


Figure 4.9: The participants reported to make use of different strategies to support their observations. The majority wrote down numbers on a piece of papers, whilst others drew a sketch of CueAuth or made use of Excel spreadsheets.

digit the user gazed at. Others mentioned that the combination of focusing on the avatar's eye movements and the screen put them off. In repeated-video attacks, the participants made use of rewinding and slowing down the videos. This was mentioned frequently in touch input and mid-air input, but not when observing eye-gaze input. In mid-air input, the participants mentioned that observing one-handed interactions was easier. Similar to single attacks on eye-gaze input, the participants perceived input using eye gaze as "hard" to attack and mentioned that slowing down the videos did not help them to recognise eye movements. Some participants reported having a vague idea of the entered digits, but could not use this information to provide successful attacks on eye-gaze input, which is in line with the successful attack rate (Security Observation 1).

Theme 2: Participants' Security Perception. When asked which input method the participants perceive as most secure, all reported that they perceive eye-gaze input as most secure, which is in line with the findings reported in the VR usability study in Section 4.6.4.4 and the original real-world study [245]. Similar to the real-world study, opinions differed when asking about the least secure method: eleven participants found touch the least secure input method, whilst five participants found mid-air the least secure input method. This slightly differs from the real-world study [245], where the participants believed that touch input is safer than mid-air input.

Theme 3: Participants' Usability Perception. The feedback received from the participants in the security study aligns with the findings reported in the VR usability study in Section 4.6.4.4 and the real-world study. Input using touch was defined as easy to use, convenient, and practical. P4 and P6 mentioned that entering a PIN can be done discreetly by covering up the hands. However, P16 mentioned that the on-screen gestures could linger on the screen after input, i.e., smudge attacks [28]. Others brought up that touch input is easy to attack and that people are already used to attack this input method. Additional comments were on the methods' hygiene and social acceptability. For example, P4 raised the concern about the infection risk in a post-pandemic world where people must touch surfaces. Some participants mentioned that providing mid-air input feels like being a fool in public and looks strange. Overall, the participants were reserved towards the social acceptability of mid-air input. P11 and P12 stated that mid-air input reminds them of playing games on a Nintendo Wii or in VR. P14 mentioned that mid-air input feels like sharing the PIN with everyone around. P12 voiced that using mid-air gestures over-complicates input a lot. The participants were reserved towards the usability of eye-gaze input. Providing input with eye gaze was considered hard to use, hard to learn, and long-winded. P5 mentioned that the concept feels weird. Furthermore, the participants noted that eye-gaze input could lead to many errors, prompting users to go through the authentication process several times. On the positive side,

P4 emphasised that eye-gaze input could support disabled people in their interaction.

Security Observation 5

Many of the mentioned comments in the VR security study match those voiced in the VR usability study (cf., Section 4.6) and the real-world study [245]. This suggests that the participants' usability and security perception of the input methods remained the same across the studies and was not influenced by the avatar and VR.

4.8.4 Discussion of RepliCueAuth's VR Security Evaluation

In contrast to CueAuth's real-world security study, where the participants watched real-world video recordings [245], the participants in the VR replication study watched VR recordings of user authentications that took place in a virtual environment using an abstract VR avatar. Additionally, the participant recruitment in the VR study took place on Prolific [394]. This is conceptually different to the original real-world security study, which recruited participants for a lab study [245]. Voit et al. [528] observed that the participants in their online method were less engaged than in studies where a researcher was present. The poor performance of six participants in the VR security study of CueAuth suggests that this phenomenon [88] was present here too. In the VR security study, the same number of participants ($N = 22$) as in CueAuth's original real-world security study was recruited. Whilst Khamis et al. [245] excluded two participants due to their poor performance (0 out of 24 attacks were successful), six participants had to be excluded in the VR security study for the same reason. There were repetitive guesses throughout the survey, for example, "1234", and repeatedly wrongly answered control questions, indicating that some participants did not participate meaningfully in the study. This suggests that the 22 participants in CueAuth's real-world study [245] felt more committed to their participation than the ones in the study of this thesis. The VR security study emphasises the importance of filtering out low-quality responses in security studies and extends the findings by Redmiles et al. [409] and Fahl et al. [133] who argued that the ecological validity can be improved by filtering out such cases.

After excluding the six participants, the VR security study results match with the findings from the real-world evaluation to a great extent. There was no evidence that the successful attack rates differed significantly between the two study types (Security Observation 1) and the performance of attacks against the different input methods followed the same pattern across both studies with similar significant differences in VR and the real world. However, whilst more general measures such as the participants' successful attack rate matched between the study types (Security Observation 1), more specific measures, such as the Levenshtein distances, differed between the studies (Security Observation 2).

To summarise, the results of the VR-based security study show that observing VR avatars during authentications reveals insights into the security of USEC prototypes, providing the first understanding of the use of VR studies for simulating real-world security research on VR replicas of real-world prototypes. The findings presented in this section, together with the VR usability study in Section 4.6, will contribute to the overarching research question of this thesis, RQ₆. RQ₃, which asked which findings of VR-based security evaluations on USEC prototypes match the conclusions from corresponding evaluations in traditional physical lab settings, will be answered in Section 4.10.2. Before contributing answers to the research questions RQ₂ and RQ₃, some limitations of the research in this chapter are discussed.

4.9 Limitations

It is important to consider the following technological and experimental design limitations when interpreting the findings of and concluding this chapter. First, a VR-based usability and security evaluation of a real-world USEC prototype, CueAuth [245], was reported. Whilst replicating CueAuth allowed studying the usability and the security of a breadth of input methods that are frequently used in security research (for example, touch: [102, 540], mid-air: [6, 25], and eye gaze: [104, 235, 272]), future research, for example, additional replication studies, are required to generalise the results to the broader USEC research field.

Second, the security evaluation is based on two specific threat models: 1) single observation attacks, and 2) repeated-video attacks. In the pre-security study in Section 4.7, the replication security study in Section 4.8, and the real-world security study [245] the observations were single-person attacks through optimal views on the authentication scheme. Using non-optimal user-defined views or more advanced threat models, for example, multiple observers [240], may result in different findings. However, as this thesis is the first work that contributes a validation of the use of VR studies for simulating real-world USEC research and aims at contributing a robust baseline exploration for follow-up investigations, it was essential to incorporate commonly used threat models that have found widespread adoption when assessing USEC prototypes (for example, [102, 108, 239, 243]).

Third, following Khamis et al.'s study design means facing the same limitations [245]. CueAuth's evaluation depended on the hardware used in the real world and the VR studies. Other hardware, such as OptiTrack systems, may lead to more accurate hand and finger tracking. However, if (remote) VR studies are to become mainstream (cf., chapter 6), they must utilise personal commodity hardware that a typical VR user would own. It is unlikely that average VR users will own a high-end tracking system like OptiTrack.

Finally, replication studies are generally challenging to conduct [550]. One of the largest replication studies attempted to replicate 100 studies and succeeded only in 39% of them

[31, 80]. Human test subjects consciously or subconsciously remember previous experiences impacting their thoughts, behaviour, and performance. Thus, experiments can result in different results due to the non-uniformity of nature [184, 447]. The VR equipment used for the studies must be noted and considered for future replication studies. Advancements in technology, specifically in VR/AR hardware and sensing capabilities, are already happening.

4.10 Chapter Conclusion

This chapter investigated, for the first time, the use of VR studies for usability and security evaluations on VR replicas of real-world USEC prototypes through three experiments. Table 4.3 summarises the main findings of the VR usability and security study and puts those in contrast to the real-world study results of CueAuth [245]. The user study in Section 4.6 focused on the usability evaluation of a virtual replica of a real-world USEC prototype. It evaluated the usability of three input methods, i.e., touch, mid-air, and eye-gaze input, in VR and investigated how the findings transfer to the real world. The participants achieved similar entry accuracies; reported similar perceived workload when authenticating using touch input, mid-air input, and eye-gaze input; and shared similar security perceptions of the input methods across the studies. However, longer input entries were found when using touch input in VR. This finding indicates that introducing virtual artefacts, such as hand and finger tracking, negatively impacts the participants' performance when interacting with USEC prototypes in VR instead of in reality. In contrast to this finding, wearable eye trackers as used in the VR usability study (cf., Section 4.6), instead of static eye trackers as used in the original real-world study [245], have a positive impact on the participants' performance, which can be misleading and may overestimate a USEC prototype's usability.

Through the pre-security study in Section 4.7 and the security study in Section 4.8, many similarities between observations performed on a virtual avatar and on a human in the real world were observed. Attack rates did not significantly differ between the study types and there was no evidence that the participants spent more or less time on their attacks in the VR security study compared to the real-world study. However, measures like the accuracy of the participants' guesses and their confidence differed significantly between the two study types, with the participants in the real-world study being closer to the correct guesses on eye-gaze input, but not when performing single-view attacks on touch input. Overall, the participants' perception of the usability and security of the input methods in both the VR usability study in Section 4.6 and the VR security study in 4.8 match to a great extent the perception of the real-world study participants (Usability Observation 4 and Security Observation 5). The validation analysis revealed many similarities between the quantitative measures of the conceptual replication studies using VR and CueAuth's real-world counterparts [245]. Similar

Table 4.3: The VR study achieved similar results regarding the participants' entry accuracy and perceived workload when interacting with CueAuth compared to the real-world study [245]. There was no evidence that the participants' attack rate and duration differed significantly between the study types. However, the two study types significantly differed in measures such as entry time, the distance of the participants' guesses to the correct PINs, and their confidence. * The reported values are for both single and repeated-video attacks with the notation *single* | *repeated*. **Bold** highlights best usability and security.

Usability Measures				Security Measures			
		Real-World Study [245]	VR Study			Real-World Study [245]	VR Study
Entry Accuracy	Touch	93.38%	89.97%	Attack Rate*	Touch	0.00% 74%	10.95% 59.38%
	Mid-Air	84.19%	80.42%		Mid-Air	0.01% 64%	0.00% 59.38%
	Eye Gaze	82.72%	83.75%		Eye Gaze	0.03% 0.05%	0.00% 0.00%
Entry Time	Touch	3.73 s	6.06 s	Levenshtein Distance*	Touch	2.83 0.50	2.03 0.59
	Mid-Air	5.51 s	5.54 s		Mid-Air	2.78 0.64	2.94 0.83
	Eye Gaze	26.35 s	16.75 s		Eye Gaze	2.81 2.63	3.55 3.45
Perceived Workload	Touch	23.25	34.83	Attack Duration*	Touch	N/A 103.9 s	103.00 s 150.98 s
	Mid-Air	39.584	34		Mid-Air	N/A 91.9 s	79.68 s 138.80 s
	Eye Gaze	46.54	30.33		Eye Gaze	N/A 163.4 s	63.48 s 150.54 s
Preferred Input Method	Touch	56	37	Attacker Confidence*	Touch	1.89 3.75	1.74 4.34
	Mid-Air	33	41		Mid-Air	2.03 3.71	1.33 4.53
	Eye Gaze	31	45		Eye Gaze	1.93 2.28	1.06 1.09

to the design implications in the real-world study [245], this thesis can deduce the following implications from the VR usability and security studies:

- **Design Implication 1:** Eye gaze is the most secure input method, but the slowest (cf., Table 4.3). This suggests that eye gaze is suitable when authentication frequency is low and subtle authentication is required, which aligns with the design implication 1 reported in the real-world study [245].
- **Design Implication 2:** When comparing eye-gaze with mid-air input in the VR usability study and VR security study, this thesis concludes that mid-air input is more usable than eye-gaze input, but eye-gaze input is more secure (cf., design implication 2 [245]).
- **Design Implication 3:** The qualitative feedback in the VR studies suggest that mid-air input is not suitable for public spaces as it requires additional space and “*looks like a jumping jack*” (P17). Therefore, novel authentication methods should aim for providing alternative input modalities if wished by end-users. This finding is also depicted in [245]’s design implication 3.

In the following, RQ₂ and RQ₃ are answered based on the research conducted in this chapter.

4.10.1 Research Question 2 (RQ₂)

The usability study in Section 4.6 contributes answers to the following research question:

RQ₂: Which findings of VR-based usability evaluations on USEC prototypes match the findings from corresponding evaluations in traditional physical lab settings?

To answer RQ₂, it was necessary to investigate how the results of a user study in VR match with the results obtained in a real-world study. In a conceptual replication user study, several common usability metrics (cf., Table 4.3) were compared between a VR usability study and the equivalent original real-world user study [245]. The results of the conceptual replication suggest that using VR studies to evaluate the usability of real-world authentication prototypes is possible. However, the use of VR is not an “all-in-one” solution, and results from VR studies may differ from results collected in the real world. For example, in contrast to the real-world study, eye-gaze input was preferred over mid-air input and touch input in the VR study. This aligns with the participants’ entry time when providing eye-gaze input and touch input. Authentications using eye gaze were significantly faster in VR than in the real world, but touch input was slower in VR than in the real world. However, the comparison between the VR study results and the original-real world study did not reveal significant differences between the participants’ perceived workload. The comments made by the participants during the interviews were mostly in line with the qualitative feedback in the original real-world study. For example, the participants were concerned about performing mid-air gestures in public in both the VR study and real-world study. Table 4.3 provides an overview of the usability and security metrics of CueAuth collected in the real world and VR.

4.10.2 Research Question 3 (RQ₃)

The pre-security study in this thesis chapter (Section 4.7) has informed the security study reported in Section 4.8. Both studies contribute answers to the following research question:

RQ₃: Which findings of VR-based security evaluations on USEC prototypes match the findings from corresponding evaluations in traditional physical lab settings?

As Section 4.7 has shown, interactions in VR are already distinguishable when using an abstract avatar design. Whilst abstract avatars are sufficient when assessing a USEC prototype’s resistance against observations, there are many situations in which more human-like avatars are to be preferred. This thesis aimed to keep the avatar design as simplistic as possible to make the use of VR studies for USEC prototyping research more broadly accessible and cost-effective whilst still allowing USEC researchers to evaluate their prototype’s security. Given that an abstract avatar design was informative enough to simulate touch input, mid-air input, and eye-gaze input in virtual environments, the next question was how well the results

of a VR-based security evaluation of a USEC prototype match the results of an equivalent evaluation in the real world.

This chapter's security evaluation in Section 4.8 suggests that the number of successful attacks against a VR avatar is largely similar to the number of successful attacks against a human in the real world. In both study types, the repeated-video attacks resulted in more successful attacks on touch input and mid-air input compared to the single attacks. Furthermore, the threat model (single attack vs repeated-video attack) did not impact the attack rate on eye-gaze input in both study types. Similar patterns were found for the Levenshtein distances and the participants' confidence between the VR study and the original real-world study [245]. The participants' confidence and their low success rates when observing eye-gaze input align with prior work that emphasised the high observation resistance of gaze-based input (for example, [156, 235, 272, 308]).

All in all, the VR-based security evaluation of CueAuth achieved promising results, with many similarities to the real-world study (cf., Section 4.10 and Table 4.3). The findings of the main security study in Section 4.8 confirm that observations on an abstract avatar provide USEC researchers with an accurate assessment of their prototype's security against observations. Furthermore, the results endorse the pre-security study findings in Section 4.7 regarding the distinguishability of different input methods and provide researchers with insights into the participants' perceptions of the input methods' usability and security.

4.10.3 Ways Forward

Through this chapter's first investigation of using VR studies for the implementation and evaluation of VR replicas of real-world USEC prototypes, this thesis has opened the door for follow-up research that reinforces the validation of using VR for USEC research and advances the more traditional USEC research methods. In the next chapter, this thesis further contributes to the validation of using VR studies for USEC research to allow the thesis to comment on how well VR-based usability and security explorations match equivalent investigations in real life and through the means of more traditional lab-based research methods. Contributing additional validation studies is important to further strengthen the argument of using VR studies for simulating real-world USEC research and complementing the existing, well-established lab studies and field studies. Additionally to the validation, the next chapter contributes, for the first time, to the realism of shoulder surfing evaluations in the lab using VR studies. Furthermore, it investigates, for the first time, how the use of VR studies allows researchers to simulate real-world scenarios in which USEC systems are eventually deployed and used by end-users, potentially contributing to ecological validity [260].

4.10.4 Contributions

In summary, the research in this chapter makes the following *empirical* as well as *methodological contributions* [556]:

- It investigates, for the first time, the use of VR studies for usability and security evaluations on VR replicas of real-world USEC prototypes. By doing so, it provides new empirical knowledge about the use of VR studies for real-world authentication research and may influence how the human-centred usable security community does science, forming an *empirical contribution* as well as a *methodological contribution* [556].
- It complements prior work that evaluated VR replicas of real-world artefacts by:
 - a) the first VR usability evaluation and online security evaluation using VR recordings of a real-world USEC prototype, and
 - b) the first validation of both usability study and security study findings through comparisons to their equivalent real-world studies.
- Finally, the research in this chapter is the first work that augments the design, implementation, and evaluation arsenal of USEC prototypes using VR and lays out a strong foundation for future research, some of which is pursued in chapter 5 and chapter 6.



USING VR STUDIES TO ADVANCE USEC RESEARCH

Chapter 5

Using VR Studies to Advance USEC Research

This chapter is based on the following two publications:

[Publication 6] Mathis, F., O’Hagan, J., Khamis, M., & Vaniea, K. (2022). Virtual Reality Observations: Using Virtual Reality to Augment Lab-Based Shoulder Surfing Research. In Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR 2022). IEEE, DOI: [10.1109/VR51125.2022.00048](https://doi.org/10.1109/VR51125.2022.00048) (Best Paper Award Nominee)

[Publication 7] Mathis, F., Vaniea, K., & Khamis, M. (2022). Can I Borrow Your ATM? Using Virtual Reality for (Simulated) In Situ Authentication Research. In Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR 2022). IEEE, DOI: [10.1109/VR51125.2022.00049](https://doi.org/10.1109/VR51125.2022.00049)

5.1 Introduction

VR studies allow simulating real-world contexts and conducting human-centred research on virtual replicas of real-world research artefacts, validated in chapter 4 for usability and security evaluations on VR replicas of USEC prototypes, and discussed in Section 2.2.5 of the Literature Review for the broader HCI field. This chapter focuses on forging forward research in the USEC research field through in-depth user studies on two core topics in human-centred authentication: *shoulder surfing research* and *in situ research*. As this chapter will show, the USEC community benefits from the use of VR studies in multiple ways: First, VR studies do not constrain USEC research to physical lab environments and available resources. The participants of user studies can be immersed into different scenarios and environments

with little effort, facilitating USEC research in multiple – often also hard-to-reach – security-sensitive environments. Second, characteristics that are typical in the wild, for example, additional bystanders and ambient noise, can be integrated into VR study environments. Therefore, VR studies allow researchers to simulate in situ research for accurate usability and security testing of USEC prototypes. Finally, using virtual replicas instead of real-world implementations facilitates replication studies and cross-country evaluations. Future VR-based user studies are not necessarily bound to physical locations, potentially contributing to large-scale usability and security testing of USEC research artefacts.

This chapter contributes additional VR studies to strengthen further and broaden the validation of using VR studies for real-world USEC research. Whilst chapter 4 has laid out promising results when replicating a real-world lab study in VR, this chapter shows how the use of VR studies contributes to closer-to-reality USEC research on research artefacts. This will be achieved by transforming real-world lab environments into VR replicas of plausible public settings in which USEC systems are typically used. Each study in this chapter involves established baselines from the literature against which the VR results are compared. For the *shoulder surfing* research in Section 5.3, a comparison of two novel shoulder surfing research methods against traditional 2D recordings is performed. For the *VR-simulated in situ authentication study* in Section 5.4, the results from simulated in situ settings in the real world and VR are compared to traditional investigations in both a real-world laboratory environment and a virtual laboratory environment.

All in all, both studies in this chapter aim to provide researchers with a bridge over the methodological gap between lab and field studies when 1) conducting usability and security evaluations on USEC prototypes and 2) evaluating USEC prototypes in their intended use case. Equivalent evaluations are challenging and often infeasible to achieve with existing, more traditional research methods due to the constraints of physical lab environments and the challenges USEC researchers face when going into the wild, as reviewed in the Literature Review and reported by the USEC experts in chapter 3. Therefore, it is important to understand better the use of VR studies for USEC research and how their use advances the broader USEC field. The lack of knowledge about how VR studies advance USEC research motivated the fourth research question of this thesis:

RQ₄ Can substitutional in situ studies using VR provide a bridge over the methodological gap between lab and field studies?

If, and only if, the use of VR studies can forge forward USEC research concerned with prototypes, then USEC researchers can justify their use for human-centred usability and security research instead of relying on more traditional laboratory studies. If there is no added value to its use, then the contributions of VR studies to the USEC field remain unclear.

Therefore, this thesis chapter will provide insights into the advantages of VR studies for USEC research on prototypes. Additionally to answering RQ₄, the findings of this chapter will contribute partial answers to the overarching research question, RQ₆, by outlining the advantages and disadvantages of using VR studies for USEC research compared to more traditional research in the lab. For example, the use of VR studies enables researchers to learn more about the participants' observation behaviour when aiming at guessing another person's PIN input (cf., Section 5.3). VR studies also provide researchers with a research method that allows simulating and conducting research in security-sensitive real-world scenarios (cf., Section 5.4.2.3), which is often infeasible in the real world due to financial, ethical, and legal constraints, as voiced by the USEC experts in chapter 3.

5.1.1 Chapter Structure

Section 5.3, the first user study of this chapter, describes the use of VR studies for shoulder surfing research. It first introduces the topic and then the authentication scenarios and the apparatus used in Section 5.3.2. It then outlines the methodology in Section 5.3.3 and reports and discusses the results in Section 5.3.5 and Section 5.3.6, together with some study-specific limitations in Section 5.3.7. Section 5.3 concludes by contributing answers to RQ₄, which are synthesised into lessons learned in Section 5.5.

Section 5.4, the second user study of this chapter, showcases and validates the use of VR studies for simulating in situ authentication research. It first introduces the current evaluation challenges when going into the field in Section 5.4.1 and then discusses the study methodology in Section 5.4.2. It then describes the hardware and the implementation in Section 5.4.3 and outlines the results of the user study in Section 5.4.5. The second user study concludes with a discussion of the results in Section 5.4.6, with some study-specific limitations in Section 5.4.7, and eventually contributes answers to RQ₄ in Section 5.4.8.

The fifth chapter of this thesis concludes with an overall discussion of the findings in Section 5.5 and outlines the ways forward to motivate the next chapter.

5.2 Ethics and Compensation

The research this thesis will discuss in Section 5.3 has been reviewed and approved by the University of Glasgow College of Science & Engineering ethics committee (*ref: #300200034*). Note that both studies in Section 5.3 and Section 5.4 were conducted in Austria due to the COVID-19 restrictions. Participants were paid €15 (€10/h) and participated in a lottery to win an additional €15. They were informed in advance of the study that the chances of winning increase with the number of successfully observed PINs/patterns. Providing a

lottery to win an additional €15 contributes to motivating participants to perform well in their shoulder surfing tasks [245, 308].

For the research in Section 5.4, the participants were paid according to the local standard (€10/h). The study has been reviewed and approved by the University of Glasgow College of Science & Engineering ethics committee (*ref: #300200295*). In this study, a single-blinded research approach was applied to ensure that the results better reflect real-world behaviour when interacting with USEC prototypes [267]. The experimental motive was not disclosed to the participants – doing so may impact their behaviour and responses. Although blinded experiments have already been conducted in the VR, USEC, and HCI research fields (for example, [10, 205, 292, 348, 561]), there are ethical considerations with not telling the entire truth to the participants. Therefore, the study’s aim was disclosed at the end of the study.

5.3 VR Studies to Augment Lab-Based Shoulder Surfing Research

5.3.1 Introduction

Accessing personal data almost anytime and everywhere has become a fundamental part of people’s daily lives. Examples include checking emails on smartphones, accessing the account balance through online banking apps, or withdrawing cash at ATMs. In many of these situations, people are required to authenticate, for example, to enter a PIN, which puts them at risk of getting observed, referred to as *shoulder surfing* [125]. Consequently, researchers looked into the shoulder surfing resistance of a large variety of authentication schemes (for example, [43, 99, 102, 239, 434]). A common approach in USEC research is to study such authentication prototypes’ security by inviting the participants to the lab, showing them two-dimensional (2D) video recordings that depict a user interacting with the system, and asking them to guess the observed PIN. These recordings show user authentications from predefined observation angles, with the intention to simulate a “best-case scenario” for an attacker that shoulder surfs the user. Previous works have shown that such 2D video recordings form a suitable baseline for shoulder surfing research [29]. However, it remains unclear if the selected perspectives represent best-case scenarios for attackers and if 2D video recordings provide realistic shoulder surfing experiences. Although studying shoulder surfing in a live setting is possible, real-time shoulder surfing studies are challenging [544], and in some cases even infeasible due to the various ethical and legal constraints [100, 529].

To draw on the success of the VR-based security study in Section 4.8, and to close the gap between commonly used 2D video recordings [102, 243] and the often hard-to-conduct real-time shoulder surfing evaluations, this study investigates the applicability of VR studies



Figure 5.1: The first study in this chapter explores the use of VR studies for shoulder surfing research. The impact of non-immersive and immersive VR observations on participants' observation performance and behaviour while shoulder surfing authentications is evaluated. To demonstrate the strengths of VR-based shoulder surfing research, this study explores three different authentication scenarios: (1) ATM authentication, (2) smartphone PIN authentication, and (3) smartphone pattern authentication.

and VR-based three-dimensional observations for shoulder surfing research. This thesis has already presented a comparison between 2D videos recorded in VR (*2DVO*, the baseline in this study) and traditional 2D real-world videos in Section 4.8 in chapter 4. Now, the first half of this chapter will advance research in this field by investigating the impact of 3D non-immersive and immersive VR observations on the participants' shoulder surfing performance and behaviour. Compared to shoulder surfing research on 2D video recordings, in 3D non-immersive and immersive VR observations the participants can freely change their observation position before observing the authentications, aiming at providing a more realistic shoulder surfing experience compared to how shoulder surfing research is currently conducted using traditional real-world video recordings.

As a result, the research in this section will showcase how VR studies enable researchers to study shoulder surfing in settings that are challenging to replicate in the lab and infeasible to research in naturalistic real-world settings. As discussed in the Introduction in Section 1.3.3, shoulder surfing is selected as security evaluation due to its common use when evaluating USEC prototypes (cf., [55, 56, 245, 530, 544]) and the impact it had on the USEC community and prototype designs in general. The three main aims of this study can be summarised as follows: (1) Exploring the strengths and weaknesses of non-immersive and immersive VR observations for shoulder surfing research (cf., Section 5.3.2.2); (2) Demonstrating how VR studies contribute to more realistic shoulder surfing research through three different authentication scenarios (cf., Figure 5.1); and (3) Discussing the findings in the light of prior works and providing lessons learned to support researchers when applying VR studies for shoulder surfing research. The research in this section contributes answers to the fourth research question:

RQ₄ Can substitutional in situ studies using VR provide a bridge over the methodological gap between lab and field studies?

If the question can be answered with a ‘yes’ and the use of VR studies contributes towards a bridge over the methodological gap between lab and field studies, then the research presented in this thesis section contributes to increased realism of USEC research that is concerned with shoulder surfing and USEC prototypes. If the question cannot be answered with a ‘yes’, further investigations would have been required. However, as this section will show, the use of VR studies provides the participants of user studies with a more realistic shoulder surfing experience than traditional 2D real-world recordings are capable of, enabling the researchers to advance shoulder surfing research and learn more about the shoulder surfers’ behaviour without necessarily impacting the results from security assessments of USEC prototypes.

5.3.2 Authentication Scenarios, Apparatus, and Implementation

Two public spaces were simulated to evaluate the suitability of VR studies for shoulder surfing research: 1) ATM authentication and 2) smartphone PIN (and pattern) authentication at a bus station (cf., Figure 5.2). These two scenarios were selected for several reasons: First, a survey by Eiband et al. [125] showed that shoulder surfing is most prominent in public spaces, especially when using smartphones. Second, ATMs are often found in public spaces, are frequently visited by people (for example, De Luca et al. [100] reported widespread ATM usage), and are challenging to research in the real world [100, 529]. Running a similar study in front of a real-world ATM is close to impossible in the detail required for in-depth authentication and shoulder surfing research. Finally, shoulder surfing forms an important threat vector in USEC research, is one of the most common used threat models when evaluating novel authentication prototypes (for example, [27, 102, 243]), and both studied authentication schemes, i.e., PIN and pattern input, form a popular security baseline in the human-centred security field (for example, for PINs: [27, 102, 158, 239], for patterns: [27, 102, 158]).



Figure 5.2: Two authentication environments were studied: PIN and pattern smartphone authentication at a bus station and an ATM authentication scenario next to a petrol station. The black circle shows the avatar’s position during the study (cf., Figure 5.1)

5.3.2.1 Apparatus and Implementation of the Authentication Scenarios

To evaluate the suitability of VR-based three-dimensional observations for shoulder surfing research, it was first necessary to collect recordings of users authenticating. The three authentication scenarios (cf., Figure 5.1) were implemented using Unity 3D (C#) to then prepare authentication recordings. As headset the HTC VIVE Tobii DEV KIT [510] was used and connected to a VR-ready laptop (*Razer Blade 15, NVIDIA GeForce RTX 2080*). A Leap Motion Controller for the hand and finger tracking [338] and an abstract avatar design with a head, body, legs, eyes, and hands were used to represent a user. The abstract avatar's dimensions and movements were mapped to a human in the real world. The abstract avatar design was informed by the research in Section 4.7 and Section 4.8 of chapter 4, suggesting that shoulder surfing studies conducted in virtual environments do not necessarily require highly realistic full-body avatars. Using an abstract avatar makes VR studies more accessible to the broader research community as it does not require additional expertise in tracking systems, i.e., OptiTrack, and experience in building avatars. The same abstract avatar (cf., Figure 5.1) was used for all three USEC systems, environments, and observation methods to contribute to internal validity. To track a smartphone's movements in the virtual environment (cf., Figure 5.1), an HTC VIVE tracker was attached to the back of a real smartphone, similar to Amano et al.'s work (cf., [19, Figure 5]). 2D video recordings and non-immersive/immersive VR recordings were then prepared for the study, as reported in Section 5.3.2.2. The participants' shoulder surfing experience was enriched with realistic environmental sounds that match the visual appearance of the VR environment, including traffic sounds and birds twittering.

A low-polygon styled city package [212], a 3D ATM model [143], and a slightly modified 3D smartphone model [124], i.e., the lock screen was replaced with a PIN and pattern authentication scheme, were used to present the participants with a VR replica of a real-world authentication environment. For the PIN-based authentication, Unity's `OnCollisionEnter` method [509], which triggers after another object collides, i.e., the user's finger, was used. Unity's `Line Renderer` component [1], which takes an array of $N \geq 2$ points in 3D space to draw a straight line between each point, was used to implement a realistic pattern-based authentication scheme. In the smartphone authentication scenarios, the user interface of the authentication scheme, i.e., the PIN/pattern layout, was only visible for the authentication duration. The authentication screen disappeared after entering a 4-symbol PIN/pattern, simulating real-world smartphone authentication where the user lands on their smartphone's home screen after unlocking their device.

5.3.2.2 Authentication Recordings

Traditional two-dimensional (2D) video recordings, the baseline of this study, are typically recorded from predefined observation angles to provide the user study participants with a best-case scenario, i.e., a clear sight on a mobile device's screen and input [43, 102, 239, 243, 434]. VR capture [420] was used to create 2D video recordings¹ of the user's input and the authentication scheme. Figure 5.1 on page 139 shows the three authentication systems the participants observed. The observation position and perspective that present the participants with a "best-case scenario" were selected through internal pilot tests. For the three-dimensional recordings, Ultimate Replay [512], a state-based replay system that records the scene at regular intervals, was used. Additional scripts to track mesh changes and keep track of the different states of Unity's Line Renderer component were implemented based on exchanges with Trivial Interactive, a small group of independent game developers who released UltimateReplay [512]. The participants then experienced the authentications for ~ 2 s - 3.5 s, similar to previous PIN/pattern-based research [6, 99], using *2D Video Observations*, *3D Observations*, and *VR Observations*. The observation methods are now described in detail:

- **2D Video Observations (2DVO, baseline).** The baseline of this study depicts the scenario where both the avatar's input and the authentication scheme were recorded using an angle that provides a shoulder surfer with a "best-case scenario", similar to how prior shoulder surfing evaluations were conducted (for example, [43, 239, 434, 530]). The participants performed their shoulder surfing observations on video recordings on a computer screen and could not manipulate the observation position and orientation. The authentications were recorded through virtual cameras in the virtual environment, as described in Section 5.3.2.2. The research in Section 4.8 in chapter 4 has shown that shoulder surfers' observation performance on VR-based two-dimensional video material aligns with the findings from a video-based real-world shoulder surfing study [245]. Therefore, the baseline against which 3D Observations and VR Observations are compared is based on 2D videos recorded in VR.
- **3D Observations (3DO, non-immersive).** The participants' initial observation view was positioned so that the camera pointed towards the avatar's back. This was done to ensure that the participants develop individual observation strategies and are required to change their initial position and perspective. The participants navigated in the environment using a traditional mouse-keyboard configuration, inspired by previous work on direct manipulations in non-immersive VR environments (for example, [131, 418]). They used the keyboard to simulate walking, i.e., translation along the x/y/z-axis,

¹The VR videos are available at <https://youtube.com/playlist?list=PLs1tzNuOyzfyQqqFILbfVi4LNGcBVcQ6X>, last accessed 22/01/2023

and the mouse to simulate head movements, i.e., rotations along the x/y/z-axis. After setting up their preferred observation position and orientation, the participants watched the authentications on a traditional computer monitor. They were not restricted to real-world conditions, which allows the study to explore if the participants exploit the affordances of 3D observations in a non-immersive VR environment (for example, being independent of gravitational force).

- **VR Observations (*VRO*, immersive).** The participants were wearing a VR headset, i.e., the HTC VIVE, and could freely move around and change their observation perspective and position. This depicts a scenario closest to in situ observations where a bystander freely moves around in physical space and shoulder surfs another person during their authentication. In comparison to *2DVO* and *3DO*, in *VRO* the participants were fully immersed in the VR environment.

5.3.3 Methodology

Several 1.5 hour in-the-lab investigations were conducted. In the role of attackers, the participants observed 648 authentications (18 participants \times 12 PINs/patterns \times 3 authentication scenarios). Participants were recruited using social media postings and word of mouth (outside of a university environment). The study followed a within-subjects design where all participants observed authentications in all three authentication scenarios: 1) 4-digit PIN entries on an ATM, 2) 4-digit PIN entries on a smartphone, and 3) 4-symbol pattern entries on a smartphone, and with all three observation methods. The order of the observation methods and the authentication scenarios was counterbalanced using a Latin Square.

As independent variable (IV) there was the observation type with three levels: *2DVO* (baseline), *3DO*, and *VRO*, and the threat model with two levels: single-view observations and repeated-view observations. Both threat models are frequently used when evaluating a USEC prototype's security (for example, [139, 239, 245]). Whilst in single-view observations the participants could observe the authentication only once, in repeated-view observations the participants could replay the authentication. The attack type alternated, as done in Section 4.8 in chapter 4. The impact of the IVs on four dependent variables was measured: **Observation Performance:** the participants' observation performance, the number of successful PIN/pattern guesses; **Levenshtein Distance:** the minimum number of single-digit edits between the participants' best guess and the correct PIN/pattern, which is commonly used in shoulder surfing research [3, 103, 158]; **Sense of Presence:** the participants' sense of presence experienced when using the different observation methods, measured using the IPQ questionnaire [451]; and **Perceived workload:** the participants' perceived workload when using the different observation methods, measured using the NASA-TLX questionnaire [194].

Demographic questions, including age, gender, and prior experience with VR, were asked using Qualtrics [397]. In-VR questionnaires [135, 395] were used to measure the participants' perceived workload (NASA-TLX [193]) and their sense of presence (IPQ [451]).

5.3.3.1 Study Procedure

The scenarios, the authentication schemes, the observation methods, and the tasks were explained to the participants before the data collection. The participants then went through an example observation where they observed an example PIN entry (for example, "1234"). Example authentications help the participants to familiarise themselves with the observation methods and the authentication schemes. They then started with the first observation method, for example, *2DVO*, and observed four authentications for each authentication context. The participants were not allowed to clip through the VR avatar in *3DO* and *VRO* as this would not be possible in the real world. However, they were not restricted from positioning themselves in, for example, front of the VR avatar because a) this can happen in the real world as well when, for example, standing at a bus station, and b) the use of virtual research artefacts enables investigating if participants make use of proxemics [186]: do observers in VR maintain a certain social distance to the person authenticating? are they aware that such observations are likely noticeable by the person authenticating? The participants could provide up to three PIN/pattern guesses for each observation, of which the guess closest to the correct PIN/pattern has been analysed. They then reported their perceived workload using the NASA-TLX questionnaire [193] and their sense of presence using the IPQ questionnaire [451]. The study concluded with semi-structured interviews on the participants' perceived performance and observation experience when using the different observation methods. Appendix D provides the semi-structured interview questions asked at the end of the study.

5.3.3.2 Data Analysis

Unless otherwise stated, an aligned rank transformation was applied to the data to correct for violations of normalcy, i.e., ART by Wobbrock et al. [555] and ART-C [126] for post-hoc pairwise comparisons (cf., ARTool in [R]²). The aligned rank transformation allows analysing multiple factors nonparametrically, which is not possible with classic nonparametric statistical tests, for example, with a Friedman test or Wilcoxon signed-rank test [555]. Therefore, the aligned rank transformation transforms non-parametric data into a form which can be analysed using a parametric statistical test [555]. Table 5.2 provides an overview of the F-ratios, together with the effect sizes, means, and standard deviations of the results.

²<https://depts.washington.edu/acelab/proj/art/>, last accessed 22/01/2023

The interview data was transcribed for the qualitative data analysis. The participant statements were split into meaningful excerpts. This process resulted in 292 participant statements, which were then systematically clustered using an affinity diagram. The lead researcher performed the initial clustering. A second researcher then reviewed the clustering independently and added tags to the clusters that required another iteration. Both researchers then met to discuss the clustering and resolve any discussion points during the review process. Three main themes that contribute to the validation of using VR studies for USEC research and advance research in this field were identified: 1) Observation Methods' Unique Characteristics, 2) *VRO* for More Realistic Shoulder Surfing Experiences, and 3) Lab vs Real-World Observations. Reporting the number of participants who shared certain opinions would be inaccurate due to the use of a semi-structured interview approach and the study's exploratory nature. Thus, frequencies are not reported. Quotes are translated from German to English where necessary.

5.3.4 Demographics

Eighteen (18) participants took part in this study (5 male, 13 female). The participants were on average 32.44 years (min = 18, max = 61, SD = 12.22). All participants reported that they have used an ATM before and own a smartphone they use daily. Slightly more than half of the participants (N = 11) mentioned that they have used VR before.

5.3.5 Results

This section first reports the participants' observation performance, represented through the percentages of successful observations and the mean Levenshtein distances. It then reports the participants' sense of presence and perceived workload when using *2DVO*, *3DO*, and *VRO*. Finally, the section provides a qualitative analysis of the semi-structured interviews and the participants' observation strategies. The results are summarised in Table 5.1 and Table 5.2.

5.3.5.1 Observation Performance and Levenshtein Distance

Participants' observations in *VRO* resulted in more successful observations (M = 93.14%, SD = 25.34%) than in *2DVO* (M = 89.35%, SD = 30.92%) and *3DO* (M = 81.40%, SD = 39.01%). To proceed with a more nuanced analysis and to gain better insights into how close the participants' guesses were to the entered PINs/patterns, the mean Levenshtein distances between participants' best guesses and the correct PIN/pattern were calculated.

ATM Authentication: For the ATM authentication scenario, the participants' observation performance was M = 94.44% (SD = 15.94%) for *2DVO*, M = 83.33% (SD = 23.90%) for *3DO*, and M = 95.59% (SD = 14.40%) for *VRO*. There was a significant effect of observation

method, $F_{(1,83)} = 4.584$, $p < 0.05$, $\eta_p^2 = 0.10$, and threat model, $F_{(1,83)} = 4.526$, $p < 0.05$, $\eta_p^2 = 0.05$, on participants' guesses and their distance to the correct PIN. There was also an interaction effect between threat model \times observation method, $F_{(1,83)} = 3.319$, $p < 0.05$, $\eta_p^2 = 0.07$. Follow-up analysis on the main effect of observation method revealed that participants' guesses on ATM authentications were closer to the correct PIN when using *VRO* ($M = 0.074$, $SD = 0.250$) and *2DVO* ($M = 0.097$, $SD = 0.288$) compared to *3DO* ($M = 0.278$, $SD = 0.470$) ($p < 0.05$). Table 5.1 provides an overview of the numbers for each authentication scenario and threat model, together with the statistical analysis in Table 5.2.

Smartphone PIN Authentication: The participants' observation performance was $M = 77.78\%$ ($SD = 30.34\%$) for *2DVO*, $M = 69.44\%$ ($SD = 36.41\%$) for *3DO*, and $M = 83.82\%$ ($SD = 26.74\%$) for *VRO*. There was a significant effect of observation method, $F_{(1,83)} = 4.95$, $p < 0.05$, $\eta_p^2 = 0.11$, and threat model, $F_{(1,83)} = 6.69$, $p < 0.05$, $\eta_p^2 = 0.07$, on the mean Levenshtein distance. Overall, the participants' guesses in *VRO* were closer to the correct PIN ($M = 0.265$, $SD = 0.448$) than in *3DO* ($M = 0.648$, $SD = 0.867$) ($p < 0.05$). There were no significant differences between the other pairs (*2DVO*: $M = 0.403$, $SD = 0.685$).

Pattern Smartphone Authentication: The participants' observation performance was $M = 95.83\%$ ($SD = 14.02\%$) for *2DVO*, $M = 91.67\%$ ($SD = 22.36\%$) for *3DO*, and $M = 100.00\%$ ($SD = 0.00\%$) for *VRO*. There was a significant effect of observation method, $F_{(1,83)} = 3.21$, $p < 0.05$, $\eta_p^2 = 0.07$, and threat model, $F_{(1,83)} = 25.53$, $p < 0.05$, $\eta_p^2 = 0.24$, on the mean Levenshtein distance. Overall, the participants' guesses in *VRO* were closer to the correct pattern ($M = 0.00$, $SD = 0.00$) than in *3DO* ($M = 0.139$, $SD = 0.371$) ($p < 0.05$). There were no significant differences between the other pairs (*2DVO*: $M = 0.083$, $SD = 0.305$).

Summary: Observation Performance

The Levenshtein distances confirm the differences in the participants' observation performance between *VRO* and *3DO*, but not between *VRO* and *2DVO*. *VRO* resulted in the most accurate observations, followed by *2DVO*.

5.3.5.2 Sense of Presence (IPQ)

There was a significant effect of observation method on the overall IPQ scores, $F_{(2,34)} = 71.429$, $p < 0.05$, $\eta_p^2 = 0.81$. Post-hoc analysis confirmed that the sense of presence was significantly higher in *VRO* ($M = 4.22$, $SD = 1.76$) than in *3DO* ($M = 2.28$, $SD = 1.93$) and in *2DVO* ($M = 1.55$, $SD = 1.77$) ($p < 0.05$). The difference between *3DO* and *2DVO* was significant too ($p < 0.05$). Figure 5.3 shows an overview of the results, featuring the subscales 1) sense of being there (PRES), 2) spatial presence (SP), 3) involvement (INV), and 4) experienced realism (REALISM).

Table 5.1: Overview of the number of successful observations (in %) and mean Levenshtein distances when using the different observation methods in the three authentication scenarios.

Variable	ATM			2D VIDEO OBSERVATIONS (2DVO)			Smartphone pattern		
	single	repeated	overall	single	repeated	overall	single	repeated	overall
Successful Observations	94.44% (SD=16.17%)	94.44% (SD=16.17%)	94.44% (SD=15.94%)	72.22% (SD=35.24%)	83.33% (SD=24.25%)	77.78% (SD=30.34%)	97.22% (SD=11.79%)	94.44% (SD=16.17%)	95.83% (SD=14.02%)
Levenshtein Distance	0.083 (SD=0.256)	0.111 (SD=0.323)	0.097 (SD=0.288)	0.639 (SD=0.888)	0.167 (SD=0.243)	0.403 (SD=0.685)	0.028 (SD=0.118)	0.139 (SD=0.413)	0.083 (SD=0.305)
	3D OBSERVATIONS (3DO)								
	ATM			Smartphone PIN			Smartphone pattern		
	single	repeated	overall	single	repeated	overall	single	repeated	overall
Successful Observations	80.56% (SD=25.08%)	86.11 (SD=23.04%)	83.33% (SD=23.90%)	55.56% (SD=37.92%)	83.33% (SD=29.70%)	69.44% (SD=36.41%)	86.11% (SD=23.04%)	97.22% (SD=11.79%)	91.67% (SD=22.36%)
Levenshtein Distance	0.333 (SD=0.515)	0.222 (SD=0.428)	0.278 (SD=0.470)	0.943 (SD=1.03)	0.361 (SD=0.564)	0.648 (SD=0.867)	0.250 (SD=0.493)	0.028 (SD=0.118)	0.139 (SD=0.371)
	VR OBSERVATIONS (VRO)								
	ATM			Smartphone PIN			Smartphone pattern		
	single	repeated	overall	single	repeated	overall	single	repeated	overall
Successful Observations	94.12% (SD=16.61%)	97.06% (SD=12.13%)	95.59% (SD=14.40%)	79.41% (SD=30.92%)	88.24% (21.86%)	83.82% (SD=26.74%)	100% (SD=0%)	100% (SD=0%)	100% (SD=0%)
Levenshtein Distance	0.118 (SD=0.332)	0.029 (SD=0.121)	0.074 (SD=0.250)	0.294 (SD=0.470)	0.235 (SD=0.437)	0.265 (SD=0.448)	0 (SD=0.000)	0 (SD=0.000)	0 (SD=0.000)

Table 5.2: F-ratios for the statistical analysis of the Levenshtein distances, IPQ scores, and NASA-TLX scores. $p < 0.05$ highlighted.

Measures	Observation Method	Threat Model	Observation Method \times Threat Model
Levenshtein Distance (ATM)	$F(1,83) = 4.584, p < 0.05, \eta_p^2 = 0.10$	$F(1,83) = 4.526, p < 0.05, \eta_p^2 = 0.05$	$F(1,83) = 3.319, p < 0.05, \eta_p^2 = 0.07$
Levenshtein Distance (Smartphone PIN)	$F(1,83) = 4.95, p < 0.05, \eta_p^2 = 0.11$	$F(1,83) = 6.69, p < 0.05, \eta_p^2 = 0.07$	$F(1,83) = 2.70, p = 0.073, \eta_p^2 = 0.06$
Levenshtein Distance (Smartphone pattern)	$F(1,83) = 3.21, p < 0.05, \eta_p^2 = 0.07$	$F(1,83) = 25.53, p < 0.05, \eta_p^2 = 0.24$	$F(1,83) = 2.62, p = 0.0789, \eta_p^2 = 0.06$
IPQ Presence Score	$F(2,34) = 71.429, p < 0.05, \eta_p^2 = 0.81$	n/a	n/a
Sense of being there (PRES)	$F(2,34) = 31.932, p < 0.05, \eta_p^2 = 0.65$	n/a	n/a
Spatial Presence (SP)	$F(2,34) = 59.61, p < 0.05, \eta_p^2 = 0.78$	n/a	n/a
Involvement (INV)	$F(2,34) = 20.592, p < 0.05, \eta_p^2 = 0.55$	n/a	n/a
Realism (REAL)	$F(2,34) = 23.944, p < 0.05, \eta_p^2 = 0.58$	n/a	n/a
NASA-TLX	$F(2,34) = 4.715, p < 0.05, \eta_p^2 = 0.217$	n/a	n/a

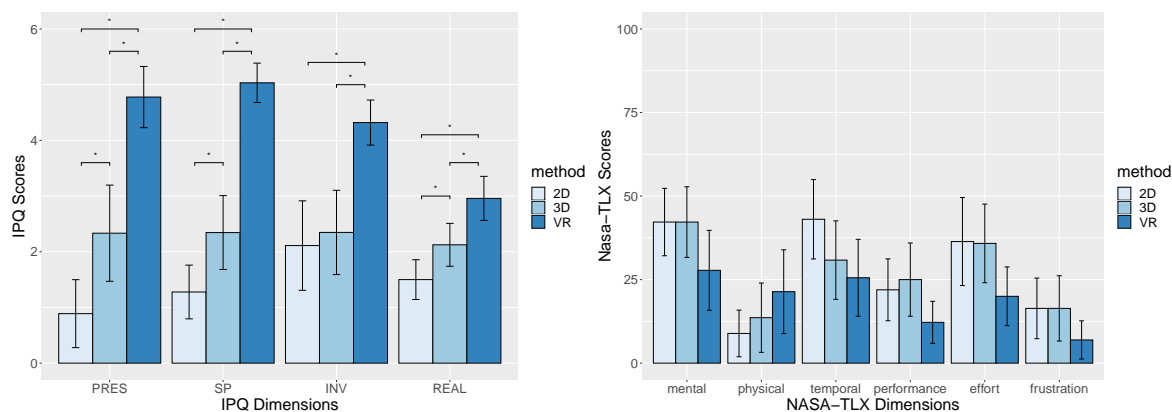


Figure 5.3: *VRO* led to a significantly higher sense of being there, higher spatial presence, higher involvement, and higher experienced realism than *2DVO* and *3DO*. There were no statistically significant differences in the participants' perceived workload when using the different observation methods. Error bars denote the 95% confidence interval (CI).

A more nuanced analysis of each IPQ subscale level was performed to better understand the significant differences:

Sense of being there. The observation methods elicited statistically significant changes in participants' sense of being, $F_{(2,34)} = 31.932$, $p < 0.05$, $\eta_p^2 = 0.65$. Post-hoc analysis revealed a statistically significant lower sense of being in *2DVO* ($M = 0.88$, $SD = 1.45$) and in *3DO* ($M = 2.33$, $SD = 2.14$) compared to *VRO* ($M = 4.78$, $SD = 1.55$) ($p < 0.05$). The difference between *2DVO* and *3DO* was also statistically significant ($p < 0.05$).

Spatial presence. The participants' experienced spatial presence differed statistically significantly between the observation methods, $F_{(2,34)} = 59.61$, $p < 0.05$, $\eta_p^2 = 0.78$. Post-hoc analysis revealed statistically significant differences in participants' spatial presence in *2DVO* ($M = 1.28$, $SD = 1.48$) and in *3DO* ($M = 2.34$, $SD = 1.99$) compared to *VRO* ($M = 5.03$, $SD = 1.18$) ($p < 0.05$). The difference between *2DVO* and *3DO* was significant too ($p < 0.05$).

Involvement. The experienced involvement was statistically significantly different across the observation methods, $F_{(2,34)} = 20.592$, $p < 0.05$, $\eta_p^2 = 0.55$. Post-hoc analysis revealed statistically significant differences in *2DVO* ($M = 2.11$, $SD = 2.15$) and in *3DO* ($M = 2.35$, $SD = 1.91$) compared to *VRO* ($M = 4.32$, $SD = 1.46$) ($p < 0.05$). There was no evidence that the participants' experienced involvement differed statistically between *2DVO* and *3DO*.

Realism. The participants' experienced realism was statistically significantly different between the different observation methods, $F_{(2,34)} = 23.944$, $p < 0.05$, $\eta_p^2 = 0.58$. Post-hoc analysis revealed statistically significant differences in participants' experienced realism in

2DVO ($M = 1.50$, $SD = 1.64$) and in *3DO* ($M = 2.13$, $SD = 1.83$) compared to *VRO* ($M = 2.96$, $SD = 1.98$) ($p < 0.05$). The difference between *2DVO* and *3DO* was significant too ($p < 0.05$).

Summary: Sense of Presence

VRO led to a higher sense of being part of the virtual environment, to a higher spatial presence, and to a higher feeling of involvement and experienced realism than *2DVO* and *3DO*. This suggests that *VRO* contribute to reasonably realistic shoulder surfing experiences, even in a laboratory setting.

5.3.5.3 Perceived Workload (NASA-TLX)

The participants' perceived workload was normally distributed for *2DVO* and *3DO*, but not for *VRO*, as assessed by Shapiro-Wilk's test ($p < 0.05$). As ANOVAs are considered to be fairly robust to small deviations from normality [49, 161, 444], a one-way repeated-measures ANOVA was performed. Participants' perceived workload was statistically significantly different between the observation methods, $F_{(2,34)} = 4.715$, $p < 0.05$, $\eta_p^2 = 0.217$. However, the post-hoc analysis did not confirm the significant differences between the observation methods ($p > 0.05$). The mean values of the participants' perceived workload were $M = 28.15$ ($SD = 15.77$) for *2DVO*, $M = 27.31$ ($SD = 14.61$) for *3DO*, and $M = 18.98$ ($SD = 17.62$) for *VRO*. Figure 5.3 shows the mean NASA-TLX values for each dimension.

Summary: Perceived Workload

There was no evidence that *VRO* or *3DO* led to a higher workload than *2DVO*. This suggests that the differences in the participants' perceived workload when using *2DVO*, *VRO*, and *3DO* are negligible.

5.3.5.4 Semi-structured Interviews

Semi-structured interviews were conducted to learn more about the participants' perceptions and performances when using the observation methods. Additionally, the interviews allow the thesis to comment on the participants' perceived differences to shoulder surfing in the wild.

Theme 1: Observation Methods' Unique Characteristics. Although *3DO* provided the participants with a more realistic shoulder surfing experience than *2DVO*, the mouse-keyboard interaction impacted their observation performance. Consequently, the "plug-and-play" characteristic of *2DVO* resulted in observations being perceived as easier than *3DO*. P11 mentioned that in *VRO* "[they] could position [themselves] in a way how

they wanted it and it was super easy to select the position; this was more difficult with keyboard/mouse” (P11). Others mentioned that in *VRO* “[you] just need to walk to a specific position” (P17). Regarding *3DO*, the participants mentioned that their experience was closer to reality than *2DVO* because “it felt more like that [they] really want to look over someone’s shoulder” (P15). P7 mentioned that “they could experiment a bit like in the real world where you can observe [the authentication] from different perspectives.” (P7). Although some participants in *2DVO* raised the lack of manipulations, there was a consensus that it was easier to observe authentications in *2DVO* than in *3DO*. The participants mentioned that the observation position and angle provided them with a clear line of sight and that their only task was to watch the recordings. In fact, some participants mentioned they found the 2D VR videos more realistic because they used *VRO* and *3DO* “in a way to really abuse them” (P9). Section 5.3.6.1 and Figure 5.4 will discuss some observation positions in more detail.

Theme 2: *VRO* for More Realistic Shoulder Surfing Experiences. In *VRO*, P3 voiced that “the [real] environment would be completely irrelevant; it does not matter if [they are] in a basement, in an attic, outside, or at the sea” (P3) and that they did not feel like being part of an experiment. Others mentioned that “with the VR headset [they] moved within the environment and it felt on a physical way more realistic” (P4). For *3DO*, the participants voiced that they did not feel being part of the environment to the same extent as in *VRO* because of the presence of reality and that they were “aware of everything that surrounded [them] in the reality” (P15). P3 explained this based on the fact that they were “sitting in front of the PC and could see stuff on the left and right side that is not related to the [authentication scheme]” (P3). For *2DVO*, the participants emphasised that their task was only to “watch” the authentications and that they were “very conscious that there is a technical device between [them] and the environment” (P4). The overall qualitative feedback suggests two extremes: whilst *VRO* contributed towards a reasonably realistic in situ shoulder surfing experience, *2DVO* and *3DO* were considered as observations from “another reality”, with *3DO* being slightly more capable of contributing to an in situ shoulder surfing experience.

Theme 3: Lab vs Real-World Observations. The participants reported that they would perform real-world observations similarly to *VRO* where they could freely move around in VR: “I can imagine that [real-world observations] work exactly how I did it in VR” (P12). However, across all participants, the message was that they would respect the social distance to the user more in the real world. P9 mentioned that “[they] would probably stay further away and do it less conspicuously” (P9). Others voiced that they ignored the social factor during the study and “only optimised [their] viewing point” (P10). P10 further voiced that they “did not pay attention on [the social] factor [and] didn’t care about the user standing there.” (P10). P11 mentioned that “at the point where I had on the VR headset I looked over

someone's shoulder and I probably would not do this [in the real world]. I would work more with the eyes instead of really looking over someone's shoulder and go that close." (P11). P4 added that in the real world "there would be other people [and that they] would probably feel being observed" (P4). P13 voiced that in the ATM scenario "the user who withdraws cash probably already acts precautiously – so you would realise when someone stays that close to you." (P13). P18 voiced that they would not "stay that close to others; they can already feel you breathing; especially in the ATM example it is very unlikely that this happens in such a way" (P18). Some participants mentioned they were unaware that such observations could happen in the real world and that *VRO* made them aware of the potential threat of shoulder surfing observations. P3 even voiced that they would now tell other people to "protect their PIN with the left hand" (P3).

In summary, *VRO* contributed to more realistic shoulder surfing experiences than *2DVO*; however, the participants mentioned that people would sense if someone is close to them. In the study, the participants did not necessarily consider the social factors (for example, proxemics [186]), in their observations (cf., the participants' tracked observation positions in Figure 5.4, visualised through black dots). Social factors can indeed take on an essential role in real-world observations. For example, Brudy et al. [61] utilised notions of proxemics to provide participants with an awareness of shoulder surfing moments and protect their information against bystanders when the system detects shoulder surfing. Section 5.3.6 will discuss the shoulder surfing behaviour of the participants in more detail.

5.3.6 Discussion

The first half of this chapter explored how the use of VR studies contributes to advanced shoulder surfing research. Through its comparisons to a well-established baseline in the literature, it further strengthens the validation of applying VR studies for USEC research. One finding was that *VRO* provided the participants with a reasonably realistic shoulder surfing experience without negatively impacting their shoulder surfing performance compared to the use of traditional 2D video recordings (cf., Section 5.3.5.1). *VRO* contributed to a higher sense of being in the environment, a greater feeling of spatial presence, a higher level of involvement, and a higher experienced realism than *2DVO*, the de facto standard approach when evaluating a USEC prototype's security.

Whilst the advantages of *VR Observations* over more traditional 2D video recordings (i.e., *2DVO*) are expected findings with the benefits of immersive VR in terms of sense of presence being known by the VR community [490, 522], *VRO*'s affordances are particularly interesting for human-centred security research. The findings imply that previous shoulder surfing studies using 2D video recordings were not capable of providing the participants of user studies with a realistic shoulder surfing experience. Therefore, shoulder surfing studies based on 2D

video recordings impacted the often desired high ecological validity, as highlighted in Section 3.3.5 in chapter 3. Despite the strengths of *VRO*, the results suggest that *2DVO* are sufficient to assess a system's resilience against observations (cf., Section 5.3.5.1). This confirms Aviv et al.'s findings when comparing *2D* video recordings with live observations [29]. In all three authentication contexts, there was no evidence that observations using *VRO* were more accurate than *2DVO*. In the next section, the impact of *3DO* on shoulder surfing experiments and the participants' observation behaviour are discussed in more detail. The participants' observation behaviour was similar across the authentication scenarios with no notable differences; therefore, the smartphone PIN/pattern visualisations are moved to Appendix D and the discussion on the participants' observation behaviour in Section 5.3.6.1 is based on the ATM authentication scenario.

5.3.6.1 VR Observations: A Blessing and Curse for USEC Research

The participants' shoulder surfing behaviour suggests that they made use of the unique characteristics of non-immersive VR in *3DO* (cf., Figure 5.4). This was apparent as follows:

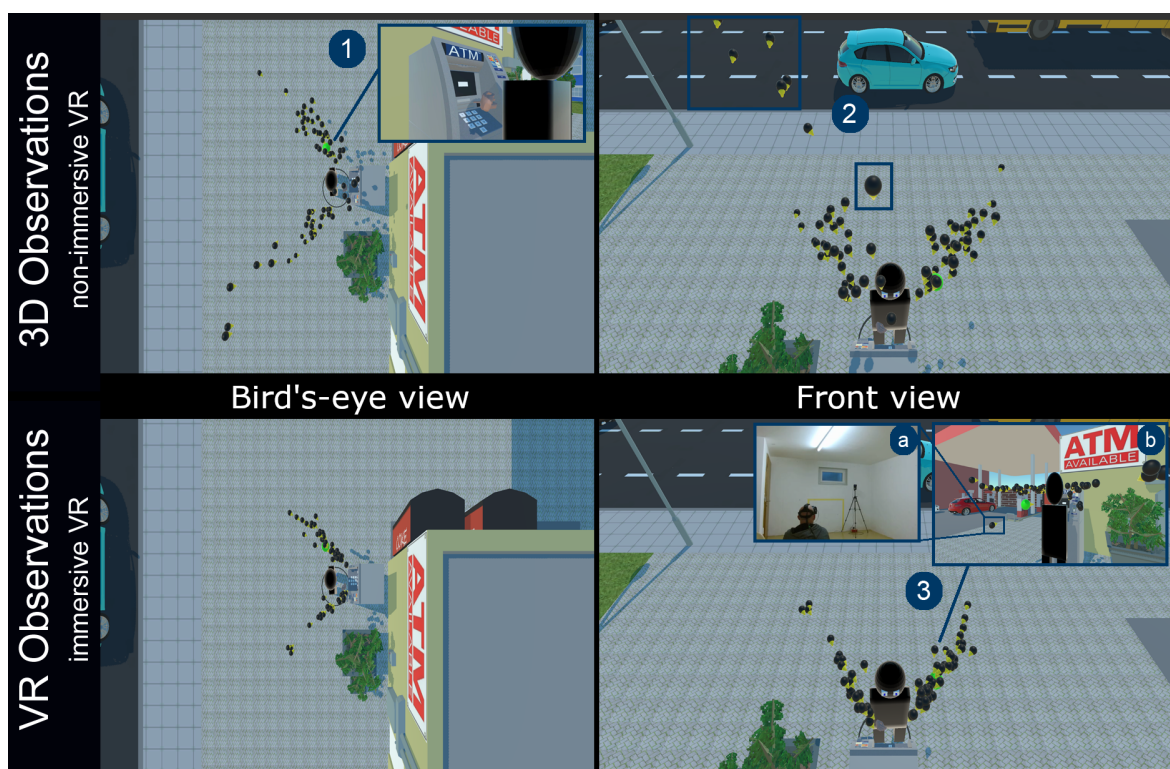


Figure 5.4: ❶ shows the reference position + orientation of *2DVO*. The participants made use of the absence of physical constraints in *3DO* (cf., ❷). The immersive VR observations showed that social factors (for example, the proximity to the VR avatar) lose relevance in such a virtual environment, which is discussed further in Section 5.5.1.1. ❸ shows a *VR observation* in which the participant pretended to tie their shoes while performing the observation (a); (b) shows the observation position through another perspective.

In *3DO*, the participants positioned themselves in several positions, many of which are challenging to reach in close-to-reality *VRO* (or in the physical real world) due to gravitational force and physical constraints. Although some of these positions seem to be unrealistic at first glance, observations from above the avatar (cf., ② in Figure 5.4) can indeed happen in the real world. For example, drones equipped with cameras [536] or surveillance cameras (CCTVs) on the corner of buildings can lead to such observation positions and angles as depicted in Figure 5.4. Furthermore, some participants linked their observations to other real-world actions. P7 brought up the example of observing ATM input in an unobtrusive way while tying their shoes (cf., ③-a in Figure 5.4).

The use of VR studies, *VRO*, and *3DO* for shoulder surfing research enables researchers to study different observation strategies in much more detail than what can be achieved with traditional 2D video recordings, i.e., *2DVO*. However, VR studies for this type of research might not be in favour of critical security evaluations when the observation method deviates from a realistic observation, such as mouse-keyboard manipulations as used in *3DO*. Figure 5.4 and the qualitative feedback suggest that the participants made use of the affordances of *3DO* to, for example, being physically independent, but *2DVO* and *VRO* led to more accurate observations and to a more precise security assessment (cf., Section 5.3.5.1).

The results suggest that when VR observation methods are introduced and the shoulder surfing resilience of a system is at the centre of the investigation, the participant-defined observation positions can greatly overestimate a prototype's resilience against observations. Taking *3DO* and ATM authentication as an example, someone could conclude that observations on ATM authentications are successful in “only” 83.33% observations, whilst both the de facto standard evaluation approach (*2DVO*) and *VRO* resulted in noticeable more successful observations (*2DVO*: 94.44%, *VRO*: 95.59%). Therefore, researchers risk being misled into thinking that their USEC prototype is more resilient against observations than it actually is.

5.3.6.2 VR Observation Methods and Their Use Cases

The literature discussed how the participants' lack of experience could lead to an underestimation of risk [544] and emphasised the importance of the participants' familiarity with the authentication methods [99, 102, 239, 263, 308]. As shown in this study, the participants' experiences are essential when researchers introduce novel observation methods for shoulder surfing research. As evidenced by the participants' feedback during the semi-structured interviews, *VRO* were perceived as highly realistic. However, the interaction with alternative methods, which differ from their real-world observation experiences (for example, mouse-keyboard manipulations as in *3DO*), harm shoulder surfing evaluations and corresponding security conclusions of USEC systems. Still, in cases where the focus is more on an exploratory shoulder surfing evaluation, such as studying the participants' observation

behaviour and their strategies, shoulder surfing methods such as *3DO* can be particularly helpful: they enable researchers to analyse situations that are challenging to research using existing means in physical lab settings. Furthermore, the use of VR studies can be beneficial to contribute to more realistic USEC research. Studies of this type are particularly promising when researchers aim to run many consecutive experiments. It is often easier to maintain and access VR replicas of security-sensitive environments and make adjustments. For example, VR studies easily allow exchanging USEC systems or immersing user study participants into different security-sensitive environments, as demonstrated by the three different authentication methods and authentication scenarios in this study.

5.3.6.3 VR Shoulder Surfing Studies and Research in the Wild

It is important to emphasise that VR studies for shoulder surfing research should not replace field studies, but rather complement them in situations where going into the field is infeasible or too challenging. As reviewed in the Literature Review in chapter 2 and highlighted by Mäkelä et al. [291]: “VR field studies situate between lab studies and real-world field studies, being closer to field studies in ecological validity, and closer to lab studies with regards to their required effort”. The shoulder surfing study presented in this section demonstrated how VR enables researchers to study human shoulder surfing on USEC systems in several contexts. Examining all three authentication contexts in the wild would require a significant amount of additional hardware (for example, tracking sensors and cameras) and is often prohibited due to the ethical and legal constraints of security-sensitive contexts [100, 529]. Whilst the USEC and HCI communities often expect field research to increase the generalisability and the ecological validity of research findings, as reported in chapter 3, it has been argued that “we [as a community] just need to be a little bit more open to what sort of solutions/evaluations we are expecting out of [something] that has not actually been deployed in the real world.” (cf., P11 in Section 3.3.5.1 in chapter 3).

The results of this work imply that, yes, the use of VR studies for shoulder surfing contributes to more realistic research and is superior to studies in more traditional lab environments; however, the researchers’ expectations of VR studies and simulations of real-world shoulder surfing cannot be to fully bridge the gap between lab studies and more naturalistic in-the-wild studies. Yet, in situations where shoulder surfing research in the field is infeasible, VR shoulder surfing studies are superior and forge forward USEC research as they allow to evaluate prototypes and learn more about potential security threats in plausible environments.

5.3.7 Limitations

There are some limitations that are worth discussing. The study evaluated the participants' shoulder surfing behaviour and provided insights into how the participants used VR's unique affordances when performing observation attacks on different USEC systems and in different environments. However, the research in this section did not account for the many additional factors, such as shoulder surfing users when interacting with different devices (for example, tablets [402]) or situations in which shoulder surfing defense strategies are applied [246]. Similar to the work by Aviv et al. [27], the research did not study text-based authentication, mainly because traditional PIN and pattern authentications are the most commonly used baseline measures in shoulder surfing and authentication research [102, 158, 239]. Future research may apply 3D VR recordings to evaluate multimodal authentication schemes (for example, eye gaze in combination with touch or mid-air input [5, 239]).

Furthermore, a non-vivid environment with no additional bystanders was used to immerse the participants into different authentication scenarios and contribute to a strong baseline evaluation of the use of VR studies for shoulder surfing research. No additional noise was introduced because one key factor of shoulder surfing research on authentication systems is to provide participants with a best-case scenario when observing authentications [43, 239, 434, 530]. In such scenarios, bystanders or, in general, more vivid environments might introduce a confounding variable which would not allow the thesis to comment on and provide a strong baseline evaluation. Finally, a photorealistic VR environment may further increase the visual realism of such a virtual environment. However, recording security-sensitive scenarios is often infeasible in the wild. For example, creating 360° real-world recordings as done by Saad et al. [428] introduces ethical and legal challenges in the context of ATM authentication. Such recordings are limited to what is possible to stage or record in the real world. Virtual replications are particularly promising at this point: they provide researchers with more flexibility in changing parts of the environment [291] and allow them to study scenarios that are challenging or even impossible to access in the real world. Observing user authentications on, for example, an ATM in the real world would go beyond what is ethically and legally possible. Therefore, *3DO* and *VRO* were compared against the de facto standard approach when evaluating authentication systems, i.e., *2DVO*, the baseline. The qualitative data hinted at similar observation behaviour to the real world (cf., Section 5.3.5.4).

5.3.8 Conclusion

The first half of this chapter investigated, for the first time, the use of VR studies for human-centred shoulder surfing research. It introduced non-immersive and immersive VR observations to advance lab-based shoulder surfing research and contributed to bridging the gap

between more traditional lab research and research in the field. Furthermore, the research demonstrated how VR and its unique affordances can be applied in the human-centred security research domain to study shoulder surfing in different authentication scenarios affordably and effortlessly. It showed that immersive VR recordings provided the participants with a reasonably realistic shoulder surfing experience without impacting their observation performance compared to commonly used 2D video recordings.

5.3.8.1 Contribution to Research Question 4 (RQ₄)

The shoulder surfing study in this chapter contributes answers to the fourth research question:

RQ₄: Can substitutional in situ studies using VR provide a bridge over the methodological gap between lab and field studies?

As discussed by De Luca et al. [100] and Volkamer et al. [529], and highlighted in chapter 3 of this thesis, conducting USEC research in the wild is challenging and often infeasible to do with the level of detail required for usability and security research. Therefore, comparisons to real-world shoulder surfing experiences were mainly performed on the participants' feedback during the semi-structured interviews in Section 5.3.5.4. Furthermore, the aim of VR studies is not to replace the existing USEC research methods, but to complement them and open the door for and facilitate hard-to-conduct research in the wild.

In summary, and to provide answers to RQ₄, the results from the VR shoulder surfing study suggest that, 'yes', the use of VR studies can contribute towards a bridge over the methodological gap between lab and field studies. The interviews revealed that the participants perceived *VRO* as more realistic than traditional *2DVO*. They reported that they felt not being part of a user study because they were unaware of the real-world study setting when using *VRO*: "the [real] environment would be completely irrelevant; it does not matter if [they are] in a basement, in an attic, outside, or at the sea" (cf., P3 in Section 5.3.5.4). Interestingly, the participants' observation performances remained relatively stable regarding the use of traditional *2DVO*, which further strengthens the use of VR studies for shoulder surfing research. Whilst results from VR studies align with the more traditional studies that use 2D (real-world) video recordings, the use of VR facilitates and advances research in this field by providing the participants with more realistic shoulder surfing experiences. The advantages of VR-based shoulder surfing studies over traditional lab studies are also beneficial for researchers. Due to the benefits of VR studies, researchers can more accurately study their USEC prototype resistance against observations and iteratively incorporate the participants' observation behaviour into the design and evaluation of novel USEC prototypes.

The use of VR studies for shoulder surfing research is particularly auspicious when the aim is to combine the strengths of live observations and video recordings. Although video recordings contribute to internal validity as the entire sample receives the same video material, live observations can, as evidenced by Schaub et al. [442], result in more critical security evaluations. VR shoulder surfing studies, as presented in this work, combine the internal validity of recordings with the ecological validity of live observations, eventually contributing to and allowing researchers to conduct realistic shoulder surfing research despite the ethical and legal challenges in the field. This means that *VRO* contributes towards a future, well-established research method that enables researchers to conduct in situ shoulder surfing research on different USEC systems and in various security-sensitive contexts, which was not possible previously using more traditional research methods. The research this thesis will discuss in Section 5.4 provides additional answers on the use of VR studies for bridging the methodological gap between lab and field studies, and further supplies the validation of this thesis regarding the use of VR studies for USEC research.

5.4 VR Studies for Simulated In Situ USEC Research

5.4.1 Introduction

As the Literature Review in chapter 2 and the USEC expert interviews in chapter 3 have shown, researchers face significant challenges when conducting research in the wild. A classic example is research on ATMs, an area of research that is challenging to perform due to ethical and legal constraints [100, 529]. Whilst a plethora of novel ATM authentication methods has been proposed (for example, [39, 99, 118]), there is a shortcoming in research methodologies that allow researchers to evaluate these systems in their corresponding environment, i.e., in situ, creating uncertainty in the value and validity of USEC research conducted in the lab, as voiced by the interviewed experts in chapter 3. Many of these authentication methods can provide end-users with more usable and secure authentication experiences than traditional 4-digit PIN authentication on a physical keypad can achieve. However, the vast majority of these systems has not been evaluated in realistic scenarios in the wild. Consequently, novel usable and secure authentication methods have often not found their way into practice.

ATMs in the United Kingdom use traditional 4-digit PIN authentication since 1976, which is subject to many vulnerabilities (for example, shoulder surfing attacks [102, 243] or thermal attacks [3, 15]). One potential reason for the lack of transitioning research artefacts into practice is the challenges around resources and links to industry partners, as learned in the expert interviews in chapter 3. It is also not ethically and legally feasible to video record people's actual PIN input in the real world, making academic research in this space particularly

challenging [24, 100, 529]. As a result, researchers either evaluate their security systems in a setup that is isolated³ from an actual authentication scenario (for example, [99, 245]), or aim to create “realistic” authentication scenarios in the lab (for example, [118]). However, it remains to be seen how the conclusions drawn from lab experiments align with the findings from more realistic in situ evaluations where the authentication scheme is part of an actual production task [437] (for example, withdrawing cash).

In situ investigations, where the authentication scheme is part of a production task, are particularly interesting because authentication is usually not a person’s primary task [437]. Although the study in Section 5.3 achieved promising results when exploring the use of VR studies for enhanced security research with shoulder surfing as an example security threat (for example, a security evaluation that aligns with existing baseline evaluations whilst providing insights into observation strategies; cf., Table 5.1 and Figure 5.4), it remains unclear how well in situ USEC research, where the usability of novel USEC prototypes is investigated in a realistic use case, can be simulated in VR. The lack of answers raises the following questions: does a USEC prototype’s usability evaluation results differ when evaluated in situ rather than isolated from a person’s production task? do people behave in VR-simulated authentication scenarios similarly as they do in the real world?

To contribute answers to RQ₄ and further address the overarching research question, RQ₆, this study reports on a user study where the participants (N = 20) experienced an ATM authentication scenario in the real world and VR. The study will compare a) the participants’ performance and their behaviour when interacting with two different ATM replications and b) how embedding an authentication scheme into its actual usage context impacts the experiment’s usability results. The results of this study contribute additional answers to the fourth research question:

RQ₄ Can substitutional in situ studies using VR provide a bridge over the methodological gap between lab and field studies?

As in Section 5.3, but this time with a focus on usability evaluations, if RQ₄ can be answered with a ‘yes’ and the use of VR studies can contribute towards a bridge over the methodological gap between lab and field studies, then the research in this section contributes to the often desired high ecological validity of USEC research and enables researchers to move their usability evaluations on USEC prototypes out of traditional lab settings into more realistic security-sensitive (VR) environments. If the answer is ‘no’, then VR studies are incapable of simulating realistic in situ usability evaluations, and future work would be needed to investigate other approaches capable of simulating in situ experiences. However, as this

³*Isolated* refers to a scenario where the participants experience the authentication independent from an actual production task [437].



Figure 5.5: The second study in this chapter investigates, for the first time, the impact of *isolated* authentications, where users authenticate in a lab environment (❶), and *in situ* authentications, where users' authentication precedes a primary task (❷), on a prototype's usability evaluation results. The research highlights the importance of simulated *in situ* USEC evaluations, demonstrates how the use of VR studies advances prototype evaluations, and enables researchers to conduct usability research in contexts that are close to impossible to study in the wild and challenging to replicate in the lab.

section will show, the use of VR studies contributes towards realistic simulations of *in situ* authentication experiences and can transition the authentication task into a secondary task that precedes a production task [435], which matches more closely to how security tasks are perceived in the real world.

5.4.2 Methodology

Whilst a common approach to evaluate novel authentication schemes is to compare their usability to traditional authentication systems (for example, [93, 239, 243, 367]), this study investigates how the context in which a novel authentication prototype is evaluated impacts the user behaviour and the prototype's usability evaluation results. So far, this thesis has not performed an empirical comparison between simulated *in situ* and *isolated* authentications, which is particularly interesting from a human-centred security perspective. Conducting detailed *in situ* authentication research in the real world is often infeasible, and authentication

is usually not considered a person's primary task [251, 267, 437]. As a result, and to further contribute to the validation of the use of VR studies for simulating real-world USEC research, the study in this section compares simulated in situ evaluations to evaluations in the lab in both the real world, which simulates a physical lab environment and a public space, and in a virtual environment (cf., Figure 5.5 and Figure 5.7).

The study followed a within-subjects design with the order of the conditions being counter-balanced using a Latin Square. The ATM environments were set up as realistic as possible to contribute to ecological validity (cf., Figure 5.7 and Section 5.4.3). Skarbez et al. [473] emphasised that a realistic scale of the space is the most crucial factor for generating a "feeling of reality" [473]. Although the studied authentication settings depict a "realistic" ATM scenario, there is, in line with the results in Section 5.3, still a gap to an ATM experience in the wild. The differences between this study's findings and in-the-wild ATM interaction observations by De Luca et al. [100] are further discussed in Section 5.4.6.2.

5.4.2.1 Studied USEC Prototype: ColorPIN

ColorPIN [99] was replicated to achieve this study's main goal of investigating the suitability of VR studies for simulated in situ USEC research and assessing the impact of VR and in situ evaluations on the users' authentication performance and behaviour. The motivation behind replicating ColorPIN [99] is manifold. First, ColorPIN is proposed as an authentication scheme for ATMs, but its original evaluation took place in isolation, which means that the users' authentications were not part of an actual ATM interaction scenario [99]. This study aims to close the gap between commonly isolated usability evaluations and hard-to-conduct in situ usability evaluations of USEC prototypes. Second, ColorPIN's intended application context, ATM authentication in security-sensitive public environments, received significant attention from the USEC community in the past and highlighted the challenges researchers experience when conducting USEC research [100, 529]. Furthermore, ColorPIN's underlying concept, i.e., one-to-one relationship between PIN length and required input, is commonly used in authentication research to not artificially increase input times (for example, [108, 245, 308, 530]).

In summary, the financial, ethical, and legal barriers when conducting USEC research in the wild (cf., chapter 3), the widespread use of ATMs [349], and ColorPIN's characteristics [99], make ColorPIN a suitable candidate for the first investigation of VR studies for simulated in situ usability evaluations of USEC prototypes.

ColorPIN: A Brief Overview. ColorPIN is a highly secure and usable ATM authentication scheme, initially proposed by De Luca et al. [99] and further studied by Bianchi et al. [40] and Lee [278]. A user enters a ColorPIN using a commercial keyboard by selecting

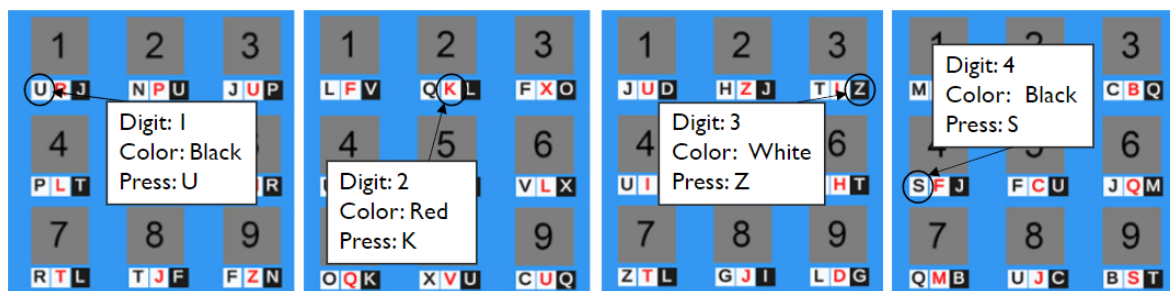


Figure 5.6: Example of a ColorPIN input which was used during the study to introduce the participants to the prototype.

a letter corresponding to a digit (cf., Figure 5.6). For example, instead of entering 1-2-3-4 on a keypad, users map their ColorPIN to coloured letters displayed below the digits on the authentication interface. To input 1(black) in Figure 5.6 the user would need to press “U” on the keyboard. The letters are randomly assigned after each input for increased security.

5.4.2.2 Independent Variables

The impact of two independent variables (IVs) on user authentications was investigated: the authentication context, isolated from a primary task vs integrated into a primary task (in situ); and authentication environment, real world vs VR. ColorPIN was replicated in VR and used as authentication scheme for all conditions to contribute towards internal validity. Figure 5.7 shows the replicated ColorPIN [99] prototype in all authentication environments and contexts. The “in the lab” conditions in Figure 5.7 represent the state-of-the-art baseline conditions, which were implemented in reality and VR. The “in the field” conditions in Figure 5.7 are inspired by Krol et al. [267] and Sasse and Fléchaïs’ [437] comments on security and its secondary role in everyday life. Therefore, in “in the field” the ColorPIN authentication prototype is embedded into its intended ATM authentication application (cf., Section 5.4.2.3).

Authentication Context (IV1). First, the extent to which the authentication context impacts the users’ authentication performance and behaviour was investigated. The research in this section distinguishes between two contexts:

- **Isolated Authentication (Lab):** *Isolated* refers to a traditional lab setting where the USEC prototype is not evaluated in the intended usage context (for example, on a desktop PC instead of an ATM). This presents the participants with an authentication isolated from a production task, the de facto standard when evaluating authentication prototypes (for example, [6, 243, 245, 530]). This authentication context aligns with the study context of the original real-world study by De Luca et al. [99] and forms the baseline in the real world (*RW Lab*) and VR (*VR Lab*).



Figure 5.7: Five simulated authentication scenarios were studied: Two in the real world and three in VR. Virtual replicas of both real-world environments were implemented. *In the lab* was treated as a baseline in this study in both the real world (*RW Lab*) and VR (*VR Lab*).

- Integrated Authentication (ATM):** ColorPIN was integrated into an actual ATM for which the USEC prototype has initially been built [99]. Integrating the USEC prototype into an ATM aims to increase authentication realism. By doing this, the users' attention is not artificially drawn to the authentication itself. There were two scenarios: one in the real world (*RW ATM*) and one in VR (*VR ATM*). Due to the required resources in the real world to simulate realistic ATM environments (for example, additional bystanders and access to a public space with an ATM), this study further demonstrates how the use of VR studies advances USEC research by an additional condition (i.e., *VR ATM Public*). *VR ATM Public* allows the thesis to comment on the impact of the environment on the participants' authentication behaviour and if external factors, such as bystanders and social density [285], affect the results from a usability evaluation.

Authentication Environment (IV2). The extent to which the environment impacts the users' authentication performance and behaviour formed the second independent variable. This investigation is based on two authentication environments:

- **Real World (RW):** Depending on the authentication context (isolated or integrated), the real-world condition depicts a traditional lab environment or a staged ATM authentication scenario. This allows the thesis to a) compare how a simulated ATM scenario matches a virtual replica of the same environment and b) investigate how the usability evaluation results align with the ColorPIN's original study setting [99].
- **Virtual Reality Simulations (VR):** Virtual replicas of the real-world environments (i.e., lab and outdoor) were created to explore the users' authentication performance and behaviour when using ColorPIN [99] in a virtual environment. This allows the thesis to compare the authentication performance and behaviour in VR to a real-world setup and pinpoint potential differences. An additional VR environment representing a simulated public space (cf., *VR ATM Public* in Figure 5.7) was implemented to demonstrate further the use of VR studies for simulated in situ USEC research.

5.4.2.3 Participant Instructions

Traditional storytelling was used to present participants with a realistic authentication scenario – a method where researchers introduce plausible authentication scenarios to increase the ecological validity of user studies [133, 267]. Whilst in *RW Lab* and *VR Lab* the participants were told to imagine they would need to use their credit card to withdraw money, in *integrated*, i.e., *RW ATM*, *VR ATM*, and *VR ATM Public*, the participants had to take out their credit card (a fake one which was provided) and navigate through the ATM user interface (UI) before authenticating. Consequently, the ATM interaction steps in *RW ATM*, *VR ATM*, and *VR ATM Public* consisted of a) inserting the credit card, b) interacting with the ATM according to the given scenario, c) authenticating using ColorPIN, and d) taking the credit card and the money out of the ATM. For *RW Lab* and *VR Lab*, the participants authenticated using ColorPIN in front of a (virtual) desktop screen. This scenario depicts a traditional usability evaluation of authentication prototypes (for example, [99, 245]). The participants were then exposed to ColorPIN and their task was to authenticate. The following ATM scenario was used for all scenarios and environments: “*Your PIN for your credit card is: [ColorPIN]. As a customer, you now want to login to your account using card and PIN code so that you can withdraw [amount of cash]. After entering your PIN, you expect that the system provides you with the requested cash and spits out the money. Please withdraw [amount of cash] now.*”.

The story remained the same across the conditions. The amount of cash the participants had to withdraw (for example, €20) and their ColorPIN changed. The participants were asked

to perform the ATM withdrawal task in a way most similar to how they would do it in the wild. This allows the thesis to receive insights into the participants' input behaviour and shielding strategy when interacting with the (virtual) ATM. Furthermore, it allows the thesis to comment on the participants' perceived differences to ATM interaction in the wild.

5.4.2.4 Study Procedure

The participants' task was to authenticate with ColorPIN using each of the five conditions, which means that participants went through 5 authentication sessions (authentication context \times authentication environment = 4 + *VR ATM Public* = 5). The participants were first introduced to the authentication scenarios, i.e., lab and ATM, and then to the authentication environments, i.e., in the real world and VR. They then underwent a training phase where they were introduced to ColorPIN before collecting data. They then went through a ColorPIN authentication in one of the environments and scenarios. After each authentication session, they reported their sense of presence using the IPQ questionnaire [451] and their perceived workload using the raw NASA-TLX questionnaire [194] (reported for both the authentication and the overall ATM interaction, if applicable). Although the use of presence questionnaires for real-world experiences is debatable [520], the reported sense of presence in the RW conditions is treated as an indication of the participants' experience and sense of being part of an ATM authentication scenario, which is further discussed along with the qualitative feedback collected during the semi-structured interviews.

After filling in the IPQ questionnaire and the NASA-TLX questionnaire, the participants were asked to verbally walk the experimenter through their interactions and tell him about their perceived primary and secondary tasks (structured interview, Section 5.4.5.5 and Appendix D). The structured interview allows the thesis to collect data about the participants' task perception, i.e., if they perceived the security task as their primary or secondary task. Participants were also asked to fill in 5-point Likert scale questions about their feeling of being part of a user study and how the context impacted their behaviour. The study concluded with 1) rankings on the realism of the different authentication contexts and environments and 2) semi-structured interviews (cf., Section 5.4.5.7 and Appendix D). The participants' security knowledge and attitude using the Security Behavior Intentions Scale (SeBIS) [122] and their technological affinity using the Affinity for Technology Interaction (ATI) scale [142] was collected to support and facilitate future replication studies.

5.4.2.5 Data Analysis

Unless otherwise stated, repeated-measures ANOVAs were used. As done in the first study of this chapter, an aligned rank transformation was applied to correct for violations of normalcy,

i.e., ART by Wobbrock et al. [555]. ART-C [126] was used for post-hoc pairwise comparisons, which were corrected using Bonferroni correction. As described in Section 5.4.2.2, there are two baselines in this study: *RW Lab* and *VR Lab*. Two-way repeated-measures ANOVAs where the independent variables were Context (Lab vs ATM) and Environment (RW vs VR) were run. This covered *RW Lab*, *VR Lab*, *RW ATM*, and *VR ATM*. Additional one-way repeated measures ANOVAs when comparing *VR Lab* to *VR ATM* and to *VR ATM Public* were conducted. There were no outliers that had to be removed (for example, measurement errors, data entry errors) – those data points that are suspected of being legitimate to be representative of the population as a whole were kept [373]. Previous work on ColorPIN showed that such outliers can be expected [278]. The structured interviews after each condition were transcribed and coded. The main themes of the interviews are presented in Section 5.4.5.5.

For the semi-structured interviews, the participants' statements were divided into meaningful excerpts. This process resulted in 280 participant statements, which the lead researcher systematically clustered using an affinity diagram. To increase the objectivity of the clustering, a second researcher reviewed it and added tags to clusters that required another iteration. Two researchers, the lead researcher and a second researcher, met to discuss the clustering and resolved any discussion points that came up during the review process. Five themes were identified through this process: 1) Reasoning of Participants' Perceived Realism Ranking, 2) Perceived Differences: ATM Authentication in the Wild, 3) Input Behaviour: The Keyboard, 4) ColorPIN Recall Strategy, 5) General Comments. The most relevant themes for evaluating the use of VR studies for real-world USEC research are reported in Section 5.4.5.7. Reporting the number of participants who shared certain opinions would be inaccurate due to the use of a semi-structured interview approach. Thus, frequencies are only reported where appropriate. Quotes are translated from German to English where necessary.

5.4.3 Apparatus and Implementation

Two software elements (in Unity, C#) were implemented to evaluate the use of VR studies for simulating in situ USEC research and compare the findings from such a VR evaluation to traditional usability testing in the lab. A fully functional 2D ATM UI was implemented for a real ATM, and a fully functional 3D ATM UI for a virtual replica of it (cf., Figure 5.7). For the real-world ATM that employs ColorPIN, a touch screen, cardboard, styrofoam, and metallic spray paint were used. The participants interacted with the ATM in an outdoor environment in front of the research lab (cf., Figure 5.7). Moving the ATM part of the study outdoors contributed to the realism of interacting with an ATM. As done in the original real-world study of ColorPIN [99], a commercial keyboard was attached to the ATM's touch screen, an embedded Samsung Galaxy Tab 3.

The ATM in the real world and VR allowed the participants to navigate through the UI as they

wished. A sensory behaviour was implemented for the real-world ATM to ensure internal validity between the two ATMs. This means that once the participants put in the credit card, the ATM's UI changed based on an external trigger initiated by the experimenter, i.e., Wizard of Oz [87]. This behaviour was fully implemented in VR. An ATM that matches the prototype in the real world [143] was used for the virtual ATM. The real-world study environment was replicated as close as possible, and the branding of a local bank was used to increase the realism of the ATMs.

Due to the baseline conditions and to further validate the use of VR studies for USEC research,



Figure 5.8: In each VR setup, there was a physical keyboard, a greenscreen, and a camera to blend the keyboard and the user's hands into virtuality, similar to McGill et al.'s work [314] and Oculus Passthrough API [106]. Pilot tests were run to position the cameras and tripods to ensure they do not interfere with the participants' interactions during the study.

the lab in the real world (*RW Lab*) was replicated in VR to present the participants with a virtual replica of a similar lab environment in VR (*VR Lab*). Implementing a baseline in both realities and performing comparisons between the real-world conditions (for example, *RW Lab* vs *RW ATM*) and the VR conditions (for example, *VR Lab* vs *VR ATM*) allows the thesis to comment on the use of VR studies for simulated in situ USEC research and to contribute towards the assessment of the validity of using VR studies for USEC research.

For the VR conditions, the Meta Quest 2 VR headset and a Logitech C920 camera were used to bring the real-world keyboard into virtuality. The camera was mounted on a mini tripod or a flexible camera holder, depending on the condition (cf., Figure 5.8). An inferred partial blending was implemented for the transition of the user's virtual hands (rendered through the Oculus Integration SDK [106]) to the user's real hands (rendered through the camera feed). This means that a view of the keyboard, i.e., a CSL wireless slim keyboard, and the user's hands were blended into VR using a chroma key shader and a green screen, similar to McGill et al.'s work [314]. The position of the user's virtual hands was checked, and if the hands did not overlap with the physical keyboard, the virtual hands were rendered, otherwise their real hands. This was piloted in advance of the study to ensure a smooth transition between the participants' virtual and real hands. Adobe's Mixamo avatar library [11] was used for the VR bystanders in *VR ATM Public* (cf., Figure 5.5 and Figure 5.7). Environmental noise was added in the VR ATM environment (i.e., people chatting and birds twittering) to contribute to the fairness of the comparison between *RW ATM* and *VR ATM*, and to contribute to the immersion of the participants into a more vivid environment in *VR ATM Public*.

5.4.4 Demographics

Participants were on average 35.45 years old ($SD = 9.46$). Thirteen participants ($N = 13$) self-identified as male, 7 as female. All participants have used an ATM before, with $M = 2.33$ ($SD = 2.03$) ATM cash withdrawals a month. Almost all ($n = 17$) had previous VR experience, briefly at a demonstration ($n = 10$), a couple of times at a friend's house ($n = 6$), or as part of their job ($n = 1$). The sample's security knowledge and attitude score [122] was $M = 3.18$ ($SD = 1.57$) on a scale from 1 to 5 (*Device Securement*: $M = 4.21$, $SD = 1.36$; *Password Generation*: $M = 3.3$, $SD = 1.59$; *Proactive Awareness*: $M = 2.44$, $SD = 1.28$; *Updating*: $M = 2.87$, $SD = 1.45$) and its technological affinity [142] from 1 to 6 was $M = 3.88$ ($SD = 1.63$).

5.4.5 Results

This section first reports the participants' authentication times, error rates, sense of presence, and perceived workload when authenticating in the real world and VR. Section 5.4.5 then

concludes with the results from the structured interviews, the answers on the 5-point Likert scale questions, and the qualitative feedback from the semi-structured interviews.

5.4.5.1 Authentication Time

The participants' authentication time from the first character entry until the last character entry was collected. This depicts the overall authentication time reported in the original ColorPIN study [99] and is a common approach when evaluating authentication methods [102, 242, 308]. There was a significant main effect of environment, $F_{(1,49)} = 27.00$, $p < 0.05$, $\eta_p^2 = 0.36$, on the participants' authentication times. Authentications were significantly faster in the real world than in VR ($p < 0.05$), with *RW Lab* ($M = 13.28$ s, $SD = 7.76$ s, $Md = 10.04$ s) being significantly faster than *VR Lab* ($M = 20.89$ s, $SD = 8.33$ s, $Md = 20.50$ s), and *RW ATM* ($M = 16.57$ s, $SD = 14.01$ s, $Md = 12.36$ s) being significantly faster than *VR ATM* ($M = 23.85$ s, $SD = 25.32$ s, $Md = 16.45$ s). There was no evidence of a significant main effect of context, $F_{(1,49)} = 0.149$, $p = 0.70$, $\eta_p^2 = 0.003$, and no interaction effect, $F_{(1,49)} = 0.313$, $p = 0.58$, $\eta_p^2 = 0.006$. When comparing *VR ATM Public* ($M = 25.55$ s, $SD = 13.73$ s, $Md = 22.57$ s) to *VR Lab* and *VR ATM*, there was a significant effect of context on the authentication times, $F_{(2,33)} = 3.676$, $p < 0.05$, $\eta_p^2 = 0.18$. Post-hoc pairwise comparisons did not confirm these significant differences ($p > 0.05$). Despite the absence of significance, there was an increase of the mean authentication times in both environments: from *RW Lab* to *RW ATM* (+24.71%), from *VR Lab* to *VR ATM* (+14.17%), and from *VR Lab* to *VR ATM Public* (+22.31%).

Results are visualised in Figure 5.9. Authentication times in *RW Lab*, which was treated as

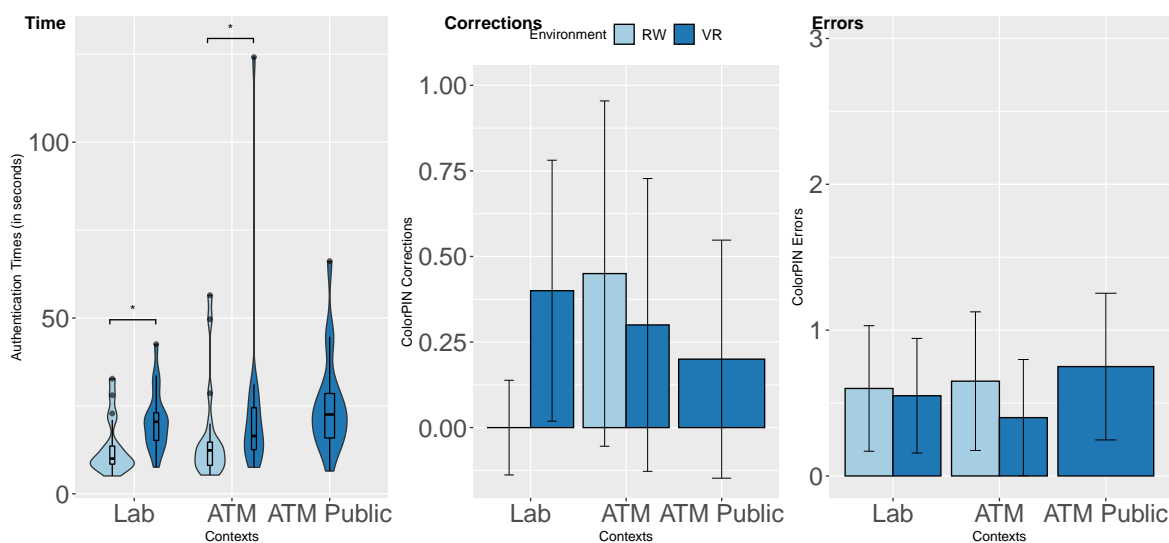


Figure 5.9: Authentications were significantly faster in reality than in VR. Authentications were slower in both environments when performed before withdrawing cash from an ATM. There was no evidence that the number of ColorPIN corrections and errors differ significantly between the conditions. Error bars denote adjusted 95% CIs [336].

the real-world baseline and replicated the original real-world study by De Luca et al. [99], were roughly the same as reported in the original ColorPIN work ($M = 13.28$ s and $SD = 7.76$ s vs $M = 13.33$ s and $SD = 1.74$ s) [99].

5.4.5.2 Error Rate (PIN Corrections and Incorrect Entries)

The error rate is reported as *corrections*, the number of corrections before submitting a ColorPIN, and as *errors*, the number of incorrect ColorPIN entries with a maximum of three tries to authenticate correctly. There was no main effect of the environment, the context, and no interaction effect on the participants' number of corrections. Table 5.3 shows the statistical analysis, including the F-ratios, effect sizes, means, and standard deviations. Corrections were lowest in *RW Lab* with no corrections at all, followed by *VR ATM Public* ($M = 0.20$, $SD = 0.68$), *VR ATM* ($M = 0.30$, $SD = 0.90$), *VR Lab* ($M = 0.40$, $SD = 0.73$), and *RW ATM* ($M = 0.45$, $SD = 1.07$). There was no evidence that the number of incorrect entries differed significantly between the conditions. The values were *RW Lab* ($M = 0.60$, $SD = 1.11$), *RW ATM* ($M = 0.65$, $SD = 1.07$), *VR Lab* ($M = 0.55$, $SD = 0.92$), *VR ATM* ($M = 0.40$, $SD = 0.92$), and *VR ATM Public* ($M = 0.75$, $SD = 0.99$). Results are visualised in Figure 5.9. The statistical analysis is summarised in Table 5.3.

Cash Withdrawal Performance. The participants' cash recall performance was analysed using Cochran's Q test. The cash recall performance depicts the extent to which the participants could recall the amount of cash they were supposed to withdraw. Insights into this metric allow the thesis to investigate how the different authentication scenarios impact the participants' memorability (i.e., their primary task performance). The participants' primary task performance, i.e. correctly recalling the amount of cash they had to withdraw, was not statistically significant between the conditions, $\chi^2(4) = 2.194$, $p = 0.70$. There were two participants in *RW Lab*, four in *RW ATM*, and five each in *VR Lab*, *VR ATM*, and *VR ATM Public* who were unable to correctly recall the amount of cash to withdraw.

5.4.5.3 Sense of Presence (IPQ) and Perceived Workload (NASA-TLX)

Table 5.5 provides an overview of the means, the standard deviations, and the statistical analysis of the IPQ and NASA-TLX values, featuring the subscales 1) sense of being there (PRES), 2) spatial presence (SP), 3) involvement (INV), 4) experienced realism (REAL), and 5) the raw NASA-TLX scores. Overall, the participants' sense of presence was significantly higher in *RW ATM*, *VR ATM*, and *VR ATM Public* than in *RW Lab* and in *VR Lab* ($p < 0.05$), and significantly higher in *VR ATM Public* than in *VR Lab* ($p < 0.05$). A more nuanced analysis on the level of each subscale is reported in Table 5.4 and Table 5.5, which followed

Table 5.3: The statistical analysis, including means, standard deviations, F-ratios, effect size, and p-values of the participants' authentication times (in seconds), number of corrections, and number of errors. $p < 0.05$ highlighted.

Measure (Two-way RM ANOVA)	(1) RW Lab	(2) RW ATM	(3) VR Lab	(4) VR ATM	Context (Lab/ATM)	Environment (RW/VR)	Context×Environment	p<0.05
Authentication Time	13.28 (7.76)	16.57 (14.01)	20.89 (8.33)	23.85 (25.32)	$F_{(1,49)} = 0.149, p = 0.70, \eta_p^2 = 0.003$	$F_{(1,49)} = 27.00, p < 0.005, \eta_p^2 = 0.36$	$F_{(1,49)} = 0.313, p = 0.58, \eta_p^2 = 0.006$	1-3,2-4
Number of Corrections	0 (0)	0.45 (1.07)	0.40 (0.73)	0.30 (0.90)	$F_{(1,57)} = 0.418, p = 0.52, \eta_p^2 = 0.007$	$F_{(1,57)} = 0.269, p = 0.61, \eta_p^2 = 0.005$	$F_{(1,57)} = 0.516, p = 0.48, \eta_p^2 = 0.009$	n/a
Number of Errors	0.60 (1.11)	0.65 (1.07)	0.55 (0.92)	0.40 (0.92)	$F_{(1,57)} = 0.420, p = 0.52, \eta_p^2 = 0.007$	$F_{(1,57)} = 0.157, p = 0.69, \eta_p^2 = 0.003$	$F_{(1,57)} = 0.650, p = 0.42, \eta_p^2 = 0.01$	n/a
Measure (One-way RM ANOVA)	(3) VR Lab	(4) VR ATM	(5) VR ATM Public	Context	p<0.05			
Authentication Time	20.89 (8.33)	23.85 (25.32)	25.55 (13.73)	$F_{(2,33)} = 3.676, p < 0.05, \eta_p^2 = 0.18$				
Number of Corrections	0.40 (0.73)	0.30 (0.90)	0.20 (0.68)	$F_{(2,38)} = 0.73, p = 0.49, \eta_p^2 = 0.04$				
Number of Errors	0.55 (0.92)	0.40 (0.92)	0.75 (0.99)	$F_{(2,38)} = 1.40, p = 0.259, \eta_p^2 = 0.07$				

Table 5.4: Means and standard deviations of the IPQ scores, the NASA-TLX scores, and the participants' responses on the 5-point Likert scale questions (1 = strongly disagree, 5 = strongly agree). The statistical analysis follows the description in Section 5.4.2.5. $p < 0.05$ highlighted.

IPQ Scores (Two-way RM ANOVA)	Context (Lab/ATM)				Environment (RW/VR)		Context×Environment		p<0.05
	(1) RW Lab	(2) RW ATM	(3) VR Lab	(4) VR ATM	(1,57)	(1,57)	(1,57)	(1,57)	
Sense of being there (PRES)	1.55 (2.16)	5.25 (0.70)	3.1 (1.04)	4.55 (1.02)	$F_{(1,57)} = 47.89, p < 0.05, \eta_p^2 = 0.46$	$F_{(1,57)} = 5.323, p < 0.05, \eta_p^2 = 0.09$	$F_{(1,57)} = 18.03, p < 0.05, \eta_p^2 = 0.24$	$F_{(1,57)} = 20.11, p < 0.05, \eta_p^2 = 0.26$	1-2,3-4
Spatial Presence (SP)	1.52 (1.49)	4.48 (1.72)	3.52 (1.78)	4.27 (1.46)	$F_{(1,57)} = 56.07, p < 0.05, \eta_p^2 = 0.50$	$F_{(1,57)} = 13.11, p < 0.05, \eta_p^2 = 0.19$	$F_{(1,57)} = 20.11, p < 0.05, \eta_p^2 = 0.26$	$F_{(1,57)} = 20.11, p < 0.05, \eta_p^2 = 0.26$	1-2
Involvement (INV)	1.91 (1.91)	2.48 (1.97)	3.35 (1.70)	3.69 (1.72)	$F_{(1,57)} = 3.05, p = 0.09, \eta_p^2 = 0.05$	$F_{(1,57)} = 40.80, p < 0.05, \eta_p^2 = 0.42$	$F_{(1,57)} = 0.19, p = 0.66, \eta_p^2 = 0.003$	$F_{(1,57)} = 0.19, p = 0.66, \eta_p^2 = 0.003$	n/a
Realism (REAL)	1.76 (1.96)	3.46 (2.04)	2.14 (1.57)	3.04 (1.71)	$F_{(1,57)} = 45.058, p < 0.05, \eta_p^2 = 0.44$	$F_{(1,57)} = 0.0017, p = 0.966, \eta_p^2 = 0.00003$	$F_{(1,57)} = 6.830, p < 0.05, \eta_p^2 = 0.11$	$F_{(1,57)} = 6.830, p < 0.05, \eta_p^2 = 0.11$	1-2,3-4
Overall Presence Score	1.70 (1.82)	3.67 (2.06)	3.05 (1.80)	3.77 (1.67)	$F_{(1,57)} = 69.403, p < 0.05, \eta_p^2 = 0.55$	$F_{(1,57)} = 18.151, p < 0.05, \eta_p^2 = 0.24$	$F_{(1,57)} = 18.147, p < 0.05, \eta_p^2 = 0.24$	$F_{(1,57)} = 18.147, p < 0.05, \eta_p^2 = 0.24$	1-2,3-4
NASA-TLX Scores (Two-way RM ANOVA)	(1) RW Lab	(2) RW ATM	(3) VR Lab	(4) VR ATM	Context (Lab/ATM)	Environment (RW/VR)	Context×Environment		p<0.05
ColorPIN only	31.79 (30.48)	34.35 (32.03)	31.71 (29.49)	33.04 (30.91)	$F_{(1,57)} = 0.491, p = 0.49, \eta_p^2 = 0.009$	$F_{(1,57)} = 0.803, p = 0.37, \eta_p^2 = 0.014$	$F_{(1,57)} = 0.0002, p = 0.99, \eta_p^2 = 0.0000033$	$F_{(1,57)} = 0.0002, p = 0.99, \eta_p^2 = 0.0000033$	n/a
ATM + ColorPIN	n/a	33.21 (28.60)	n/a	35.17 (30.29)	n/a	$F_{(1,19)} = 0.33, p = 0.57, \eta_p^2 = 0.017$	n/a	n/a	n/a
5-Point Likert Scale Questions	(1) RW Lab	(2) RW ATM	(3) VR Lab	(4) VR ATM	(5) VR ATM Public	Friedman			p<0.05
Feeling of being part of a user study	3.70 (1.23)	2.95 (1.16)	3.35 (1.31)	3.15 (1.11)	3.05 (1.16)	$\chi^2(4) = 12.670, p < .05$			1-2;1-4;1-5
Awareness of the experimenter	2.85 (1.31)	2.70 (1.31)	2.55 (1.36)	2.65 (1.28)	2.50 (1.43)	$\chi^2(4) = 0.874, p = 0.928$			n/a
Impact of experimenter's presence on performance	1.50 (0.59)	1.60 (0.92)	1.40 (0.58)	1.45 (0.74)	1.55 (0.74)	$\chi^2(4) = 1.538, p = 0.82$			n/a
Impact of experimenter's presence on behaviour	1.30 (0.46)	1.55 (0.74)	1.40 (0.38)	1.60 (0.92)	1.55 (0.74)	$\chi^2(4) = 4.155, p = 0.385$			n/a
Impact of the secondary task on the primary	2.55 (1.24)	2.95 (1.40)	2.85 (1.35)	2.75 (1.22)	2.85 (1.24)	$\chi^2(4) = 1.957, p = 0.744$			n/a
Impact of the primary task on the secondary	1.9 (1.18)	1.95 (1.07)	1.75 (0.89)	1.90 (0.83)	1.75 (0.77)	$\chi^2(4) = 1.784, p = 0.775$			n/a

Table 5.5: Results of the one-way RM ANOVA on the IPQ and NASA-TLX scores of the VR conditions. $p < 0.05$ highlighted. The $p < 0.05$ column shows pairwise comparisons.

IPQ Scores					
(One-way RM ANOVA)	(1) VR Lab	(2) VR ATM	(3) VR ATM Public	Context(Lab/ATM/Public)	$p < 0.05$
Sense of being there (PRES)	3.10 (1.04)	4.55 (1.02)	4.85 (1.24)	$F_{(2,38)} = 22.41, p < 0.05, \eta_p^2 = 0.54$	1-2;1-3
Spatial Presence (SP)	3.52 (1.78)	4.27 (1.46)	4.60 (1.18)	$F_{(2,38)} = 8.880, p < 0.05, \eta_p^2 = 0.32$	1-2;1-3
Involvement (INV)	3.35 (1.70)	3.69 (1.72)	4.14 (1.61)	$F_{(2,38)} = 3.822, p < 0.05, \eta_p^2 = 0.17$	1-3
Realism (REAL)	2.14 (1.57)	3.04 (1.71)	3.03 (1.77)	$F_{(2,38)} = 8.71, p < 0.05, \eta_p^2 = 0.31$	1-2;1-3
Overall Presence Score	3.05 (1.80)	3.77 (1.67)	4.04 (1.64)	$F_{(2,38)} = 19.275, p < 0.05, \eta_p^2 = 0.50$	1-2;1-3

NASA-TLX Scores					
(One-way RM ANOVA)	(1) VR Lab	(2) VR ATM	(3) VR ATM Public	Context (Lab/ATM/Public)	$p < 0.05$
ColorPIN only	31.71 (29.49)	33.04 (30.91)	40.04 (30.03)	$F_{(2,38)} = 2.65, p = 0.084, \eta_p^2 = 0.12$	n/a
ATM + ColorPIN	n/a	35.17 (30.29)	40.88 (27.78)	$F_{(1,19)} = 1.48, p = 0.24, \eta_p^2 = 0.07$	n/a

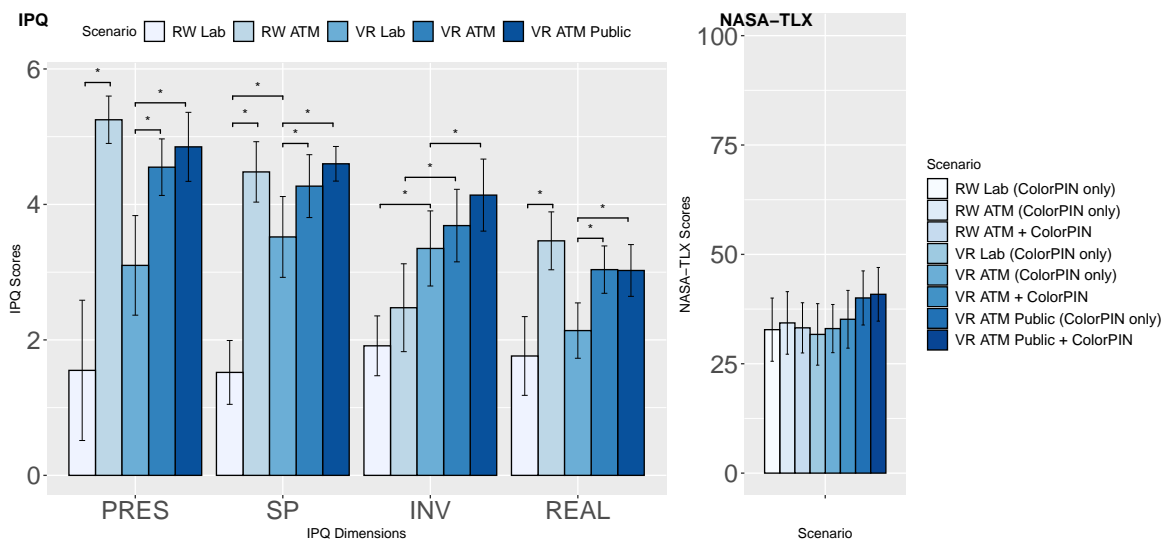


Figure 5.10: The participants’ sense of presence was significantly higher in *RW ATM*, *VR ATM*, and *VR ATM Public* compared to *RW Lab* and *VR Lab*. There was no evidence that the mean raw NASA-TLX values differed significantly between the conditions.

the approach described in Section 5.4.2.5. There were statistically significant main effects in all IPQ subscales, with the participants’ sense of being part of an ATM authentication scenario, spatial presence, and realism being statistically significantly higher in the simulated in situ ATM experiences than in *RW Lab* and in *VR Lab*. For the NASA-TLX values, there was no evidence that the participants’ perceived workload differed significantly between the conditions ($p > 0.05$). Figure 5.10 and Table 5.5 provide an overview of the statistical analysis. Table D.1 in Appendix D shows the individual NASA-TLX subdimensions.

5.4.5.4 5-Point Likert Scale Questions

The participants were asked on 5-Point Likert scales (1 = strongly disagree, 5 = strongly agree) if a) they felt being part of a laboratory study during the authentication, b) they were aware of the experimenter, c) the experimenter’s presence impacted their performance and

behaviour, and d) recalling the PIN made it more challenging to complete the other cash withdrawal steps, and vice versa. A Friedman test with post-hoc Wilcoxon signed-rank tests (Holm-Bonferroni corrected [206]) resulted in a significant difference between the conditions and the participants' feeling of being part of a laboratory study, $\chi^2(4) = 12.670$, $p < 0.05$. The participants' feeling of being part of a laboratory study was statistically significantly less in *RW ATM* ($Z = -2.375$, $p = 0.018$), *VR ATM* ($Z = -2.484$, $p = 0.013$), and *VR ATM Public* ($Z = -2.365$, $p = 0.018$) compared to *RW Lab*. Table 5.4 shows the means and standard deviations. All in all, the main takeaway from the 5-Point Likert scale scores and the structured interviews in Section 5.4.5.5 is that there was evidence that the participants' "feeling of being part of a user study" was stronger in *RW Lab* than in the other conditions, which highlights the affordances of staged in situ environments and VR studies to contribute towards perceived realism of user studies when evaluating USEC prototypes.

5.4.5.5 Structured Interviews

Structured interviews were conducted after each authentication procedure to learn about the tasks the participants perceived as their primary and secondary tasks. Additionally, the participants were asked about potential differences to their prior ATM interaction experience. The full questionnaire is available in Appendix D, with the statements highlighted with (*) addressed in Section 5.4.5.4. The analysis is summarised in Table 5.4.

Topic 1: Primary and Secondary Task Perception. When isolating the ColorPIN authentication system from an actual production task (*RW Lab* and *VR Lab*), there was a notable amount of participants who perceived entering their ColorPIN as their primary task (*RW Lab*: $n = 12$; *VR Lab*: $n = 17$). When ColorPIN was part of an overall production task where participants performed a task before and after the actual authentication, i.e., in *RW ATM*, *VR ATM*, and *VR ATM Public*, the participants mentioned less often that they perceived entering their ColorPIN as their primary task (*RW ATM*: $n = 9$; *VR ATM*: $n = 8$; *VR ATM Public*: $n = 8$). This means that whilst the three ATM conditions did a slightly better job in providing participants with a realistic authentication context than *RW Lab* and *VR Lab*, the ATM environments were still unable to provide fully realistic authentication experiences. Some participants still perceived the authentication as their primary task, which deviates from the real world (cf., [267, 438]).

Topic 2: Participants' Perceived Differences to Real-World ATM Experiences. When the participants were asked about their perceived differences to a real-world ATM withdrawal in their everyday life, there was one comment that frequently appeared across the conditions. The participants emphasised that the USEC prototype, i.e., ColorPIN, differed

from the ATM authentication system they are familiar with (i.e., traditional 4-digit PIN authentication, $n = 9$ for *RW Lab*, $n = 15$ for *RW ATM*, $n = 12$ for *VR Lab*, $n = 14$ for *VR ATM*, $n = 10$ for *VR ATM Public*). This finding implies that when researchers study novel USEC prototypes, it remains challenging to present the participants of user studies with highly realistic real-life scenarios due to the prototype's novelty, which is further discussed in Section 5.4.6.2.

In *RW Lab* and *VR Lab*, the participants mentioned that they were sitting in front of a PC ($n = 8$ for *RW Lab*, $n = 4$ for *VR Lab*) and that this led to a different experience than being part of an ATM interaction scenario ($n = 10$ for *RW Lab*, $n = 12$ for *VR Lab*). About half of the participants ($n = 11$) mentioned that in *RW ATM* the fidelity and location of the ATM deviated from an ATM withdrawal scenario in the wild ($n = 7$ for *VR ATM*, $n = 2$ for *VR ATM Public*). Some participants mentioned that using VR (for example, putting on the headset) is something they would not do in real life ($n = 6$ for *VR Lab*, $n = 5$ for *VR ATM*, $n = 4$ for *VR ATM Public*), which means that the additional VR hardware used to simulate reality can impact the participants' perception of reality. Some participants ($n = 7$) mentioned that they would take precautions when they see other people next to them, which they reported not having done in the study.

5.4.5.6 Perceived Realism Ranking

The participants were asked to rank the different conditions based on their perceived realism (1 = best, 5 = worst). Raw scores were multiplied by a weight factor ($\times 5$ for rank 1, $\times 4$ for rank 2, etc.) and then summed up to compute weighted scores. *RW ATM* achieved the highest score (85) with *VR ATM Public* (80) on rank two and *VR ATM* on rank three (70). The baseline conditions were perceived as the least realistic ATM contexts, with *RW Lab* slightly more realistic (37) than *VR Lab* (31). The ranking shows that in both reality and virtuality, the replicated ATMs improved the participants' perceived authentication realism, which aligns with the responses about their feeling of being part of a user study (cf., Table 5.4).

5.4.5.7 Semi-structured Interviews

Semi-structured interviews were conducted at the end of the study to capture the rich nuances of the participants' experiences more qualitatively. The data has been analysed as described in Section 5.4.2.5. The full questionnaire is available in D in Appendix D. Follow-up questions were asked when needed.

Theme 1: Reasoning of Participants' Perceived Realism Ranking. The participants perceived *RW ATM* as most similar to an ATM withdrawal experience in the wild, as

reported in Section 5.4.5.6. P1 voiced that *“the real-world ATM was the most realistic because there was something in front of me, I could really feel the card.”* (P1). Others mentioned that they perceived the real-world ATM as the most realistic one because *“you cannot get closer to that where you have the ATM 1:1 in front of you”* (P11). There were discussions around the realism of an ATM scenario where other people are close to the person who interacts with the ATM. P19 perceived *RW ATM* and *VR ATM* as more realistic than *VR ATM Public* because they had not experienced an ATM authentication scenario where other people were relatively close to the ATM. Others perceived *VR ATM Public* as more realistic than *VR ATM* and explained this around the fact that *“ATMs are usually at locations where much more is going on”* (P4). For both *RW Lab* and *VR Lab*, the participants mentioned that they felt *“like playing a game; you sit in front of a keyboard and enter a PIN”* (P14) and that the baseline conditions do not represent a realistic ATM withdrawal scenario: *“I never withdraw cash in front of a desktop monitor”* (P1).

Theme 2: Perceived Differences: ATM Authentication in the Wild. When asked about any differences to an ATM authentication in the wild, the participants voiced that they were familiar with the actions they had to do: *“the actions I had to do were very similar to the real world; take the card, put in the card, take it out – it is the same mechanism”* (P18). However, they frequently brought up that the USEC prototype, i.e., ColorPIN [99], did not represent a realistic authentication scenario and how they authenticate in real life: *“it was quite realistic, I mean you enter a different PIN - the ColorPIN - which is different to the real world”* (P14). This was mentioned for both the staged real-world ATM scenario and for the two VR ATM scenarios. For *VR ATM* and *VR ATM Public*, the participants hinted at the lack of haptic feedback: *“in VR all the haptics were missing, and also to identify the distance when interacting with the touch screen”* (P12). Some participants, for example, P6, mentioned that they probably need more exposure to VR and that VR’s novelty led to a different feeling compared to their prior real-world ATM withdrawal experience: *“the difference is at the beginning to get familiar with the technology, to see where the hands are and when you interact with the keyboard.”* (P6).

Theme 3: Input Behaviour: The Keyboard. About half of the participants used touch typing when authenticating using ColorPIN, independent of the environment. Some participants voiced that they usually use touch typing when providing keyboard input, but their interaction behaviour was different in the study: *“[I] only used one finger because that is how I do it when interacting with an ATM”* (P10) and *“like I’d type on a traditional ATM, there wasn’t much difference”* (P6). Interestingly, P1 mentioned that they only used touch typing in *RW Lab* and *VR Lab* because these two environments provided them with a feeling of being part of a workplace rather than an ATM environment. The importance of contextual

factors and user behaviour in USEC prototype research is further discussed in Section 5.4.6.1.

5.4.6 Discussion

When conducting, for the first time, simulated *in situ* research on a USEC prototype in VR and a staged real-world setting, it was found that both RW and VR exhibit similar evaluation patterns when comparing *isolated* with simulated *in situ* authentications. The participants' sense of presence increased significantly from the laboratory settings, i.e., *RW Lab* and *VR Lab*, to the ATM environments, i.e., *RW ATM*, *VR ATM*, and *VR ATM Public*, in both the real-world environment and virtuality. The participants felt less part of a user study in all ATM conditions compared to *RW Lab*, the de facto standard evaluation methodology when assessing novel authentication prototypes. This suggests, together with the participants' qualitative feedback, that applying VR studies to simulate in situ authentication research contributes to a considerable high authentication realism. Furthermore, authentication times increased by 24.71% from *RW Lab* to *RW ATM*, by 14.17% from *VR Lab* to *VR ATM*, and by 22.31% from *VR Lab* to *VR ATM Public*. The increased authentication times and the participants' comments during the interviews suggest that in situ evaluations impact USEC prototypes usability evaluation findings. There was no evidence that authentications using ColorPIN are more or less demanding in VR than in the real world, which supports the use of VR replicas for real-world USEC research and further adds to the validation of using VR studies for human-centred usability and security research.

Although the use of VR studies for simulating in situ research achieved promising results, some typical user behaviour in the wild, such as shielding PIN entries [24, 100], was not present in this study, which is further discussed in Section 5.4.6.2.

KEY LESSON 1

VR studies enable simulating in situ real-world USEC research on VR replicas, even in contexts that are often infeasible to research in the wild. Yet, it is important to acknowledge potential behavioural differences between the participants in (VR) lab studies and those in more naturalistic in-the-wild scenarios.

5.4.6.1 There Is More to Context Than Authentication

USEC prototypes that are proposed and designed for specific contexts, for example, for public displays [99, 242, 245] or mobile devices [41, 239, 243], should, if feasible, be evaluated in their intended usage scenario. Johnson [226] argued that the conventional usability laboratory is not able to adequately simulate conditions in the wild and cannot “*provide for the wide range of competing activities and demands on users that might arise in a natural setting*” [226].

However, although evaluations in the lab are affordable, often relatively easy to conduct, and allow for quick prototype iterations and evaluations, there are often no follow-up studies in the wild, as noted in the expert interviews in chapter 3. As a result, it remains unclear how research prototypes can be applied and used in their intended usage context. In the simulated in situ study reported in this chapter, several participants voiced that they used touch typing in *RW Lab* and *VR Lab*, but not in *RW ATM*, *VR ATM*, and *VR ATM Public*, and explained this around the fact that they perceived *RW Lab* and *VR Lab* as “sitting in front of a PC at work”. Future research is encouraged to contribute to usability evaluations in different contexts as they do, as evidenced by the findings in this chapter, impact the users’ performance, sense of presence, and behaviour. In situ research is particularly important because user behaviour is a key factor in security failures [435] and the users’ behaviour in the lab might not depict their behaviour in more realistic settings in the wild. If USEC prototypes are studied in traditional (physical or virtual) laboratory settings only, as simulated with *RW Lab* and *VR Lab*, the USEC research community will not be able to identify and capture the causes of undesirable user behaviour.

KEY LESSON 2

Context is a key factor when evaluating USEC prototypes and can impact a prototype’s usability evaluation results and how users interact and behave. Using VR studies to replicate real-world scenarios that are hard to research using other means contributes to more realistic and effective USEC research than more traditional laboratory studies.

5.4.6.2 Despite the Use of VR Studies, Achieving High Realism is (Still) Hard

Although the virtual ATM scenarios outperformed the VR baseline, for example, resulting in a higher sense of presence and perceived realism (cf., Section 5.4.5.3 and 5.4.5.6), eliciting in-the-wild user behaviour using VR studies remains a challenge. This was apparent in the study as follows: participants mentioned that both the real-world ATM and the two VR ATM replicas provided them with a high level of realism and that all three setups came close to their prior in-the-wild ATM interaction experience. However, the USEC prototype, ColorPIN [99], impacted their perceived realism and behaviour. There was a consensus that the lab setup did an excellent job in replicating an ATM scenario, but that the novelty of the prototype made them realise they are still in a user study and that there is a mismatch to an ATM authentication in the wild (cf., Section 5.4.5.5). Whilst it can be argued that the novelty effect can be reduced by replacing ColorPIN [99] with a more traditional authentication system, doing this would hinder researchers from drawing any conclusions on the usability of novel USEC prototypes and would restrict such user studies to already deployed systems.

The use of VR studies contributes to more realistic simulations of in situ USEC research than traditional lab studies are capable of, especially when studying security-sensitive contexts that are otherwise challenging to access in real life. However, the study's results show that when novel prototypes are introduced, the participants will likely not behave as they would in the wild. None of the participants shielded their PIN entry. In contrast, observational studies in the real world showed that about a third of ATM users usually apply PIN entry shielding [100]. This implies that the use of VR studies for simulated in situ USEC research cannot fully replace studies in the wild. Yet, as evidenced by the results, the use of VR studies can advance research in the lab and support researchers in studying scenarios that are challenging (and sometimes even infeasible) to investigate in the wild.

KEY LESSON 3

Staged real-world environments and VR replicas contribute to a high sense of authentication realism. However, evaluating novel USEC prototypes in highly ecologically valid contexts remains challenging due to the nature of user studies and the novelty of research prototypes.

5.4.7 Limitations

Some research decisions as part of this study are worth discussing. First, the study was conducted within an ATM authentication scenario, a context that is challenging to study in the real world [100]. Whilst this shows that using VR studies enables researchers to study USEC prototypes in security-sensitive contexts in great detail, the research in this thesis section cannot provide an exhaustive list of advantages and disadvantages of using VR studies for the full breadth of USEC research. Other contexts and USEC prototypes, for example, biometric airport systems [436] or authentication systems for doors [323], are worth investigating to further establish the use of VR studies for simulated in situ usability evaluations and exploit VR's full potential for USEC research. Furthermore, Volkamer et al. [529] highlighted differences in people's ATM interaction behaviour across countries. Future work might want to run a cross-country study of a VR-based in situ research approach to compare the results with the findings reported in this PhD and with ColorPIN's original study results [99].

Finally, as technology improves and the society becomes more acquainted with VR, more advanced VR headsets and more realistic authentication environments will likely increase the participants' perceived realism when interacting with VR replicas of real-world USEC prototypes. The study in this thesis was conducted in 2021 using the Meta Quest 2, which has to be noted and considered when aiming to replicate the findings. Advanced VR headsets (for example, the Meta Quest Pro), increased display resolutions, more extensive field of views,

and better keyboard support (for example, Meta's Tracked Keyboard SDK⁴) may contribute to even more realistic (virtual) study environments and interaction experiences.

5.4.8 Conclusion

For the first time, the second study of this chapter explored the suitability of VR studies for simulating in situ USEC research. It evaluated ColorPIN [99], a novel USEC prototype, in two authentication contexts (in the lab and in public) using two authentication environments (in the real world and VR). The study found different usability evaluation results between a traditional lab environment and a more realistic ATM authentication environment in both the real world and VR. Interestingly, the participants felt less being part of a user study in all ATM scenarios compared to the lab environment in the real world, which suggests that the use of VR studies contributes to more realistic USEC research than studies in more traditional (physical) lab environments. Based on the investigation of using VR studies for simulating in situ USEC research, the second part of this chapter showcased how VR supports and facilitates USEC research in contexts that are otherwise challenging (or even infeasible) to study in the real world. The study findings emphasise the impact of different authentication contexts on the usability evaluation findings of USEC prototypes and highlight the need for the USEC community to study novel prototypes in their intended usage scenario rather than in physical, often limited, laboratory environments.

5.4.8.1 Contribution to Research Question 4 (RQ₄)

The VR-based in situ study contributes answers to the following research question:

RQ₄: Can substitutional in situ studies using VR provide a bridge over the methodological gap between lab and field studies?

In a similar vein as the shoulder surfing study (cf., Section 5.3), the simulated in situ study and its results are based on staged authentication scenarios, but this time in the real world and VR. Staging such authentication scenarios means, as evidenced by the results, that there is still a gap to naturalistic research in the wild. However, by using VR studies to simulate in situ research, this thesis demonstrates how authentication scenarios can be studied in environments beyond the traditional ones in physical laboratory settings. As shown in Section 5.4, the use of VR studies allows researchers to present user study participants with a lab environment similar to how research in the lab is currently conducted, i.e., a traditional lab or office space. This is depicted through the *VR Lab* condition in Figure 5.7. However, if beneficial for and required

⁴<https://developer.oculus.com/documentation/unity/tk-overview/>, last accessed 22/01/2023

by the research questions, VR studies enable immersing the participants into highly realistic USEC scenarios and gradually adding contextual elements into the user study environments (for example, additional bystanders as simulated in *VR ATM Public*; cf., Figure 5.7).

Most importantly, utilising VR studies allows researchers to immerse the participants into realistic authentication scenarios, which is important when studying the usability of USEC prototypes as in real life “security as a task is secondary to a main purpose” [267]. The study’s findings in this section imply that, ‘yes’, the use of VR studies for simulating real-world research contributes towards providing a bridge over the methodological gap between lab and field studies. Despite some differences to observations in the real world (for example, the participants’ lack of shielding behaviour), VR studies revolutionise how USEC research involving prototypes is currently being conducted in physical labs by allowing researchers to study USEC prototypes and the participants’ behaviour in plausible real-world scenarios.

5.5 Chapter Conclusion

This chapter explored, for the first time, the use of VR studies for two core research topics in USEC: human shoulder surfing research (cf., Section 5.3), one of the most common threat models when evaluating the security of novel USEC artefacts; and simulated in situ USEC research (cf., Section 5.4), a long-lasting research challenge due to the financial, ethical, and legal constraints when aiming to conduct highly ecologically valid USEC research in the field.

Although the two studies presented in this chapter contribute to the applicability and validation of the use of VR studies for simulating real-world research, their main goal was to advance research in the USEC field and showcase the potential of VR studies for USEC research. The VR shoulder surfing study in Section 5.3 has introduced two novel shoulder surfing research methods: non-immersive *3D Observations* and immersive *VR Observations*. It evaluated both shoulder surfing methods against more traditional VR-based 2D recordings, which have been previously validated against real-world recordings in Section 4.7 and Section 4.8 of chapter 4. The use of *VR Observations* provides a novel research method to evaluate the security of USEC prototypes and learn more about the participants’ behaviour when aiming to observe other people’s input. Building upon the comments by Krol et al. [267] and Aviv et al. [27], who argued for live shoulder surfing research instead of using recorded videos, the use of VR studies combines the strengths of live observations with the strengths of the consistency across study samples due to the nature of VR recordings.

All in all, the shoulder surfing study in Section 5.3 has demonstrated how the use of VR studies enables researchers to access and study different authentication environments, opening the door for the research community to use VR’s unique affordances to advance human-centred security research through more realistic shoulder surfing studies and security evaluations.

The simulated in situ study in Section 5.4 has shown that contextual factors, such as the medium (reality or virtuality) and the environment (lab or public), impact the results from usability evaluations of USEC research prototypes. The findings in Section 5.4 are of particular relevance as the simulated in situ investigations have found empirical evidence of the impact of the scenario on a USEC prototype’s usability results. Simulating security as a secondary task that aligns with how security is treated in real life [267,437] has been tackled by several experts in the field (for example, by role-playing scenarios or storytelling [443,455]). However, VR studies provide researchers with a novel research method that allows them to simulate and immerse their user study participants into various plausible real-world scenarios.

5.5.1 Research Question 4 (RQ₄)

The overarching findings from this chapter, including the VR shoulder surfing study and the simulated in situ study, are brought together here in the context of the fourth research question:

RQ₄: Can substitutional in situ studies using VR provide a bridge over the methodological gap between lab and field studies?

As discussed in the individual studies in Section 5.3.8.1 and Section 5.4.8.1, this chapter concludes that, ‘yes’, the use of VR studies for simulating real-world USEC research contributes towards providing a bridge over the methodological gap between lab and field studies. It is important to note that the ecological validity discussions in the expert interviews (cf., chapter 3) around lab studies and field studies should not be treated as categorical. Instead, studies of those types – including the use of VR studies for simulating real-world research – are likely to be found on a continuum, similar to how mixed reality experiences exist on a reality-virtuality continuum [330] (cf., Figure 5.11). Therefore, expanding on Mäkelä et al.’s comments about VR field studies “being closer to field studies in ecological validity,

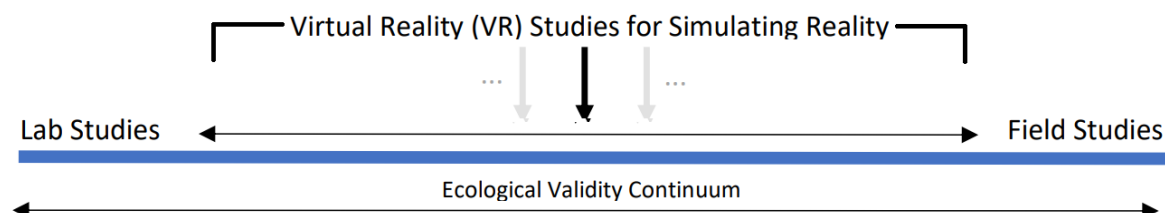


Figure 5.11: A one-dimensional visualisation of the location of VR studies on a theoretical ecological validity continuum, ranging from fully controlled lab studies to more naturalistic field studies. The visualisation is based on [291, Figure 6] and a SOUPS 2022 Lightning talk [301]. It showcases the application continuum of VR studies, as demonstrated with *VR Lab*, *VR ATM*, and *VR ATM Public* in Section 5.4 of this thesis (cf., Figure 5.7).

and closer to lab studies with regards to the required effort” [291], this thesis argues that the ecological validity proximity of VR studies to the existing research methods, such as lab studies and field studies, is highly contextual and study-specific. For example, *VR Lab*, an example of the use of VR studies for simulating a real-world lab environment, exhibits more characteristics of a traditional lab environment than of a field environment. In contrast, *VR ATM Public* is closer to an authentication scenario in the wild than *VR Lab*. However, the *VR Lab* condition is particularly interesting because VR studies can be used for simulating real-world lab environments without necessarily contributing towards ecological validity. Depending on the nature of a researcher’s interests and their overarching research aim (for example, is a naturalistic scenario required? or is a simplistic virtual lab environment sufficient for initial usability testing?), VR studies can be used for various research interests and research questions. It is important to stress here that whilst high ecological validity is important, achieving the highest possible ecological validity might not always be a researcher’s aim when, for example, conducting early usability testing or pinpointing differences between simulated lab evaluations and field evaluations.

The following three lessons synthesise the key takeaways from the research in this chapter:

5.5.1.1 Lesson 1: Account For Real-World Factors if They are of Relevance and Consider How the Corresponding Research Findings Transfer to the Real World

VR studies greatly advance shoulder surfing research and simulated in situ USEC research. They enable researchers to get insights into the participants’ observation strategies and how people behave when interacting with USEC research artefacts in their intended use case. However, the results from such VR studies highly depend on how well reality is emulated. Researchers are encouraged to control for proxemics [186] in virtual environments if social factors are relevant to their research question(s). Contrary to prior work that found that the users’ perception of personal space in the real world is similar to that in a virtual environment [30,197], the shoulder surfing study in Section 5.3 revealed that when participants optimise their shoulder surfing observations, social factors and the proximity to the user authenticating lose relevance and may even be ignored in VR. Considering the existing VR community discussions is important when aiming for close-to-reality shoulder surfing behaviour and authentication experiences in virtual environments. For example, Slater [475] argued that the effect of both “place illusion” and “plausibility illusion” (i.e., “essentially the extent to which a scenario complies with a user’s expectations” [472,473]) contributes to realistic behaviour in virtual environments and that improved visual realism can enhance realistic behavioural responses [477].

The studies in this chapter demonstrated how the use of VR studies increases the realism of

shoulder surfing research (Section 5.3) and enables and facilitates simulated in situ research on USEC prototypes (Section 5.4). However, it is important to keep in mind that hinting at similar user behaviour to the real world is, due to the introduced challenges when conducting in-depth security and privacy research in the wild [100, 306], often only possible using qualitative research methods (cf., Section 5.3.5.4 and Section 5.4.5.7). Therefore, comparisons to real-world observations are important to draw when using VR studies. Such comparisons can be implemented by outlining the differences to the results of baseline conditions and qualitative data collection methods that rely on the participants' prior real-world experiences.

5.5.1.2 Lesson 2: Consider the Use of VR Studies When the Aim Is to Contribute Towards Reasonably “Realistic” Shoulder Surfing and Authentication Experiences, but Keep Baseline Measures such as 2D Video Observations, *RW Lab*, and *VR Lab*

As evidenced by the participants' qualitative feedback and the sense of presence reported in the shoulder surfing study in Section 5.3, *VRO* led to more realistic shoulder surfing experiments than *2DVO*. However, this thesis and the existing literature provided empirical evidence that traditional *2DVO* already provide a suitable baseline measure when assessing a USEC prototype's resilience against observations (cf., [29] and Section 5.3.5.1). Whilst novel shoulder surfing methods, for example, *3DO* and *VRO*, may be used to contribute towards more realistic shoulder surfing experiences, they do not necessarily outperform traditional *2DVO* (and might even be misleading, cf., Section 5.3.6.1).

Similarly, for the simulated in situ authentication study in Section 5.4, it is essential to allow researchers to compare their results against well-established baselines (for example, to evaluations in traditional lab environments as simulated with *RW Lab* in the real world and *VR Lab* in VR). Without such comparisons, it is challenging to comment on the value and validity of in situ USEC studies based on staged real-world scenarios and VR simulations.

The long-term goal of establishing VR studies as a research methodology, as presented in this chapter, must be to allow researchers to “neglect” the comparisons to baseline conditions in the real world. However, when doing so, the results from “pure VR simulations” need to be treated carefully. The use of VR studies for simulating real-world research is still in the beginning and likely requires long-time community input from HCI, USEC, and VR researchers to show its full potential. This thesis chapter has put forward strong building bricks to promote and support future VR investigations on real-world USEC prototypes. It is important to set clear expectations and identify early on whether or not it is helpful to employ a VR-based research approach. In situations where investigations in the wild are infeasible, VR-based USEC research can be particularly promising. To make the results more tangible and support replications and comparisons to prior works, it is recommended to keep baseline

conditions such as 2D Video Observations, *RW Lab*, and *VR Lab* in the not-too-distant future.

5.5.1.3 Lesson 3: Apply VR Studies as a Research Method When Contexts are Challenging to Access in the Real World

VR-based shoulder surfing studies and simulated in situ studies, as presented in Section 5.3 and Section 5.4.2.3, are not an alternative to naturalistic research in the field, but they complement and advance the more traditional research methods. Studies of this type enable researchers to study scenarios that are otherwise challenging to access in the field and resource-intensive to replicate in the lab. The use of VR studies for human-centred usable security research does not require having physical access to security-sensitive contexts and provides researchers with more control of the study environment, eventually contributing to high internal validity. Virtual environments are more affordable and faster to build, deploy, and evaluate than equivalent real-world scenarios [291].

The use of VR studies for human-centred usability and security research is particularly promising when pandemics (for example, COVID-19) significantly impact the safety and well-being of the researchers and the user study participants. Whilst the studies presented in this chapter took place in laboratory environments, future work is encouraged to look at more distributed research approaches. Remote VR experiments introduce practical and ethical concerns [486], but they can “continue to forge forward with experimental work” [488].

As this thesis will show in chapter 6, applying VR studies in a remote research approach is particularly promising for usable security research when the aim is to simulate and evaluate more future-oriented, speculative real-world USEC prototypes and experiences.

5.5.2 Ways Forward

This chapter showcased two promising applications for the use of VR studies in the human-centred usability and security domain. It further contributed to a core contribution of this thesis by adding to the validation of the use of VR studies for USEC research through comparisons to baselines in both the shoulder surfing study in Section 5.3 and the simulated in situ study in Section 5.4. So far, this thesis has shown how lab-based VR studies facilitate and forge forward usability and security research. However, an essential next step to contribute to VR’s full potential for USEC research that typically takes place in the lab must be to demonstrate how the use of VR studies can contribute towards location-agnostic USEC research on VR replicas of real-world prototypes. In other words, how can the use of VR studies enable researchers to transfer real-world prototypes into virtuality and deploy virtual replicas of prototypes and study environments online to recruit participants from across the world?

As Chapter 6 will demonstrate, VR studies can indeed contribute towards a future where USEC evaluations on real-world prototypes are first transferred into VR and then deployed online. Chapter 6 will combine *traditional remote VR research* [339, 340, 400] with using *VR as a proxy for real-world research*, as proposed in chapter 3 and validated in chapter 4 and chapter 5. It will present a location-agnostic VR user study with the participants located in nine different countries, and evaluates the usability and social acceptability of two novel, speculative real-world USEC prototypes. As the last chapter of this thesis, it demonstrates how VR studies complement and advance the research methods available to the USEC community to study real-world prototypes and human factors.

5.5.3 Contributions

The contributions of this thesis chapter are three-fold. Both studies in this chapter contribute to the lessons outlined in Section 5.5 and provide *empirical contributions* as well as *methodological contributions* [556] to the USEC and HCI fields. The first study-specific set of contributions is based on the VR shoulder surfing study in Section 5.3:

- It provides an *empirical contribution* [556] through its proposal and human-centred investigation of the use of non-immersive and immersive VR observations for shoulder surfing research on USEC research artefacts.
- It showcases through three different authentication scenarios how the use of VR studies contributes towards more realistic shoulder surfing research than traditional lab studies, providing a *methodological contribution* [556] that may improve how the USEC and HCI communities design, build, and analyse research artefacts in the future.

The second set of contributions is based on the simulated in situ study this thesis discussed in Section 5.4. It makes the following *methodological* and *empirical contributions* [556]:

- It proposes and investigates, for the first time, the use of VR studies as a research method for simulated in situ USEC research on VR replicas of real-world prototypes (i.e., *methodological contribution* [556]). Furthermore, it evaluates its methodology through a replication and comparison study of ColorPIN [99], providing an *empirical contribution* [556] to the USEC and HCI research fields.
- It supplies the validation of applying VR for USEC research, shows how simulated in situ evaluations lead to a sense of realism, and provides strong fundamental work to enable researchers to study novel USEC prototypes in their intended usage contexts.

VI

REMOTE EVALUATION OF REAL-WORLD USEC
PROTOTYPES USING VR STUDIES

Chapter 6

Remote Evaluation of Real-World USEC Prototypes Using VR Studies

This chapter is based on the following two publications:

[Publication 8] **Mathis, F.**, O’Hagan, J., Vaniea, K., & Khamis, M. (2022). Stay Home! Conducting Remote Usability Evaluations of Novel Real-World Authentication Systems Using Virtual Reality. In Proceedings of the International Conference on Advanced Visual Interfaces (AVI 2022). ACM, DOI: [10.1145/3531073.3531087](https://doi.org/10.1145/3531073.3531087)

[Publication 9] **Mathis, F.**, Zhang, X., O’Hagan, J., Medeiros, D., Saeghe, P., McGill, M., Brewster, S., & Khamis, M. (2021) Remote XR Studies: The Golden Future of HCI Research?. In CHI 2021 Workshop on XR Remote Research, URL: http://fmathis.com/publications/chi2021_workshop_remoteXR.pdf

6.1 Introduction

So far, this thesis research has shown how VR studies contribute to more realistic USEC research compared to more traditional research in physical laboratory settings. For example, as shown in chapter 5, this was achieved by immersing the participants into various plausible VR replicas of real-world authentication scenarios. However, the studies in chapter 4 and chapter 5 relied on a physical lab setup to prepare the VR material (cf., Section 4.8 in chapter 4) and run the actual user studies (cf., Section 4.6 in chapter 4 and Section 5.3 & Section 5.4 in chapter 5). The latter setup, running the VR studies in physical labs, means that the participants were recruited within a local environment. In the context of this PhD research, the participant recruitment happened mostly in local areas within Austria and



Figure 6.1: This chapter proposes *Remote Virtual Reality for simulating Real-world Research* (RVR³) to evaluate novel real-world prototypes. Two real-world USEC prototypes for public displays (i.e., *Hand Menu* (③) and *Tap* (④)) were implemented and compared against *Traditional 4-digit PIN authentication* (①) and *Glass Unlock* (②) [553]. The individual prototypes are described in Section 6.3.

the United Kingdom due to the researcher’s residence. Whilst the results from such local studies are valuable and inspire follow-up research, the local participant recruitment is an ongoing challenge in the broader HCI and USEC fields. Linxen et al. [284] found that 73% of CHI study findings are based on Western participant samples, which represent less than 12% of the world’s population [284]. Whilst many research methods exist to reach out to broader populations and learn about people’s behaviour and opinion, such as interviews, online surveys, or field studies, these methods are often impractical to evaluate prototypes that involve hardware components. Recent comments by Schmidt et al. [445] and Alt [16] emphasised the need to move traditional lab research out of the lab and introduce research methods that do not rely on user studies in the lab. From the researchers’ perspectives, the use of VR “provides an opportunity to recreate your research environment virtually and let your participant access this from home. The researcher thereby is not limited to a lab environment but can rebuild arbitrary settings, including but not limited to public spaces, cars, homes or work environments” [445]. Figure 6.1 shows the VR research environment implemented for and used in this chapter, together with the different VR replicas of the real-world USEC prototypes, which are described in more detail in Section 6.3.

The previous chapters have provided the first evidence of using VR studies for USEC research and showcased two promising applications within the human-centred authentication field. To further showcase the potential of VR studies for USEC research, this chapter puts forward, for the first time, a **remotely** conducted VR user study, i.e., from a distance, to evaluate real-world prototypes, hereafter referred to as *Remote Virtual Reality for simulating Real-world Research* (RVR³). RVR³ is capable of targeting user study subjects from multiple countries, eventually contributing to a sample’s diversity. Virtual replicas distributed as part

of RVR³ do not require physical storage space, are easy to maintain, and require no access to hardware prototypes or contexts that are hard to reach in the real world. Finally, RVR³ is beneficial when direct interaction between the researchers and the participants is challenging or even prohibited (for example, due to COVID-19) [309, 488]. RVR³ has the potential to revolutionise research on USEC prototypes that is usually conducted in the physical lab. It provides researchers with a novel research method to move research on real-world prototypes out of the lab, similar to how online surveys have found widespread application in the broader USEC field and have made significant contributions to the field [406, 408].

This chapter presents research that applies remote VR studies to create virtual replicas of and evaluate real-world USEC prototypes. As such, it is the first work that applies VR studies as a research method for assessing real-world research artefacts without comparing the results to a real-world counterpart evaluation. Whilst this means the findings from such VR-based studies need to be treated carefully as results in the real world might differ, the future of VR studies is to apply VR as a research method to design, implement, and evaluate virtual implementations of potential real-world research artefacts without the need to run replication evaluations in the real world. Conducting USEC research on real-world prototypes holistically in VR is particularly interesting as researchers can apply speculative design research [127, 262] on devices that are not yet widely available (for example, ubiquitous immersive “always-on” technology [227]). As this chapter envisions a future of designing, implementing, and evaluating real-world USEC prototypes entirely in VR, it combines traditional remote VR research (for example, [339, 340, 431]) with using VR as a proxy for real-world research, and evaluates the usability and social acceptability of two novel real-world USEC prototypes: *Hand Menu* and *Tap*. The prototypes and the role of augmented reality (AR) for advanced real-world authentication in public spaces are described in more detail in Section 6.3.

In summary, the research in this chapter provides promising insights into the usability and social acceptability of novel real-world USEC prototypes for public displays. It showcases how traditional usability research on USEC prototypes can be moved out of physical laboratory environments, contributing answers to RQ₅:

RQ₅ Can the use of VR studies move traditional USEC research on real-world prototypes out of physical labs?

6.1.1 Chapter Structure

Section 6.2 describes the ethics, the compensation, and the data collection. Section 6.3 describes the USEC prototypes and the authentication context. The chapter then describes the USEC prototypes’ implementations in Section 6.4 and outlines the research methodology in Section 6.5. It then reports the study findings in Section 6.6 and discusses some study-specific

limitations in Section 6.8. The chapter concludes by discussing the findings in the light of prior work in Section 6.7 and by contributing answers to RQ₅ in Section 6.9.

6.2 Ethics, Compensation, and Data Collection

Ethical approval was sought and received from the University of Glasgow College of Science & Engineering ethics committee (*ref*: #300210038). Participants were recruited using social media (for example, Twitter and XRDRN¹) and word of mouth. The participants were paid according to their local standard (for example, £15 for participants from the UK). They used the Meta Quest 1 or Meta Quest 2 headset to participate – no other VR headsets were eligible to use for this study. Data was temporarily stored on the participants' VR headsets as *.csv* files. The data was locally stored to ensure no automatic file transfer is required for conducting such a remote VR study. The participants uploaded the stored *.csv* files at the end of the study session into an anonymised folder onto the University of Glasgow cloud storage. Asking the participants to upload the *.csv* files directly at the end of the study was done to avoid receiving potential “fake data” [488] and to allow the participants to reach out for help if needed for the file transfer. The participants had access to a *.PDF* that explained how to uninstall the study application from their device after the study. Uninstalling the study application deleted all stored user study data from their device and ensured no study traces were left behind. Additional data (for example, demographics, SEBIS [122], ATI [142]) were collected using Qualtrics [397]. Participant IDs (P1 to P25) were used to ensure anonymity. Appendix E contains all the study material and instructions used in this remote VR study.

6.3 Investigated USEC Prototypes and Context

Two novel authentication prototypes were studied to investigate VR's feasibility in conducting remote research on simulated real-world USEC prototypes: *Hand Menu* and *Tap*. Both prototypes use augmented reality to present users during their authentication with a unique and private PIN layout (cf., Figure 6.2). This makes authentications resilient against shoulder surfing. Both *Hand Menu* and *Tap* allow for touch-less user authentication, avoiding touching public surfaces, which can pose a considerable risk in the transmission of bacteria and viruses [413]. In line with previous work, the keypad layouts are randomised once at the start of each 4-digit PIN authentication due to security [553]. The randomisation is applied in all authentication prototypes, i.e., *Hand Menu*, *Tap*, and *Glass Unlock* [553], except in the traditional 4-digit PIN authentication (the first baseline in this chapter). The investigated USEC prototypes and the implemented baselines are now described in more detail:

¹<https://www.xrdrn.org/>, last accessed 22/01/2023

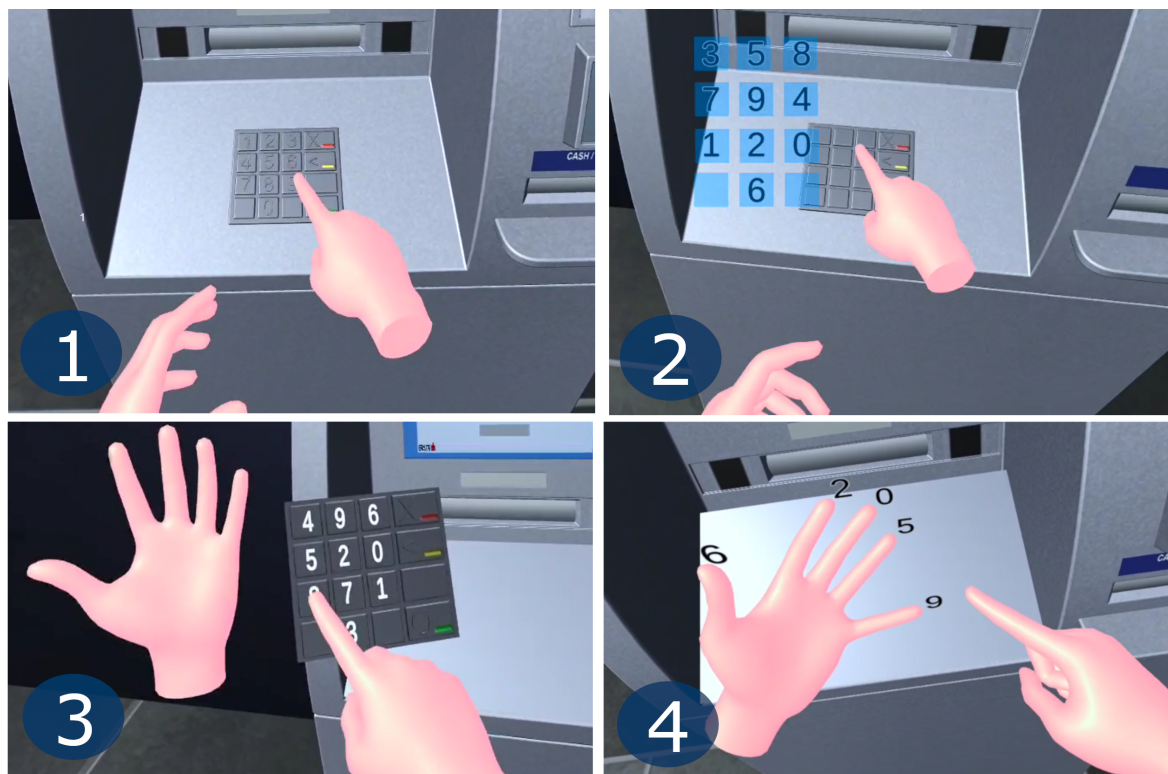


Figure 6.2: In ❶, the participants use a traditional keypad to authenticate. *Glass Unlock* [553] (❷) makes use of AR to present the participants with a private keypad layout. In *Hand Menu* (❸) and *Tap* (❹), the participants provide input on a hand-attached AR keypad (❸) or on augmented digits that are attached to their fingertips (cf., ❹).

- **Traditional Authentication + Glass Unlock:** Two USEC prototypes were implemented as baselines: 1) *Traditional* 4-digit PIN authentication and 2) *Glass Unlock* [553], an AR-based authentication prototype. Using traditional 4-digit PIN authentication as a baseline condition is a common approach in authentication research (for example, [27, 99]). *Glass Unlock* (10Key) [553] was added as a second baseline because both *Hand Menu* and *Tap* make use of the underlying concept of using AR for advanced authentication in public. Furthermore, the study aimed to investigate the participants' behaviour and *Glass Unlock's* usability when the prototype is simulated in VR and used for a different context than initially proposed for, i.e., smartphone unlocking [553]. In *Glass Unlock*, the user provides input on a traditional keypad, but this time with unlabelled buttons, i.e., the keypad has no digits. Instead, the randomised keypad layout is presented using the user's AR glasses (cf., Figure 6.2–❷).
- **Hand Menu Authentication (*Hand Menu*):** Instead of entering a PIN on a physical keypad, a one-time randomised keypad layout is augmented next to the user's wrist on which they provide mid-air input. Augmenting a keypad and applying one-time randomisation to the digits ensures the prototype's resilience against observations [553] and allows for touchless input. In summary, *Hand Menu* is a hand-attached user

interface that follows Microsoft's HoloLens 2 "*Hand menu*" implementation [329] and allows for fast and secure input (cf., Figure 6.2-③).

- **Tap Authentication (*Tap*):** In *Tap*, digits are augmented above the fingertips of the user's non-dominant hand in random order (cf., Figure 6.2-④). Only the user can see this mapping through their AR glasses. This allows for the underlying mapping of the digits, i.e., the number assigned to a finger, to be unknown by a bystander. To provide input, the user taps with their right index finger on their left-hand fingertips. Each finger of the user's left hand allows input of two digits depending on the current mode (A or B). Each mode covers five digits (for example, mode A: 6,2,0,5,9; mode B: 1,3,7,8,4). *Tap* makes use of pinch gestures² in the user's dominant hand to switch between the modes and enable them to correct and submit their PIN. Pinch gestures are commonly used for AR interaction [46, 385]. The user can switch between the two modes (A and B) by performing a pinch gesture between the thumb and index finger of the right hand. By using these two modes, the user can access all ten digits. For example, in Figure 6.2-④ the little finger allows input of the digit "9" in mode A. To delete the last digit entered, the user performs a pinch gesture with their right hand's thumb and middle finger. This gesture can be repeated multiple times to delete digits continually. Confirming the entered PIN was implemented using a pinch gesture between the thumb and ring finger of the user's right hand.

As real-world context, a VR replica of an ATM authentication scenario was implemented. Figure 6.3 shows the virtual room, the ATMs, and the virtual avatars that aimed to represent a realistic European ATM cash withdrawal scenario. Implementing an ATM cash withdrawal scenario allows the thesis to evaluate different real-world USEC prototypes in their intended usage scenario using a remote VR research approach. Furthermore, a VR replica of an ATM cash withdrawal scenario demonstrates how USEC research on public displays can be performed without having physical access to security-sensitive environments and systems.

6.4 Implementation and Apparatus

Virtual replicas of the fictional real-world USEC prototypes were implemented in VR using Unity 3D (C#). Oculus Integration and the Oculus hand tracking SDK [362] were used for the hand tracking. The Oculus' sample hand tracking implementation [362] was modified to provide the participants with a keypad with auditory and visual feedback when providing input. For *Glass Unlock* [553], the participants were presented with a virtual keypad layout in an egocentric view, i.e., the virtual keypad mapping was linked to the participants' head

²<https://docs.microsoft.com/en-us/hololens/hololens2-basic-usage>, last accessed 22/01/2023



Figure 6.3: An ATM authentication scenario was implemented to present the remote study participants with a realistic and plausible real-world environment. The figure shows an overview of the VR 3D environment and the participant's position (red box) when interacting with the USEC prototypes and the ATM during the study.

movements. This simulates the situation where a person wears AR glasses (for example, the Microsoft HoloLens 2). Oculus' OVRHand and OVRSkeleton [362] were used to augment a keypad next to the user's wrist when performing an open palm hand gesture in *Hand Menu* and map the digits to the user's fingertips in *Tap* (cf., Figure 6.2). The augmented keypad in *Hand Menu* was aligned to the right side of the user's wrist (non-dominant hand). In *Tap*, a 0.35 s delay between subsequent digit entries was added to avoid accidental inputs, determined

through pilot tests. In summary, *Hand Menu* and *Tap* simulate AR-based authentication systems that augment a) a virtual keypad next to users' wrist (for *Hand Menu*) or b) virtual digits on top of users' fingertips (for *Tap*).

For the authentication environment, a 3D ATM model [143] and 3D objects from Snaps Prototype [371] were used (cf., Figure 6.3). The simulated real-world scenario consists of five ATMs, one customer chatting with a bank employee, and one customer interacting with one of the ATMs (cf., Figure 6.1 and Figure 6.3). Adobe's Mixamo avatar library [11] was used for the bystanders in the virtual environment (cf., Figure 6.3). Environmental sound, i.e., people chatting, was added to enrich the experience and contribute to a more realistic ATM interaction scenario in public. The virtual ATM is fully functional and allows the participants to input a virtual credit card and navigate through a traditional European ATM user interface. After successful PIN input, the ATM outputs the credit card and the cash. In-VR menus and questionnaires (for example, for NASA-TLX [193] and the SUS [60]) were implemented to guide the participants through the study and not to break their VR experience [395]. Figure 6.4 shows the implemented VR menus. As study apparatus, the participants either used the Meta Quest 1 or Meta Quest 2 as a VR headset. Additionally, they used a Zoom-capable device and Oculus Casting to share their screen and provide the experimenter with a real-time view of their VR experience.



Figure 6.4: A menu in VR guided the participants through the remote VR study, ensuring to not break their VR experience [395] and to avoid switching between a) interacting in VR and b) filling in surveys outside of the VR experience multiple times during the study.

6.5 Methodology

The study application, an .apk file that can be installed on the Meta Quest 1 or Meta Quest 2, was distributed together with an installation guide describing how to install the study environment on the VR headset. A Zoom meeting was then scheduled for a 1.5 hour within-subjects user study. The Zoom meeting guided the participants through the remote VR study and allowed the researcher to conduct rich semi-structured interviews at the end of the study. The participants' demographics were collected before the user study session using Qualtrics [397]. The applied framework in this study follows one of the primary approaches for conducting remote VR studies: standalone VR application for development, direct download for distribution, and social media/ mailing lists for the participant recruitment [400, Fig. 5]. Overall, the study methodology can be defined as a remotely conducted immersive VR study to evaluate VR implementations of real-world USEC prototypes.

6.5.1 Independent and Dependent Variables

The two novel USEC prototypes, *Hand Menu* and *Tap*, were compared against the VR simulations of the *Traditional* and the *Glass Unlock* (10Key) [553] authentication prototypes. Therefore, the study had one independent variable (IV), the USEC prototype, with four levels: 1) *Traditional*, 2) *Glass Unlock*, 3) *Hand Menu*, and 4) *Tap*. Traditional usability metrics when evaluating novel USEC prototypes in the authentication field were measured: the participants' input time, the number of PIN corrections, and the number of incorrect PIN entries. Additionally, the participants were asked about their perceived workload using the NASA-TLX questionnaire [193] and their user experience using the UEQ [275]. The usability of the prototypes was measured using the SUS questionnaire [60]. The participants were asked additional 5-point Likert scale questions (for example, "*Input using this method is [usable in public].*") to allow the thesis to comment on the social acceptability of novel AR-based USEC prototypes when used for authentication in public.

Collecting insights about the social acceptability of novel AR-based authentication methods that introduce mid-air input, i.e., *Hand Menu*, and hand gestures, i.e., *Tap*, is important as prior work showed that location and audience have a significant impact on a user's willingness to perform gestures in public [416]. Assessing the social acceptability is also particularly important when eventually adopting novel authentication methods (for example, Social Compatibility: "*an authentication mechanism has to be designed to be compatible with such social factors and norms*" [94]). The study concluded with a usability, security, and combined usability and security ranking, and with a semi-structured interview (cf., Appendix E). The sample's security behaviour using the Security Behavior Intentions Scale (SeBIS) [122] and its technological affinity using the Affinity for Technology Interaction (ATI) scale [142] is

reported to support replication. Furthermore, the participants' sense of presence using the IPQ [451] and the TPI (dimension: social realism) questionnaire [287] is reported. After the participants experienced the VR environment for roughly one minute, their sense of presence was collected. Collecting the sense of presence along with the qualitative data from semi-structured interviews allows the thesis to comment on the participants' sensation of being in a real place and if the scenario was perceived as plausible and actually occurring [475].

6.5.2 Study Design and Task

Each participant experienced all USEC prototypes (within-subjects experiment). The order of the prototypes was counterbalanced using a Latin Square. Participants were first introduced to the different prototypes using a slide deck. They then experienced the virtual environment and filled in the IPQ questionnaire [451] and the TPI questionnaire [287]. The participants then entered three 4-digit PINs (for example, "1234") using the first authentication prototype as part of a training session. The ATM's user interface provided an authentication video in VR and all necessary details for the authentication sessions (for example, the PIN to enter). After the training, the participants went through a series of five PIN authentications. These authentications are referred to *isolated* authentications as the authentication itself is not part of a production task [305, 437]. This means the participants task was to enter 4-digit PINs using the corresponding authentication method.

After five successful authentications, they experienced the USEC prototype as a supporting task, i.e., in situ, as motivated by chapter 5, similar to how security tasks occur in the real world [437]. The decision of multiple authentication experiences was motivated by previous social acceptability research [416]: multiple exposures allow participants to develop preferences and contribute to more realistic results when asking about the prototypes' social acceptability [416]. Studying the social acceptability of the USEC prototypes is particularly important because if USEC systems are not usable and socially acceptable, people will use less secure alternatives. Furthermore, immersing the participants in a real-world setting allows the thesis to situate the results on social acceptability within real social experiences and better understand the participants' opinions and perceptions [416].

The participants had to a) take the (virtual) credit card, b) put the credit card into the ATM, c) authenticate using the corresponding authentication method, d) select the amount of money to withdraw, and e) take the card and the cash out of the ATM. They then reported their perceived workload [193], their user experience [275], rated the USEC prototype's usability [60], and filled in 5-point Likert scale questions [245]. The same procedure, including the training, was repeated for the other USEC prototypes. The study concluded with a usability and security ranking and with a semi-structured interview.

6.5.3 Data Analysis

Unless otherwise stated, one-way repeated-measures ANOVAs (for parametric data) and Friedman tests (for non-parametric data) were run. Post-hoc tests were Bonferroni corrected to correct for multiple comparisons. The semi-structured interviews were audio recorded and transcribed. The lead researcher went through all interviews to split the participants' statements into meaningful excerpts. A group of independent researchers ($N = 5$) then conducted an initial affinity diagram using Miro [333], an online collaborative whiteboard platform. The lead researcher first introduced the USEC prototypes and the interview questions. The team then grouped the participant statements into themes. All researchers were instructed to divide participant statements into two (or more) statements if required. The lead researcher finalised the affinity diagram based on an initial 2-hour session with the other researchers. This process resulted in an affinity diagram of 778 participant statements. The main findings of the semi-structured interviews are reported in Section 6.6.8. Reporting the number of the participants who shared certain opinions would be inaccurate due to the use of semi-structured interviews; thus, frequencies are only reported where appropriate.

6.5.4 Demographics

The results in this chapter are based on 25 participants (15 male, 9 female, 1 non-binary) who participated from overall nine different countries: 12 from the United Kingdom, four from France, three from the United States of America, and one each from Spain, Belgium, Finland, Czech Republic, Canada, and Singapore. The participants were on average 25.76 years (min = 17, max = 35, $SD = 4.36$). All participants were right-handed, except one with no marked preference for using the right or left hand. To participate, 19 participants used the Meta Quest 2, six the Meta Quest 1. The participants have VR experience of up to 5 years and 11 months ($M = 20.58$ months, $SD = 23.036$). All mentioned that they had used an ATM before. Their security behaviour score [122] was $M = 3.37$ ($Md = 4.0$, $SD = 1.47$) on a scale ranging from 1 to 5 (Device Securement ($M = 4.25$, $SD = 1.34$), Password Generation ($M = 3.19$, $SD = 1.48$), Proactive Awareness ($M = 2.66$, $SD = 1.36$), and Updating ($M = 3.63$, $SD = 1.1$)). Their technology affinity score [142], from 1 to 6, was $M = 4.20$ ($SD = 1.43$).

6.6 Results

This section outlines common authentication metrics, including input times, the number of corrections, and the number of incorrect PIN entries. First, *isolated* authentications are reported. Then, authentications that were part of an ATM interaction experience (*in situ*), as described in Section 6.5.2, are reported. The section then reports the participants'

perceived workload and experience when interacting with the prototypes. Additionally, the SUS scores [60], the participants' usability and security ranking, the responses to the 5-point Likert scale questions, and the main findings of the semi-structured interviews are reported.

6.6.1 Input Times

Input times from the first digit entry to the last input are reported. Only successful authentications w/o corrections were considered for the analysis to ensure internal consistency and a fairer comparison between the prototypes. There was a significant difference of input times between the USEC prototypes, $F_{(3,69)} = 67.33$, $p < 0.05$, $\eta_p^2 = 0.745$. *Traditional* (M = 3.70 s, SD = 1.31 s) and *Hand Menu* (M = 3.17 s, SD = 0.95 s) were significantly faster than *Glass Unlock* (M = 5.29 s, SD = 1.75 s) and *Tap* (M = 7.10 s, SD = 1.64 s) ($p < 0.05$). *Glass Unlock* was also significantly faster than *Tap* ($p < 0.05$). For *in situ*, there was a significant main effect of USEC prototype on input times, $F_{(3,27)} = 12.67$, $p < 0.05$, $\eta_p^2 = 0.585$. Input times differed significantly between *Traditional* (M = 3.85 s, SD = 1.91 s) and *Tap* (M = 6.65 s, SD = 1.89 s), between *Hand Menu* (M = 3.46 s, SD = 1.63 s) and *Glass Unlock* (M = 4.35 s, SD = 1.75 s), and between *Hand Menu* and *Tap* ($p < 0.05$). Table 6.1 provides an overview.

6.6.2 Number of Corrections

The number of corrections differed significantly between the prototypes, $\chi^2(3) = 13.45$, $p < 0.05$. *Tap* resulted in significantly more digit corrections (M = 0.70, SD = 0.76) than *Hand Menu* (M = 0.06, SD = 0.14). There was no significant difference between the other pairs (*Glass Unlock*: M = 0.30 (SD = 0.37), *Traditional*: M = 0.35 (SD = 0.44)). For *in situ*, there was no evidence that the number of corrections differed significantly, $\chi^2(3) = 3.16$, $p = 0.367$. The values were M = 0.24 (SD = 0.66) for *Traditional*, M = 0.40 (SD = 0.82) for *Glass Unlock*, M = 0.12 (SD = 0.44) for *Hand Menu*, and M = 0.38 (SD = 0.77) for *Tap*.

6.6.3 Number of Incorrect PIN Entries

There was no evidence that the number of incorrect PIN entries differed significantly between the USEC prototypes, $\chi^2(3) = 7.16$, $p = 0.067$ (*Traditional*: M = 0.11 (SD = 0.20), *Glass Unlock*: M = 0.07 (SD = 0.13), *Hand Menu*: M = 0.07 (SD = 0.13), *Tap*: M = 0.22 (SD = 0.28)). The same was found for *in situ* authentications, $\chi^2(3) = 3.86$, $p = 0.277$ (*Traditional*: M = 0.16 (SD = 0.37), *Glass Unlock*: M = 0.04 (SD = 0.20), *Hand Menu*: M = 0.08 (SD = 0.40), *Tap*: M = 0.04 (SD = 0.20)). Table 6.1 provides an overview of the values.

Table 6.1: Authentications in *Traditional* and *Hand Menu* were faster than *Glass Unlock* and *Tap*. Statistical analysis shows that *Glass Unlock* and *Tap* did not necessarily result in more PIN corrections and entry errors. The statistical analysis followed the description in Section 6.5.3. $p < 0.05$ highlighted. The $p < 0.05$ columns show pairwise comparisons.

	Isolated				Statistical Analysis	p<0.05
	(1) <i>Traditional</i>	(2) <i>Glass Unlock</i>	(3) <i>Hand Menu</i>	(4) <i>Tap</i>		
Input Times	3.70 (1.31)	5.29 (1.75)	3.17 (0.95)	7.10 (1.64)	$F_{(3,69)} = 67.33, p < 0.05, \eta_p^2 = 0.745$	1-2;1-4;2-4;3-4;2-3
PIN Corrections	0.35 (0.44)	0.30 (0.37)	0.06 (0.14)	0.70 (0.76)	$\chi^2(3) = 13.45, p < 0.05$	3-4
PIN Entry Errors	0.11 (0.20)	0.07 (0.13)	0.07 (0.13)	0.22 (0.28)	$\chi^2(3) = 7.16, p = 0.067$	n/a
	In Situ				Statistical Analysis	p<0.05
	(1) <i>Traditional</i>	(2) <i>Glass Unlock</i>	(3) <i>Hand Menu</i>	(4) <i>Tap</i>		
Input Times	3.85 (1.91)	4.35 (1.75)	3.46 (1.63)	6.65 (1.89)	$F_{(3,27)} = 12.67, p < 0.05, \eta_p^2 = 0.585$	1-4;3-4
PIN Corrections	0.24 (0.66)	0.40 (0.82)	0.12 (0.44)	0.38 (0.77)	$\chi^2(3) = 3.16, p = 0.367$	n/a
PIN Entry Errors	0.16 (0.37)	0.04 (0.20)	0.08 (0.40)	0.04 (0.20)	$\chi^2(3) = 3.86, p = 0.277$	n/a

6.6.4 Perceived Workload (NASA-TLX) and User Experience (UEQ)

The participants' perceived workload differed significantly between the USEC prototypes, $\chi^2(3) = 35.98, p < 0.05$. *Glass Unlock* (M = 32.90, SD = 16.75) and *Tap* (M = 48.07, SD = 22.36) resulted in a significantly higher perceived workload than *Traditional* (M = 20.47, SD = 16.12) and *Hand Menu* (M = 16.77, SD = 12.80) ($p < 0.05$). A more nuanced analysis, together with all means and standard deviations, is reported in Table 6.2.

Hand Menu received a positive user experience evaluation (> 0.8 [449]) in all dimensions of the UEQ questionnaire [275] (cf., Figure 6.5). *Tap* received a neutral evaluation ($-0.8 < \text{score} < 0.8$ [449]) except for stimulation and novelty (> 0.8). The UEQ dimensions for all USEC prototypes are visualised in Figure 6.5 and the statistical analysis is reported in Table 6.2.

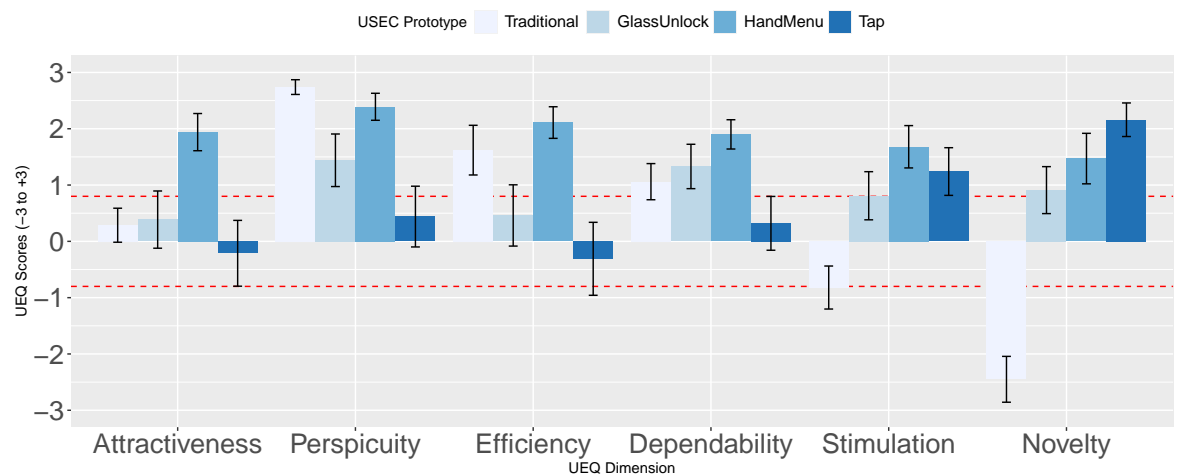


Figure 6.5: The visualisation shows all dimensions of the UEQ questionnaire. Black error bars denote 95% CI. Dotted red lines denote UEQ's +/-0.8 threshold [449].

Table 6.2: The table shows the NASA-TLX scores, the 5-point Likert scale scores, the UEQ scores, and the USEC prototypes' usability scores. $p < 0.05$ highlighted. The cells for the 5-point Likert scale questions are highlighted in green when the mean score is ≥ 3.75 , orange when the mean score is $2.5 \leq \bar{x} < 3.75$, and red when the mean score is below 2.5. Cell highlighting is inverted for error-proneness.

	(1) <i>Traditional</i>	(2) <i>Glass Unlock</i>	(3) <i>Hand Menu</i>	(4) <i>Tap</i>	Friedman Test	$p < 0.05$
NASA-TLX [193]						
Mental Demand	15.80 (20.34)	54.00 (28.10)	20.00 (25.21)	58.00 (31.72)	$\chi^2(3) = 33.61, p < 0.05$	1-2;1-4;2-3;3-4
Physical Demand	23.40 (25.11)	26.80 (24.74)	20.40 (20.61)	49.80 (30.63)	$\chi^2(3) = 18.32, p < 0.05$	1-4;3-4
Temporal Demand	25.20 (21.77)	21.40 (17.23)	19.80 (23.21)	33.20 (26.88)	$\chi^2(3) = 15.42, p < 0.05$	3-4
Performance	14.40 (16.91)	15.40 (13.53)	9.40 (9.50)	30.80 (29.46)	$\chi^2(3) = 10.14, p < 0.05$	3-4
Effort	19.80 (21.04)	46.80 (28.83)	19.00 (16.89)	61.80 (30.17)	$\chi^2(3) = 38.74, p < 0.05$	1-2;1-4;2-3;3-4
Frustration	24.20 (26.95)	33.00 (28.39)	12.00 (18.37)	54.80 (31.61)	$\chi^2(3) = 28.10, p < 0.05$	1-4;2-3;3-4
Overall	20.47 (16.12)	32.90 (16.75)	16.77 (12.80)	48.07 (22.36)	$\chi^2(3) = 35.98, p < 0.05$	1-2;1-4;2-3;3-4
5-point Likert Scale						
Ease	4.48 (0.92)	3.32 (1.03)	4.72 (0.54)	2.48 (1.19)	$\chi^2(3) = 47.62, p < 0.05$	1-2;1-4;2-3;3-4
Naturalness	4.36 (0.81)	2.68 (1.18)	3.88 (1.20)	2.36 (1.22)	$\chi^2(3) = 34.39, p < 0.05$	1-2;1-4;2-3;3-4
Pleasantness	3.20 (1.22)	3.16 (1.14)	4.36 (0.81)	2.88 (1.20)	$\chi^2(3) = 18.65, p < 0.05$	1-3;2-3;3-4
Speed	4.00 (1.15)	2.68 (1.44)	4.24 (0.72)	2.64 (1.22)	$\chi^2(3) = 22.91, p < 0.05$	1-2;1-4;2-3;3-4
Error-proneness	2.48 (1.33)	3.80 (1.26)	2.44 (0.82)	4.04 (1.24)	$\chi^2(3) = 34.18, p < 0.05$	1-2;1-4;2-3;3-4
Usable in Public	3.36 (1.41)	4.36 (0.81)	4.40 (0.76)	3.64 (1.29)	$\chi^2(3) = 10.32, p < 0.05$	(not confirmed)
Comfortable in Public	2.96 (1.43)	3.96 (1.02)	4.16 (0.85)	2.96 (1.65)	$\chi^2(3) = 16.06, p < 0.05$	1-3;3-4
UEQ [275]						
Attractiveness	0.29 (0.72)	0.39 (1.21)	1.94 (0.78)	-0.21 (1.39)	$\chi^2(3) = 34.21, p < 0.05$	1-3;2-3;3-4
Perspicuity	2.74 (0.31)	1.44 (1.11)	2.39 (0.57)	0.44 (1.28)	$\chi^2(3) = 47.71, p < 0.05$	1-2;1-4;2-3;3-4
Efficiency	1.62 (1.05)	0.46 (1.29)	2.11 (0.67)	-0.31 (1.54)	$\chi^2(3) = 40.556, p < 0.05$	1-2;1-4;2-3;3-4
Dependability	1.06 (0.76)	1.33 (0.93)	1.90 (0.62)	0.32 (1.13)	$\chi^2(3) = 31.31, p < 0.05$	1-3;2-4;3-4
Stimulation	-0.82 (0.91)	0.81 (1.01)	1.68 (0.89)	1.24 (1.00)	$\chi^2(3) = 50.32, p < 0.05$	1-2;1-3;1-4;2-3
Novelty	-2.45 (0.97)	0.91 (0.99)	1.47 (1.06)	2.16 (0.71)	$\chi^2(3) = 58.26, p < 0.05$	1-2;1-3;1-4;2-4
Hedonic Quality	-1.64 (1.24)	0.86 (0.99)	1.58 (0.98)	1.70 (0.98)	$\chi^2(3) = 53.65, p < 0.05$	1-2;1-3;1-4
Pragmatic Quality	1.81 (1.03)	1.08 (1.19)	2.13 (0.64)	0.15 (1.35)	$\chi^2(3) = 46.07, p < 0.05$	1-4;1-2;2-3;3-4
SUS [60]	84.5 (11.39)	70.2 (17.45)	90.5 (7.64)	50.3 (21.06)	n/a	n/a

6.6.5 System Usability Scale (SUS)

The SUS scores are reported as a standard metric for calculating the relative usability of the USEC prototypes [427]. No statistical analysis has been conducted on the SUS scores; instead, the SUS scores were analysed descriptively. *Hand Menu* yielded an “excellent” SUS score [34] of $M = 90.5$ ($SD = 7.64$), followed by *Traditional* with $M = 84.5$ ($SD = 11.39$). *Glass Unlock* and *Tap* yielded an average SUS score between “OK” and “GOOD” [34], with $M = 70.2$ ($SD = 17.45$) for *Glass Unlock* and $M = 50.3$ ($SD = 21.06$) for *Tap*.

6.6.6 Usability/Security Ranking and 5-Point Likert Scales

Weighted scores (i.e., rank 1×4 , rank 2×3 , etc.) were calculated to report the usability, security, and combined usability and security ratings of the USEC prototypes. *Hand Menu* achieved

the highest usability score (82), followed by *Traditional* (72), *Glass Unlock* (56), and *Tap* (40). *Tap* was perceived as most secure (80), followed by *Hand Menu* (79), *Glass Unlock* (65), and *Traditional* (26). For combined usability and security, *Hand Menu* achieved the highest score (92), followed by *Glass Unlock* (65), *Tap* (52), and *Traditional* (41). This means that the participants liked *Hand Menu* the most and *Traditional* the least. Whilst *Tap* was perceived as secure, its usability impacted the participants' preference, which is discussed in more detail in Section 6.7.1.

For the 5-point Likert scale questions, there was a significant difference between the prototypes' ease, $\chi^2(3) = 47.62$, $p < 0.05$, naturalness, $\chi^2(3) = 34.39$, $p < 0.05$, pleasantness, $\chi^2(3) = 18.65$, $p < 0.05$, speed, $\chi^2(3) = 22.91$, $p < 0.05$, error-proneness, $\chi^2(3) = 34.18$, $p < 0.05$, extent to which they were perceived as usable, $\chi^2(3) = 10.32$, $p < 0.05$, and comfortable to use in public, $\chi^2(3) = 16.06$, $p < 0.05$. Table 6.2 shows all pairwise comparisons.

6.6.7 Sense of Presence (IPQ and TPI)

The participants' sense of presence when experiencing the VR ATM scenario was measured. Their sense of presence was $M = 3.59$ ($SD = 1.80$) and their perceived social realism (TPI [287]) was $M = 4.89$ ($SD = 1.27$). The values for the IPQ's dimensions were $M = 4.2$ ($SD = 1.55$) for sense of being part, $M = 4.61$ ($SD = 1.33$) for spatial presence, $M = 2.99$ ($SD = 1.76$) for involvement, and $M = 2.77$ ($SD = 1.75$) for experienced realism. This suggests that whilst our participants' sense of being part and spatial presence was "acceptable" [326], the involvement and experienced realism scores were only "marginally acceptable", which can probably be explained by the lack of interactions in VR prior to the measurement.

6.6.8 Semi-structured Interviews

An affinity diagram, which was used to structure the qualitative data from the interviews as described in Section 6.5.3, resulted in six main themes:

Theme 1: Differences to Real-World Authentication. The participants mentioned that "most of the techniques are really similar to how [they] would imagine they are being implemented in the real world" (P19) and that their interaction experience and behaviour was "fairly similar to reality" (P14), "quite the same as I would experience in real" (P17), and "realistic in how it worked" (P8). However, P22 mentioned that "it is always easier with a real [ATM] machine" (P22) and that they missed the physical keypad in *Glass Unlock* and *Traditional*. The lack of haptic feedback when providing input on the (physical) keypad was brought up by many participants. P13 voiced that *Traditional* and *Glass Unlock* use a physical

keypad in the real world; therefore, they “*expect the [physical] buttons to be there*” (P13). Another topic that came up frequently was the hand and finger tracking’s accuracy and that this may be different in reality: “*most errors I did would disappear in the real world*” (P22); “*maybe I did more mistakes because of VR*” (P9); “*usually in real life I use multiple fingers and [authenticating] is a bit quicker*” (P2).

Theme 2: Prior Real-World ATM Experience. There was a consensus that the VR environment provided the participants with a good simulation of a plausible ATM interaction scenario and that the “*ATM was very convincing [and] felt like authenticating on the machine*” (P22). P7 mentioned that “*in virtual reality, I had the same mindset as if I were to be in front of a real ATM*” (P7). P3 referred back to the training session and stated that “*if I had to do a training session in the real world, I would expect what I did [...] everything was matching my expectations*” (P3). However, some participants mentioned that they were less aware of the (virtual) context around them: “*I have this tunnel effect in VR that I do not have in the real world*” (P22) and that they could tell “*it is not real*” (P12). P5 voiced the graphic fidelity gave the impression it was only a simulation. P18 mentioned that although the scenario, sound, and setting were good, the visual representation impacted their perceived realism.

Furthermore, there were differences between the participants’ previous experience of ATM interactions and the study environment. P23 voiced that their ATM experience differs significantly from the user study environment: “*I don’t usually see people in the bank [...] I don’t even see the receptionist for the bank*” (P23). P14 mentioned that they usually go to a drive-by ATM where they drive up with their car, open the car window, and reach out of the window to withdraw cash with their smartphone. Similarly, P25 mentioned that “*this is very much down to my personal setup [but] I can use my phone to withdraw money from an ATM*” (P25). P21 mentioned that ATMs are sometimes located in open spaces, making them more nervous when withdrawing cash. P2 voiced that they “*usually get [money] from like a hole in the wall [...] usually they have them on the street*” (P2).

Theme 3: The USEC Prototypes’ Usability. There were usability comments on all four VR USEC prototypes:

- **Traditional:** The participants found *Traditional* as usable and mostly referred to their familiarity with the prototype. P15 voiced that *Traditional* is “*sort of intuitive [...] partly because that’s something we’ve used for ages*” (P15). P19 mentioned that “*people are already familiar with such a system*” (P19) and that it is a “*common model*” for them because of prior exposure. All in all, the participants’ comments suggest that they were already familiar with *Traditional*, reinforcing the decision to treat *Traditional* as a baseline to simulate 4-digit PIN authentication on an ATM keypad.

- **Glass Unlock:** In line with the original *Glass Unlock* study [553], the participants reported that *Glass Unlock* requires an attention switch between the augmented keypad and the physical keypad: “*I’ve tried to switch between the two [layouts], but I am terrible at memorising numbers [...] memorising the keypad layout was not possible for me*” (P14). The fact that input in *Glass Unlock* still used a “traditional” keypad influenced some participants as they had, due to habituation, already established a mental model of the original keypad layout: “*sometimes, because of habits [...] I have the mapping of 1,2,3,4,5,6 in my mind, which causes a bit of confusion*” (P24). Overall, the participants perceived *Glass Unlock* as socially acceptable “*because it’s very similar to traditional [authentication]*” (P19) and because “*it is quite discreet and not embarrassing [when authenticating]*” (P8).
- **Hand Menu:** The participants perceived *Hand Menu* as fast and easy to use. P21 voiced that it takes the advantages of AR but is “*still not too complicated, easy to understand, and easy for people to adapt from the real world to AR, which is important from a product perspective*” (P21). P24 mentioned that *Hand Menu* “*is really close to what we are used to in real life, but provides a little bit more flexibility*” (P24). Some participants found that *Hand Menu* is slightly less acceptable in public because of “*poking the air, which is a little bit weird right now [...] less conventional input, maybe in like 30 years that’s not the case anymore then, but we’re not there yet.*” (P6). P14 stated that “*if someone was to walk by me as I’m at the ATM they think I’m a freak because I’m just tapping in the air and I’m wearing these crazy glasses*” (P14).
- **Tap:** *Tap* received negative usability comments due to the hand tracking accuracy and its complexity to recall gestures: “*people have to learn a lot [...] I wasn’t able to remember what combination it was to delete or confirm*” (P11). Some participants voiced that *Tap* is comparable slow and does not feel “natural” (P8). P5 mentioned that they sometimes have to hold something in their hand when using an ATM, which makes *Tap* inconvenient to use. The participants questioned *Tap*’s social acceptability. P20 mentioned that “*each hand gesture has a different meaning in a different culture [...] a **pinch gesture between middle finger and thumb** in the Buddhism culture has [the meaning of] trying to mediate*” (P20). P25 brought up that *Tap* “*is a little bit awkward and I might feel a bit stupid doing that in public*” (P25). P6 voiced that “showing the middle finger” was the easiest way for them to provide input but that they would be less inclined to use *Tap* in public due to such inappropriate gestures.

Theme 4: Perceived Security of AR-based Authentication Prototypes. The participants perceived the AR authentication prototypes, i.e., *Glass Unlock*, *Hand Menu*, and *Tap*, as more secure than *Traditional*. Whilst shoulder surfing [125] was frequently mentioned

as a potential threat when using *Traditional*, the participants voiced for the other prototypes that they “*probably need to get some view of the AR experience*” (P6) or perform a man-in-the-middle attack because “[*the systems*] *require some network communication*” (P20). P18 mentioned that they are usually aware of bystanders when authenticating in public, but that the use of AR could influence their awareness of the surrounding: “*with AR, I think that breaks a little bit of the reality, so you are more prone to the security problem.*” (P18). Overall, the participants mentioned either getting access to the user’s AR view by a) trying to catch a glimpse of what is rendered on the glasses (P4) or b) hacking the system (P7), or conducting a man-in-the-middle attack (P20) to capture information transferred from the AR glasses to the situated display (in this case an ATM).

Theme 5: VR-based Real-World Studies: Pros and Cons. There was overall positive feedback on the remote VR user study when simulating the real-world ATM environments and the USEC prototypes. P4 mentioned that using a VR-based introduction to novel real-world prototypes is particularly promising for someone who “*is nervous about doing this on the street or in a (real) bank*” (P4) and that it could be particularly helpful to “*teach kids who have their first experience using these interfaces, like getting their first bank card*” (P4). P20 mentioned that “*AR is still hard to use in real life and [the VR setting] gives a very good setup for reconstructing [an ATM] situation*”. Others voiced “*it is easier to get more people to try it, because you can have multiple people using it at the same time*” (P25), that implementing and evaluating all different USEC prototypes in reality would be expensive (P21), and that a remote VR study allowed them to participate in this research, despite being in another country (P16). P23 mentioned that experiencing the USEC prototypes in VR changed their initial preference: “*I didn’t expect Glass Unlock to be good [...] I thought Hand Menu would be my favourite, but it turns out Glass Unlock was actually my favourite; so I’m happy that I get to experience all three in virtual reality, before I can apply it to real life.*” (P23).

The participants raised some concerns about the lack of interaction fidelity and that VR implementations might not accurately represent how input on a physical keypad in *Traditional* and *Glass Unlock* would work in reality: “*this is still simulated, maybe in a real use case you would have different opinions*” (P19). P25 further voiced that the VR implementations of the four real-world USEC prototypes come with “*lower fidelity than if you actually build four machines [in reality]*” (P25).

Theme 6: The Participants’ Real-World Study Environment. The participants participated in the VR study from various locations³: from their living room (8), their home

³Note that the study setup, i.e., the VR experience, depicted in Figure 6.3, was the same across all the participants. The idea of participating from several physical locations and testing USEC prototypes in VR is not to evaluate simulated real-world USEC prototypes in, for example, bedroom settings, but to enable the

office (7), their bedroom (5), a research lab (4), and their private gym (1). The participants voiced there was nothing that significantly impacted their VR experience during the study. Although some participants mentioned minor issues with the Oculus Guardian [363] when configuring their Quest before the actual user study or during the training session, they “*did not pay attention to the [real-world] surrounding at all*” (P15). P22 brought up the problem of bumping into real-world obstacles, which required some preparation before the data collection: “*I just have to be careful not bumping into my desk when approaching the ATM*” (P22).

6.7 Discussion

The remote VR study showcased RVR³'s potential to move USEC research on VR replicas of real-world prototypes out of a physical laboratory. The study results provide a glimpse into the usability and social acceptability of two novel USEC prototypes: *Hand Menu* and *Tap*. Authentications using *Tap* took significantly longer and were more demanding than *Traditional*. However, there was no notable difference between *Traditional* and *Hand Menu* regarding perceived workload, input speed, number of digit corrections, PIN entry error rate, and pragmatic quality (cf., Table 6.1 and Table 6.2). *Hand Menu* resulted in an “excellent” SUS score, achieved the highest usability score, and received an overall positive UEQ evaluation. However, the perceived social acceptability was less prevalent for *Hand Menu* and *Tap* due to the use of AR glasses and mid-air input, which some participants perceived as inappropriate in public. As put by P14: “*if AR glasses can get a better form factor, so Nreal and the Snap Spectacles [glasses] are decent, [...] once people start using them [...] and when apple comes out with [their AR glasses], that's when we get more of that acceptance.*” (P14). There were mixed comments about the perceived realism of the authentication context, which is discussed further in Section 6.9.1.2.

Whilst the results of the user study imply that people are reluctant in adopting *Hand Menu* and *Tap* due to social acceptability concerns, the results can be decisive and trend-setting for the future of USEC prototypes. RVR³ provided the first insights into the use of novel, not yet widely available technology, for advanced user authentication in public, which will be further discussed in Section 6.7.1 and this chapter's conclusion in Section 6.9.

6.7.1 RVR³: A Complementary Research Method

Schmidt et al. [445] and Alt [16] highlighted the HCI and USEC communities' recent interest in moving human-centred research out of physical labs. Evaluating VR replicas of real-

participants to participate from their preferred location, potentially contributing to larger and more diverse samples. No real-world environments were recorded by the VR headset.

world USEC prototypes using a remote VR research approach, as performed in this chapter, notably advances human-centred research. So far, evaluating hardware prototypes outside of a research lab was often infeasible due to deployability issues, as found in the expert interviews in chapter 3. Whilst HCI prototypes can be evaluated in physical lab environments, corresponding studies often lack realism [118] and exhibit small and homogeneous samples [66, 258, 284]. Using *Remote Virtual Reality for simulating Real-world Research* (RVR³) as a proxy for real-world research opens a new world of opportunities for the broader HCI and USEC research communities. Researchers can recruit user study subjects from different countries, scale up their sample sizes, and adjust their research artefacts, including the study environments and the prototypes, without purchasing or building special hardware.

However, despite the prior HCI works that validated VR's use for empirical real-world research (for example, [291, 528]) and the first validation of the use of VR studies for USEC research presented in chapter 4 and chapter 5, it is important to acknowledge potential technical limitations of VR and have a clear vision of what can be expected from evaluations that are conducted on virtual artefacts in VR instead of on real-world prototypes in real life. For example, investigating *Tap*'s usability in the lab using better hand tracking technology (for example, an OptiTrack system [370]) may impact the participants' usability perception and their performance. At this point, as put by one of the experts during the expert interviews in chapter 3, "*we [as a community] just need to be a little bit more open to what sort of solutions/evaluations we are expecting out of [something] that has not actually been deployed in the real world*" (P11).

In summary, RVR³ forms a promising research method to implement and evaluate VR replicas of real-world USEC prototypes and move user-centred research out of physical laboratories. However, care needs to be taken when interpreting the results from VR user studies that simulate real-world research as, similar to lab studies and organised field studies, they still do not necessarily represent naturalistic user behaviour in real-life settings as shown in chapter 5 and indicated by the participants during the semi-structured interviews in this chapter.

6.8 Limitations

There are some study-specific decisions that are worth discussing. First, this chapter did not put forward an empirical assessment of the USEC prototypes' security as Winkler et al. [553] argued that private near-eye displays, i.e., AR glasses, allow for secure interaction by design. However, future work may want to conduct exhaustive security evaluations and empirical security assessments of the USEC prototypes introduced in this chapter before widely deploying them in the wild. Furthermore, to understand the USEC prototypes' usability, a large number of common usability measures (for example, NASA-TLX [193], SUS [60],

and UEQ [275]) and in-depth qualitative data from interviews were consulted. However, some studies [100, 189] also consider the participants' preparation times, i.e., the mental time it takes until a user performs input. The primary contribution of this chapter was not to introduce novel USEC prototypes but to showcase how novel USEC prototypes can be evaluated using RVR³. Therefore, the study was not set up to precisely measure the preparation times.

6.9 Chapter Conclusion

This chapter closes this PhD research by answering RQ₅ and introducing *Remote Virtual Reality for simulating Real-world Research* (RVR³), a novel research approach to evaluate VR replicas of real-world USEC prototypes. A remote VR user study with 25 participants was conducted to assess the usability and social acceptability of two real-world USEC prototypes: *Hand Menu* and *Tap*. The remote VR study provided a glimpse into the usability and social acceptability of AR-based authentication systems, fictional prototypes based on prior work that may (or may not) find application in the near-distant future. Authentications were moderately fast in both *Hand Menu* (up to M = 3.17 s, SD = 0.95 s) and *Tap* (up to M = 6.65 s, SD = 1.89 s); however, the participants criticised the prototypes' social acceptability and mentioned that people might feel reluctant to use AR-based systems these days. The chapter highlighted the VR's affordances to move traditional lab-based research on VR replicas of real-world prototypes to the participants' homes. It demonstrated RVR³'s potential to complement and, sometimes, replace research in traditional physical laboratory environments, opening the door for the HCI and USEC communities to evaluate VR implementations of novel real-world prototypes out of the lab.

The next section answers RQ₅ and outlines two promising research applications for *Remote Virtual Reality for simulating Real-world Research* (RVR³) in the USEC domain.

6.9.1 Research Question 5 (RQ₅) and Research Applications

This chapter contributes answers to the following research question:

RQ₅: Can the use of VR studies move traditional USEC research on real-world prototypes out of physical labs?

Based on the remote VR study presented in this chapter, yes, VR studies can move traditional USEC research on real-world prototypes out of physical labs. This chapter provided empirical evidence that a research infrastructure exists, including the deployment of VR applications, the participant recruitment, and having access to commercially available VR headsets, that allows

USEC researchers to deploy their virtual replicas of study environments and USEC prototypes online. Whilst the research in chapter 5 has shown how the participants' experiences can be moved out of physical lab environments, this chapter showcased how entire user studies on VR implementations of real-world prototypes can be moved to the participants' homes using VR studies, revolutionising and facilitating research on USEC prototypes that is usually conducted in physical laboratories. To further contribute to RQ₅ and highlight the long-term potential of VR studies, this chapter concludes by outlining two promising research applications for RVR³ that can forge forward USEC research involving hardware prototypes.

6.9.1.1 Research Application 1: Longitudinal Studies

This chapter presented a remote VR study to evaluate the usability and social acceptability of novel real-world USEC prototypes. The study showed that, yes, VR studies can be applied in the USEC research field to move research on VR implementations of real-world prototypes out of the lab. However, the study revealed that the participants' familiarity with prior systems could impact the study results. This was apparent in the study as follows: P24 mentioned that their prior experience impacted them when using *Glass Unlock*: “because of habits [...] I have the mapping of 1,2,3,4,5,6 in my mind”. At this point, RVR³ can be particularly valuable by tasking participants to authenticate using *Glass Unlock* once every day (for several months). This would allow the researchers to obtain learning effects and receive more realistic usability assessments of USEC prototypes than existing research methods are currently capable of. For example, conducting longitudinal studies where the participants must commute to a physical lab to interact with the research artefacts is often not feasible over several months. Longitudinal studies are rare in the broader HCI field, with over 85% of studies lasting a day or less [258]. RVR³ addresses the shortcoming of the current HCI and USEC landscape by opening the door for the communities to conduct longitudinal user studies without much resources and effort. For example, the use of VR studies, once well-evaluated across the broader HCI and USEC research fields, can lead to unsupervised user research where the general public can participate in user studies anytime and anywhere (for example, as already put into practice by Mottelson et al. [340]).

6.9.1.2 Research Application 2: Cross-Country Studies

RVR³ has enabled recruiting participants from nine different countries. However, a formal cross-country comparison of the USEC prototypes' usability and social acceptability was out of the scope of this PhD research. Based on the participants' qualitative feedback in Section 6.6.8, conducting large-scale geographically agnostic comparisons is one promising application for remote VR studies that can further contribute to the transition of research

findings into practice. Such a cross-country evaluation can further highlight VR's strengths for remote investigations on VR replicas of real-world prototypes and allows identifying the impact of social and technological factors on the results of prototype evaluations. The combination of user evaluations and research on VR replicas on real-world USEC prototypes is particularly interesting because technology is often designed and evaluated using western, educated, industrialised, rich, and democratic “*WEIRD*” samples [284]. RVR³ provides an excellent opportunity for researchers to broaden the international representation of the participant samples and work towards impactful HCI systems that are universally useful and engaging. However, in the same breath, researchers must be careful when implementing VR replicas of real-world prototypes and environments. Their mental model (for example, how an environment should look like) may not necessarily align well with the expectations of an international participant sample (cf., Section 6.6.8). Whilst RVR³ allows researchers to conduct cross-country evaluations, research is required to identify the challenges associated with large-scale cross-country studies on implemented VR simulations of real-world prototypes.

6.9.2 Contributions

This chapter makes the following empirical *artefact contributions* [556] as well as *methodological contributions* [556] to the USEC and HCI research fields:

- For the first time, it provides a *methodological contribution* [556] through its investigation of the usability and social acceptability of implemented VR replicas of novel real-world USEC prototypes in a remote VR user study (RVR³). Whilst prior work moved traditional VR research online [340, 400] and investigated the VR's feasibility for real-world USEC research in the lab (chapter 4 and chapter 5), this chapter extends research in this space by conducting a VR-based usability study of implemented VR replicas of real-world USEC prototypes in a fully remote VR setup.
- Furthermore, it provides an *artefact contribution* as well as an *empirical contribution* [556] through its proposal of two novel USEC prototypes (*Hand Menu* and *Tap*) for authentication in shared and social spaces and its usability and social acceptability results from a remote VR study with 25 participants located in nine different countries.
- Finally, it discusses the results based on the “out-of-the-lab research” comments by Schmidt et al. [445] and Alt et al. [16] and concludes with two promising research applications when studying real-world USEC prototypes using RVR³.

VII

SUMMARY AND REFLECTION ON THESIS RESEARCH

Chapter 7

Summary and Reflection on Thesis Research

Research must continue to be the centerpiece of intellectual life, and our commitment to research must grow, because our problems are growing.

– Ernest L. Boyer –

7.1 Introduction

This thesis made the following statement in the Introduction in chapter 1:

This thesis made the following statement in the Introduction in chapter 1: *“This thesis explores the suitability of virtual reality (VR) studies in supporting human-centred usability and security research. VR studies enable researchers to augment human-centred research methodologies that are constrained to conditions that can be physically replicated in the lab. This thesis presents a novel complementary research method for human-centred usability and security research by exploiting VR’s characteristics to expand the possibilities of evaluating USEC prototypes. It first identifies existing research challenges in usable security (USEC) (**identify**, chapter 3). It then validates the use of VR studies for human-centred usability and security research (**validate**, chapter 4). This thesis concludes with investigations on how VR studies can augment and move USEC research out of the lab (**advance**, chapter 5 and chapter 6), unfolding the advantages and disadvantages of VR studies for human-centred usability and security evaluations of real-world prototypes.”*

In the subsequent empirical chapters, chapter 3 to 6, research was presented that supports the thesis statement and contributes answers to five research questions, RQ₁ to RQ₅:

RQ₁ What are the challenges that USEC experts experience when designing, implementing, and evaluating security and privacy-enhancing prototypes?

RQ₂ Which findings of VR-based usability evaluations on USEC prototypes match the findings from corresponding evaluations in traditional physical lab settings?

RQ₃ Which findings of VR-based security evaluations on USEC prototypes match the findings from corresponding evaluations in traditional physical lab settings?

RQ₄ Can substitutional in situ studies using VR provide a bridge over the methodological gap between lab and field studies?

RQ₅ Can the use of VR studies move traditional USEC research on real-world prototypes out of physical labs?

Answers to the questions were provided in concluding sections within each chapter. Table 7.1 provides an overview of the chapters and the places where this thesis has answered the RQs.

Table 7.1: Overview of the research questions that were answered in the individual chapters, chapter 3, chapter 4, chapter 5, and chapter 6. Chapter 7 addresses the overarching research question, RQ₆, based on the sum of all studies.

Research Question (RQ)	Thesis Chapter	Thesis Section	
RQ ₁	chapter 3	Section 3.4 and Section 3.5.1	[Publication 3]
RQ ₂	chapter 4	Section 4.10.1	[Publication 4]
RQ ₃	chapter 4	Section 4.10.2	[Publication 4 and 5]
RQ ₄	chapter 5	Section 5.3.8.1, 5.4.8.1, and 5.5.1	[Publication 6 and 7]
RQ ₅	chapter 6	Section 6.9.1	[Publication 8]
RQ ₆	chapter 7	Section 7.2.4	

In the following, the research in the individual chapters is synthesised into research recommendations to provide answers to an overarching research question, RQ₆. So far, the thesis has reviewed the literature in chapter 2, provided the first validation research of the use of VR studies for USEC research (cf., chapter 4 and chapter 5), and showcased how VR studies revolutionise and forge forward USEC research that involves prototypes (cf., chapter 5 and chapter 6). The central contributions of this thesis are now summarised in Section 7.2. Section 7.2.4 then presents five lessons learned and research recommendations when using VR studies for human-centred usability and security research on VR implementations of real-world prototypes.

7.2 Central Contributions

This thesis makes novel contributions which inform the design and the evaluation of USEC research artefacts. Its main contributions are mapped onto a three-stage research arc (cf., Figure 1.1): (1) Identify: Expert interviews on the existing USEC research challenges and ways forward; (2) Validate: Investigation of the use of VR studies for simulating real-world USEC research involving prototypes; and (3) Advance: Evaluation of novel research methods and artefacts to push forward USEC research and contribute towards facilitating USEC research that is concerned with prototypes. This section summarises these contributions.

7.2.1 Identify: Insights from USEC Experts

In chapter 2, this thesis found that many challenges make USEC research laborious and sometimes even infeasible. Individual works have touched on what makes USEC research particularly laborious. For example, both De Luca et al. [100] and Volkamer et al.'s [529] work highlight the ethical and legal constraints associated with observational research in security-sensitive contexts. However, what was missing was a structured attempt to elicit challenges experienced by USEC experts that go beyond their experience in individual research projects. The identified key challenges in the broader USEC field raise awareness of the current shortcomings and provoke change in how the USEC community tackles research problems, some of which this thesis put forward as the first investigation of using VR studies for empirical USEC research. Whilst some of the challenges in chapter 3 are known by the community, they are not necessarily reported in publications and are often only accessible to researchers who are privileged enough to be part of the “hallway chatter” at international HCI and USEC conferences. Therefore, the research in chapter 3 is the first work that provides this kind of “hallway knowledge” view of the challenges faced by USEC researchers who design, implement, and evaluate research prototypes to the broader research community.

7.2.2 Validate: VR Studies for USEC Research

The exploration and validation of the use of VR studies for USEC research, to which five empirical studies contributed, reported in Section 4.6, Section 4.7, and Section 4.8 in chapter 4, and Section 5.3 and Section 5.4 in chapter 5, is a core element of this thesis and makes significant contributions to the field. This thesis contributes an experimental comparison of traditional real-world research in the lab with the use of VR studies for simulating real-world environments and research artefacts. Furthermore, it contributes two VR-based investigations on simulated shoulder surfing research and simulated in situ authentication research. Both

studies supply the validation of the use of VR studies for USEC research and contribute novel, well-evaluated research approaches for evaluating USEC prototypes.

7.2.3 Advance: VR Studies' Potential for USEC Research

Through the two empirical studies in Section 5.3 and Section 5.4 in chapter 5, this thesis contributes to advancing USEC research. It introduces novel research methods for 1) shoulder surfing research in public, a common threat model when conducting security research on prototypes, and 2) in situ research on real-world prototypes, showcasing how the use of VR studies facilitates and enables more realistic USEC research. Both studies in chapter 5 are the first empirical works that apply VR studies for USEC research and combine the internal validity of controlled laboratory experiments with the more realistic settings of field experiments. Finally, this thesis introduces and executes the idea of remote VR studies for the evaluation of future-oriented real-world USEC prototypes, referred to as *Remote Virtual Reality for simulating Real-world Research* (RVR³). For the first time, a remote VR study that simulates a real-world environment and novel AR-based USEC prototypes demonstrates how VR replicas of real-world research artefacts can be evaluated in a remote setting, paving the way for more large-scale, longitudinal, and impactful USEC research involving prototypes.

7.2.4 Synthesised Thesis Research: Research Recommendations

This thesis contributes five lessons learned and research recommendations based on the empirical research presented in chapter 3 to chapter 6. The lessons learned and the research recommendations provide answers to the overarching research question, RQ₆, and aim to support, facilitate, and advance USEC research.

RQ₆ What are the advantages and disadvantages of using VR studies for USEC research involving prototypes compared to traditional USEC research in physical laboratories and in the field?

7.2.4.1 Lesson 1: VR Studies for Research in the Lab and in the Field

The USEC expert interviews in chapter 3 revealed that contributing towards high ecological validity is a fundamental and ongoing challenge in USEC. Researchers face significant challenges when deploying research artefacts in the wild and providing user study participants with authentic use cases when evaluating novel prototypes. Whilst traditional lab studies are commonly used in USEC, they are often incapable of simulating real-life scenarios.

Even if in-the-wild research is feasible, the interviews in chapter 3 showed that some USEC researchers consider their (academic) career as more important than going the “extra mile” of field studies compared to more “straightforward” research in laboratory settings. For example, one of the experts criticised the mindset of some researchers who share the opinion of “why should I have to go out and do something differently [...] it’s a lot harder, it’s a lot more work and as long as I can get this stuff published why should I bother?” (P8). The use of VR studies for USEC research cannot change the “publish or perish” mindset [315] of some researchers and, as discussed in Section 5.4.6.2 in chapter 5, cannot fully elicit user behaviour typical in the field. However, it provides a novel research method to conduct simulated in situ USEC research that may not be feasible in the real world and hints at people’s preferences and opinions when experiencing USEC prototypes in plausible scenarios. As shown in Section 5.4 and visualised in Figure 5.11, the use of VR studies can be mapped onto a “ecological validity continuum” and, depending on the research interests, VR studies can be designed as being closer to lab studies or field studies. This allows researchers to simulate different scenarios and evaluate research artefacts under various circumstances, as showcased in chapter 5 and chapter 6.

The flexibility of VR studies provides, for the first time, researchers with a holistic research method that enables transitioning along an ecological validity continuum and incorporating the pros and cons of both (virtual) labs and more naturalistic (virtual) public spaces, as demonstrated in chapter 5. For example, VR studies allow researchers to conduct detailed human factors research in ATM authentication contexts, which is often impossible using other means. However, it is important to acknowledge that VR studies cannot fully simulate research in the wild due to, for example, the nature of user studies, i.e., demand characteristics [105,372], and the novelty of research artefacts, as shown in Section 5.4.6.2. Yet, the empirical research in chapter 5 and chapter 6 has shown that applying VR studies for USEC research is promising and allows researchers to assess research artefacts across different environments and in otherwise hard-to-reach real-world contexts.

Research Recommendation 1

▷ chapter 5, chapter 6

This thesis recommends applying VR studies for research in (virtual) lab environments (for example, *VR Lab* in Section 5.4) and more naturalistic research in simulated field environments that is particularly challenging to conduct in the real world (for example, *VRO* in Section 5.3 and *VR ATM Public* in Section 5.4). It is recommended to utilise VR studies as a powerful research method for usability and security evaluations at both ends of the ecological validity continuum (cf., Figure 5.11).

7.2.4.2 Lesson 2: When and How To Use VR Studies for Simulated Real-World Research

Selecting the most suitable research method for a given problem is challenging and often influenced by many factors, including the researcher's research interests and financial, ethical, and legal constraints. A promising research method, although not yet widely applied in HCI and USEC, is triangulation: the explicit use of multiple methods, measures, and approaches for empirical research and evidence [414, 548]. For example, the use of low-cost surveys with video prototypes is effective for gathering initial results about the social acceptability of multimodal research artefacts [547]. In the next step, the initial findings from surveys can be supported through in-depth focus groups, studies in the lab, or by using a remote VR research approach where participants are immersed in realistic use cases. Individual research methods have different strengths and weaknesses, some of which contribute towards generalisability whilst others are more rigorous and contextual [112].

Studies on VR research artefacts and in virtual environments are valuable when research in the real world is infeasible or particularly challenging due to ethical, legal, and financial constraints. If research in the real world is feasible, VR studies can be part of a triangulated research approach and greatly support (or challenge) the findings from controlled laboratory studies and more naturalistic research in the wild. The research in chapter 4 emphasised the importance of emulating reality and its limitations when moving usability evaluations on prototypes into virtuality. Relying on results from VR studies only can be misleading, as discussed in Section 4.6.5.3 in chapter 4; however, VR studies for simulating the real world are valuable when researchers aim at evaluating the usability of prototypes assuming ideal conditions (for example, no tracking issues for gaze-based input, as shown in chapter 4). If the evaluation of input methods should consider the existing technological limitations, additional implementation work is required to better emulate real-world conditions and other external factors that may impact human behaviour and performance.

All of the discussed factors, including the comments by Mäkelä et al. [291] about “when to consider virtual field studies”, the research on supervised and unsupervised VR studies [277, 339, 340], and the development, deployment, and recruitment process of VR studies [400, 487], must be considered when making a final call about the most suitable research method for a given research problem and research question. Questions like “which input methods do people prefer when providing input on public displays?” and objective usability and security evaluations of various USEC prototypes are feasible to answer by using (remote) VR studies, as shown in chapter 4, chapter 5, and chapter 6. However, pointing at highly realistic human behaviour might only be partially possible using semi-structured interviews, and it is only fully explainable through observations in the wild (for example, as discussed in Section 5.4.6.2 in chapter 5 with the example of shielding PIN entries).

Research Recommendation 2

▷ chapter 3, chapter 4, chapter 5, chapter 6

This thesis recommends the application of VR studies as a complementary research method alongside traditional lab studies and field studies. VR studies can be utilised to inform follow-up studies or to provide support for, or challenge, previous research. Ideally, the use of VR studies for simulating real-world environments and research artefacts considers three main factors:

1. The research question(s) and the overall research aim. For example, is the aim of the study to compare different input methods? Or is the aim to point at realistic human behaviour in the wild? In other words, what conclusions do the researchers aim to draw from research on VR replicas of real-world prototypes and in VR simulations of real-world environments?
2. The complementary research methods, for example, traditional lab and field studies, to support or challenge the findings from VR studies, and vice versa.
3. How the research can be best conducted respecting the financial, ethical and legal constraints, and the development, deployment, and participant recruitment.

7.2.4.3 Lesson 3: VR Studies to Simulate Speculative Future Realities

The USEC expert interviews in chapter 3 highlighted the need for future-oriented research, where use cases of USEC prototypes and methods are more speculative or avant-garde. They mentioned that it is crucial for them to only invest in equipment that is likely to provide a promising future use case. VR studies form a valuable and affordable research method for simulating and studying a potential future of USEC systems that goes beyond state-of-the-art. For example, instead of building “throwaway” prototypes and setting up study environments for individual studies, VR studies can replicate a range of USEC prototypes and real-world environments with little effort, as shown in chapter 5 and chapter 6. Furthermore, applying VR studies allows the implementation and in situ evaluation of more speculative, forward-looking USEC prototypes, as demonstrated with *Tap* and *Hand Menu* in Section 6.3.

As part of this PhD research, various USEC prototypes for public displays and smartphones, virtual laboratory environments, and virtual public spaces were implemented and empirically evaluated in chapter 4, chapter 5, and chapter 6. Without using VR studies, it would have been necessary to have access to various physical devices (for example, an ATM, a smartphone, AR glasses) and different physical environments and social constructs (for example, an ATM in public) to simulate real-world usage of the USEC prototypes and allow for in situ usability and security research. In contrast, VR studies have allowed simulating reality with one VR

headset across different studies without additional hardware, demonstrating how VR studies for simulating real-world USEC research can cover a broad range of prototypes and more futuristic scenarios, i.e., “immersive speculative enactments” [467], without having physical access to those. Immersive Design Fiction, often seen as an extension of the concept of Design Fiction to VR [322], raises questions about the transferability of results obtained in VR to real-world conditions. As this thesis has shown in chapter 4 and chapter 5, many findings from usability and security evaluations in VR transfer to the real world, which is further supported by the research in other subdomains (for example, [291, 344, 376, 528]). Conducting similar human-centred usability and security research as done in this thesis in a real-world context is difficult and sometimes even impossible due to the ethical and legal constraints, which sometimes makes it impossible to assess the ecological validity of data collected through simulations (for example, in VR) [114]. Instead, Dole and Ju [114] argue using the concept of “face validity” as a proxy for judging the researcher’s ecological validity: “when participants take a simulation seriously, one should feel more confident that the study’s results will apply to the real world” [114]. For example, if a participant claims to have experienced a VR replica of a real-world research artefact previously (for example, *Traditional* in chapter 6), then it can be argued that the simulation had real consequences on their experiences and opinions and was perceived seriously enough to draw implications of the research findings [114, 467]. Thus, VR studies for remote research on VR implementations of speculative real-world USEC prototypes provide answers to questions that would otherwise be challenging to find due to the resource-intensive nature of this research and the financial, ethical, and legal constraints. By applying VR studies early on in the research process, researchers can already learn about the social acceptability of novel USEC prototypes for security-sensitive contexts and investigate VR implementations of more futuristic real-world systems that are not yet available or mature enough for testing in the real world.

That being said, the implications of VR studies that aim at exploring possible future realities by creating speculative scenarios in VR might not hold true in the distant future. The technology acceptance [256, 293], including the social acceptability, the ethical implications, and the social implications, may change over time. For example, Williamson et al. [416] showed that “users will develop preferences and change their acceptance rates after multiple trials” and that the user acceptance of gestures increases after even one positive experience.

Research Recommendation 3

▷ chapter 4, chapter 6

This thesis recommends the use of VR studies for simulating future realities and more speculative use cases for USEC prototypes. However, it is recommended to replicate research after a period of time to account for technological advances and changes in the social acceptability of novel interactive systems and experiences.

7.2.4.4 Lesson 4: VR Studies to Enable and Facilitate Location-Agnostic Research on VR Replicas of Real-World Research Artefacts

The research in chapter 4 and chapter 5 has shown how physical spaces can be replicated in VR and used for user testing in different physical locations. For example, Section 5.4 in chapter 5 represented a public space environment in virtuality without having access to the real-world environment. However, there is still a methodological gap in research that enables and facilitates location-agnostic research. In other words, how can the use of VR studies contribute towards location-agnostic research by moving real-world spaces into virtuality and simultaneously moving user testing out of physical laboratory settings? Chapter 6 envisioned such a future where real-world USEC prototypes and environments are first transferred into virtuality and then deployed online to allow for location-agnostic research on USEC prototypes. Such a remote VR research approach revolutionises existing USEC research methods on real-world prototypes and opens a new world of opportunities for the USEC community. For example, remote VR studies can be applied for longitudinal large-scale and cross-country evaluations on VR replicas of real-world prototypes (cf., Section 8.1.2.3).

Sampling participants from across the world also means that participants have different expectations of real-world environments, even more so than when recruiting from a local area only. The qualitative feedback in Section 6.6.8 in chapter 6 on the participants' real-world experiences suggests that large-scale deployments of VR study environments must consider the variety of a user study samples' expectations and infrastructures. For example, as shown in Section 6.6.8, one participant reported withdrawing cash from an ATM from within their car, which does not necessarily align with the ATM scenario utilised in chapter 6. VR studies for simulating real-world research allow the recruitment of large and diverse samples, but they introduce unique challenges when aiming at simulating reality and learning more about behaviour and reality perception from participants across the world. Therefore, the participants' perceived realism of VR replicas of real-world scenarios and prototypes can be affected by local infrastructures that differ across countries and is important to capture in future research.

Research Recommendation 4

▷ chapter 6

This thesis recommends the use of remote VR studies for location-agnostic research on VR replicas of real-world prototypes. However, when using *Remote Virtual Reality for simulating Real-world Research* (RVR³) for cross-country VR study deployments, it is recommended to report and consider the participants' social norms, expectations of reality, and their sociotechnical infrastructures and processes when interpreting the research findings and informing follow-up research.

7.2.4.5 Lesson 5: Replication Studies in VR and in the Real World

The empirical research presented in this thesis has shown that, yes, the use of VR studies for simulating real-world USEC research is feasible and results in valuable USEC research that revolutionises how research on real-world prototypes is currently being conducted. Chapter 4 and chapter 5 have shown that many of the research findings based on VR studies align with their complementary real-world evaluations. However, due to the current VR technology and the novelty of the research artefacts, some of the results deviate from research in the real world, as shown in chapter 4 when comparing CueAuth’s usability in VR to the original real-world study [245] and in chapter 5 when comparing the VR users’ interaction behaviour to observations in the wild. As a result, it remains important to contribute replication studies of and comparison studies to real-world evaluations in the near future. Comparison studies between virtuality and reality require time and effort due to the two-pronged research approach that requires the design and implementation of user studies in VR and the real world. However, in return, they further supply the validation of VR studies as a powerful research method for the broader research field and enable the community to interpret and better understand research resulting from studies based on VR artefacts of real-world systems.

Chapter 6 envisioned a future of VR studies on simulated real-world research artefacts without formal comparisons to real-world counterpart evaluations. The strengths of such a “VR only” research approach has been demonstrated by other works in the literature, for example, to conduct HCI research in a virtual aeroplane environment [350] or to manipulate height [144]. However, despite the existing literature and chapter 6, which provided the first insights into the usability and social acceptability of novel USEC prototypes for user authentication on public displays, it is recommended to contribute additional comparison studies between virtuality and reality in the near future. This will root and further strengthen the use of VR studies as a fundamental, well-evaluated research method for the larger HCI, USEC, and neighbouring research domains.

Research Recommendation 5

▷ chapter 4, chapter 5, chapter 6

This thesis recommends accompanying VR studies with evaluations and comparisons to the real world in the near future. If comparisons to observations and experiments in real life are challenging, it is recommended to incorporate at least one baseline condition in VR, similar to how the research would have been conducted in the real world, as done with *VRO* and *VR Lab* in chapter 5 and *Traditional* in chapter 6.

7.3 Limitations

Limitations of the individual studies were discussed in each chapter. Here, more general limitations of this thesis are discussed. Future research directions are described in chapter 8.

7.3.1 The Focus on Authentication Research

The use of VR studies for USEC research was evaluated within the authentication research field, a major research domain in USEC [155] and the most addressed research topic in the broader USEC field [110]. By studying USEC prototypes that incorporate different input methods, for example, CueAuth with touch, mid-air, and eye-gaze input [245] and ColorPIN with keyboard input [99], this thesis makes contributions that are relevant to a variety of other HCI and USEC prototypes. However, the USEC research field has demonstrated a variety of research interests beyond authentication. For example, research on privacy-preserving prototypes that protect people's privacy in public spaces is often concerned with other metrics (for example, reading speed [402] and text comprehension [241]). VR studies might not be viable and beneficial for all USEC subdomains, particularly when the research in the real world does not necessarily involve hardware or requires no in situ evaluations. However, they can facilitate and support research on USEC prototypes and inform and advance future research that aims to apply VR studies for the broader HCI and USEC research challenges.

7.3.2 The Impact of VR Technology on Research

As shown in the research in Section 4.6 in chapter 4 and Section 5.4 in chapter 5, some of the VR findings deviated from the study results from the real-world studies. These differences, particularly the differences in Section 4.6 in chapter 4, are most likely of technological nature, highlighting that the used VR technology may have not been mature enough to deliver the same experiences someone would experience in reality. Does this mean VR studies are not ready yet for human-centred research because the underlying technology is not mature enough? No, not really. Yet, it means that some of the findings presented in this thesis are based on the used hardware, as discussed in, for example, Section 4.6.5 in chapter 4. Therefore, some of the differences between the VR study results and the real-world counterparts might not hold true in the distant future due to the advances in VR technology. Typical VR challenges such as hand tracking accuracy are yet to be resolved but will likely lose importance as the technology develops over time. This, in return, can contribute to improved usability of some of the tested USEC prototypes (for example, *Tap* in chapter 6). This thesis used the HTC VIVE VR headset, an integrated Tobii eye tracker, a Leap Motion, and the Meta Quest 1/2 VR headset when applying VR studies for simulating real-world research artefacts and

environments. The impact of the technology on the research findings might also depend on the participants' familiarity with said technology. Section 8.1.2.1 discusses the importance of additional extensive replication studies as soon as advanced VR technologies are available and have transitioned more broadly into people's everyday lives.

7.3.3 The User Study Samples: How Representative Are They?

Different recruitment channels were used to validate the applicability of VR studies for USEC research. For the lab studies, which covers the research in Section 4.6 in chapter 4 and Section 5.3 and Section 5.4 in chapter 5, the participants were mainly recruited within the researcher's social circle outside of an academic environment. For the studies that were conducted online, Prolific [394] and social media, including <https://www.xrdrn.org/> (last accessed 22/01/2023), were used for participant recruitment. No specific focus was put into recruiting more diverse groups from different backgrounds and demographics. Furthermore, sample sizes are heavily discussed in the HCI and USEC research communities, with twelve participants being the most common sample size across studies accounting for a full 10% of all CHI studies published in 2014 [66]. The sample sizes in the core studies of this thesis range from twelve participants for the expert interviews in chapter 3 to 25 participants for the remote VR study of real-world USEC prototypes in chapter 6. Although these sample sizes are common for research in laboratory settings, it is essential to acknowledge that this thesis cannot comment on the impact of more extensive and diverse samples on the research findings. To contribute to a long-term solution of the sample size and sample homogeneity discussions in (and beyond) HCI and USEC, the research in chapter 6 has outlined and discussed two research applications, which this thesis picks up again in Section 8.1.2.3 in chapter 8.

7.3.4 The Lengths of the User Studies

Most user studies in this thesis took around 1.5 hours and the participants were exposed to the individual study conditions for a fraction of the time. Although the thesis made use of training sessions to introduce and familiarise the participants with the research artefacts and the virtual environments, comments on the effect of longitudinal research on the use of VR studies cannot be made beyond some of the qualitative data received in the semi-structured interviews (for example, in Section 5.4.5.7 and Section 6.6.8). Koeman et al. [258] highlighted the need for longitudinal studies in HCI and that human-centred research studies longer than a day are rare, with 85% of the CHI 2022 papers studying participants for a day or less. The need for longitudinal research is where the remote VR study in chapter 6 can inform and inspire future research. Section 8.1.2.3 in chapter 8 discusses how the use of VR studies can facilitate longitudinal, cross-country studies on VR implementations of real-world prototypes.

VIII

THE END OF A JOURNEY: CONCLUSION, FUTURE
RESEARCH, AND FINAL REMARKS

Chapter 8

The End of a Journey: Conclusion, Future Research, and Final Remarks

Life is like riding a bicycle. To keep your balance you must keep moving.

– Albert Einstein –

8.1 Conclusion and Ideas for Future Work

Before closing the book and moving on to another stage of life, this thesis concludes the PhD research in Section 8.1.1. It proposes promising future research directions in Section 8.1.2 that further facilitate and contribute towards cutting-edge USEC research. The thesis closes with a final thought about VR studies for real-world research in Section 8.2.

8.1.1 Conclusion

Advanced research methods that support and facilitate USEC research concerned with prototypes are required. For the first time, this thesis investigated the applicability and validity of using VR studies for simulating real-world USEC research on VR implementations of real-world prototypes. Two novel research methods were introduced to advance lab-based shoulder surfing research and enable simulated in situ evaluations of USEC prototypes. These two research methods and, more generally, the use of VR studies for real-world USEC research forge forward human-centred usable security research and contribute to bridging the long-lasting research gap between traditional lab studies and more naturalistic field studies.

Furthermore, this thesis introduced and showcased the use of *Remote Virtual Reality for simulating Real-world Research* (RVR³) for location-independent USEC prototype evalua-

tions. RVR³ supports researchers in addressing homogeneous and small user study samples of real-world prototype studies by enabling the recruitment of participants from across countries. Although the transferability of the VR studies' results depends on the VR technology, this thesis has shown that VR studies for simulating real-world research are beneficial in multiple ways. VR studies make a great addition to the USEC's research arsenal by laying out, for the first time, a research method that allows researchers to create VR replicas of real-world prototypes and environments to then evaluate those through a remote research approach.

To conclude, utilising VR studies for USEC research revolutionises how traditional USEC research on prototypes is currently being conducted and contributes to bridging the methodological gap between laboratory studies and field studies. This thesis put forward a novel USEC research method that has the potential to find widespread adoption and push forward the research in this space through more realistic and impactful human-centred usability and security analyses compared to traditional studies in physical laboratory environments.

8.1.2 Future Research Directions

Many future research paths could further root and strengthen the use of VR studies for real-world research and magnify the impact such a novel research method can have on the broader research field. Researchers have access to a plethora of VR artefacts, libraries, and software development kits that are publicly available and facilitate the development of VR studies for research purposes. For example, RemoteLab [277], Ubiq-exp [487], Microsoft's Rocketbox avatar library [169], Unity's Asset Store [493], and Meta's Oculus Integration SDK [106] support researchers in the future use of VR studies for real-world research. Some specific applications for (remote) VR research were discussed in the previous chapters of this thesis (for example, in Section 6.9.1 in chapter 6). Addressing the bigger picture of this thesis, the most promising future research directions are summarised and discussed here.

8.1.2.1 Research Direction 1: Future (Replication) Studies

First and foremost, the individual studies in each thesis chapter discussed study-specific limitations, which are worth addressing in future works. For example, the key challenges identified in the expert interviews in chapter 3 are based on a USEC expert sample that draws on a large set of publications at top-tier venues. The opinions and challenges of more junior researchers, who might have more hands-on experience but whose work has not yet reached publication, were only partially captured. Future work targeting a more junior sample and incorporating institutions and industries in the discussions is required to identify additional USEC key challenges and provide a holistic overview of the challenges that impact the transition of USEC research into practice. However, due to interviewing experts in the

field, the key challenges of this thesis provide valuable insights into the USEC field and its challenges when designing, implementing, and evaluating prototypes.

Furthermore, this thesis has replicated several previously published USEC prototypes to evaluate the suitability of the use of VR studies for real-world USEC research (for example, CueAuth [245] and ColorPIN [99]). Follow-up replication studies of the VR studies in this thesis can shed further light on the impact of the technology on research findings collected in VR instead of in reality. For example, the study in Section 4.6 in chapter 4 has shown how the impact of VR technology impacts the usability evaluation findings of novel USEC prototypes and that the transferability of results from VR to the real world highly depends on how well the reality is emulated. Follow-up replication studies in the not-too-distant future (\approx five to ten years) can extend the previous research findings. It is vital to consider the impact of societal and technological factors on the results of replication studies. The advancement of VR technology over time may impact the results from VR studies that simulate reality. Additionally, as VR technology becomes more widespread available over time, the society's exposure to extended reality (be it VR or AR) and their experience will increase and likely impact their behaviour, performance, and preferences. Therefore, it is essential to distinguish between and consider both the impact of societal factors and the technological advancements when applying VR studies for future replication studies.

8.1.2.2 Research Direction 2: VR Studies for the Breadth of Human-Centred Real-World Research

This thesis contributed towards solutions of some USEC key challenges that were previously identified in the expert interviews in chapter 3. The contributions in chapter 4, chapter 5, and chapter 6 are based on empirical research from 2019 until 2022. Whilst the sum of the three years of in-depth empirical research contributed to five research recommendations that went through several iterations, they are not carved in stone and are likely to be further extended over time. It is not unlikely that technological advances and people's familiarity and interest in VR technology will impact some of the recommendations outlined in this thesis. Advancements in technology and people's familiarity with VR, along with future work, may further support the recommendations and result in additional research recommendations that fill the gaps unknown at the time of this PhD research.

Extending the research recommendations presented in this thesis is an exciting future research direction to establish the use of VR studies for simulating the entire spectrum of human-centred real-world research. Two of the main challenges of using VR studies for simulating real-world research are to decide a) when and where comparisons to the real world are required and b) what expectations and conclusions researchers can (and should) draw from evaluations on VR replicas and in virtual environments. Individual research projects contributed to the validation

of using VR for real-world research that is concerned with human factors and research artefacts (for example, the work by Voit et al. [528], Savino et al. [439], Weiß et al. [539], and Mäkelä et al. [291]). To contribute towards a future where VR studies complement and advance the breadth of real-world research that is concerned with human factors and research artefacts, a logical next step must be to collectively build upon the recommendations presented in this thesis and establish the use of VR studies as a complementary, well-evaluated research method for the sum of human-centred research, including the USEC, HCI, and their neighbouring research communities.

8.1.2.3 Research Direction 3: Longitudinal and Cross-Country Research

Research artefacts are often not studied for an extended period of time. Koeman [258] highlighted the dominance of lab studies that only last between 60 minutes to 90 minutes. Only a minority of user studies in HCI ($\approx 14\%$) involve participants for longer than a day [258]. Such short-term studies cannot account for learning and novelty effects, potentially resulting in “prematurely embracing or disregarding new concepts” [258]. Learning effects can notably impact research artefacts [99, 102] and there is empirical evidence that people with different backgrounds exhibit different interaction behaviour. For example, Volkamer et al. [529] found that significantly more subjects shield their PINs when paying with their cards in Germany, as compared to the United Kingdom and Sweden. When applying the Security Behavior Intentions Scale [122] to 3,500 participants across seven countries, it was found that people from Asian countries tend to exhibit less secure behaviour than, for example, people from the United States [440]. These findings emphasise the importance of cross-country research as results in one country may not transfer to the many others. However, the expert interviews in chapter 3 revealed that USEC researchers are often reluctant to research USEC prototypes beyond individual lab studies.

“I’ve seen this in rebuttals [...] when I write a review about something [...] and they are like oh well so many other people have published lab studies, why should I have to go out and do something differently [...] it’s a lot harder, it’s a lot more work and as long as I can get this stuff published why should I bother?” - P8

As a result of lab-based research, the corresponding user study samples are often homogeneous and based on a “WEIRD” sample [284]. Taking the research published at CHI from 2016 to 2020 as an example, the majority of research contributions (73%) are based on Western participant samples [284], with over 100 countries not contributing study participants [284, Fig. 2]. Despite the shortcomings of current user study samples for human-centred research [66, 258], it has to be acknowledged that studying geographically diverse samples may not always be required. For example, there is value in user studies that involve specific groups of people and focus on more specific use cases (for example, authentication methods for

people with visual impairments [115,232] or when learning more about the adoption of HCI systems at local institutions [82]). Furthermore, moving all research online and conducting longitudinal cross-country evaluations is neither possible nor desirable. An estimated 2.7 billion people, one-third of the world's population, remain unconnected to the Internet in 2022 [518] and have no access to online platforms and hardware suitable for remote VR studies of similar nature as presented in chapter 6 in this thesis.

Future research must contribute to research methods that enable people to participate in research anywhere and anytime to deepen the understanding of HCI and USEC systems and processes. The use of remote VR studies as applied in chapter 6 allows the research community to amplify the impact of VR studies on the human-centred research field. Future work is required to identify a) the full potential of remote VR studies, b) the challenges associated with longitudinal cross-country evaluations on VR implementations of real-world research artefacts, and c) the places where individual researchers and the community as a whole require and benefit from support.

8.2 A (More Personal) Final Thought: Are VR Studies Taking Over?

To complete this PhD thesis, the following paragraphs contain a more personal thought from the author of this work; therefore, it is written in the first person perspective.

At the end of my PhD, after presenting one of my last PhD studies at an international conference, a researcher approached me with the following question (paraphrased for clarity): “Why do you not claim that your work – using VR studies for simulating real-world research – should replace, rather than complement, traditional lab studies?”. In the following, I share my response with everyone who has read this far to make my thoughts accessible to the broader research community:

I believe that neither lab studies, field studies, (remote) VR studies, nor any other empirical research method will ever be the golden future of human-centred research if considered in isolation. All empirical research methods have their advantages and disadvantages, and it should be of interest to the individual researcher and the research community as a whole to make use of all available research methods and resources to eventually magnify the impact of the research we, as a research community, can have in people's lives.

For the first time, this thesis has proposed and evaluated the use of VR studies as a complementary research method for USEC research on real-world prototypes. Now it is on the broader research community to decide whether the effort and time spent on validating and

showcasing the pros and cons of using VR studies for real-world USEC research will be fruitful and facilitate and positively contribute to USEC research in the long run.

“The goal of academic research in usable security should be to help speed the discovery (and therefore the adoption) of techniques that simultaneously improve both usability and security. Research does this by developing underlying theories that are both explanatory and generative, discovering and validating new techniques, and creating pedagogies for training future practitioners. Ideally, research should allow new UPS techniques to be rapidly tested and either improved or discarded in the lab, rather than having developers test those ideas on unwitting customers.” – Garfinkel and Lipford [155, p. 4]

The quote above by Garfinkel and Lipford on page four of their book called “Usable Security: History, Themes, and Challenges” [155, p. 4] beautifully describes the ideal goal of academic research. The combination of various empirical research methods, including VR studies, can help speed up the discovery, development, and evaluation of usable, secure, and privacy-preserving HCI systems. Taking a step back and glancing again at the comment about claiming that VR studies for simulating real-world research should “replace, rather than complement, traditional lab studies”, it becomes clear to me that aiming to “replace” existing empirical research methods does not contribute towards a future where HCI systems, to which USEC prototypes belong, facilitate and enrich people’s lives.

“It is only when we have convergent information about the same problem gained from different methods that we can talk of accrual of knowledge. But such an accrual, of convergent substance from divergent means, -when and if we can achieve it- is far more robust and generalizable than results of any one study could be.” – McGrath et al. [316]

A breadth of research methods is required to holistically understand when, how, and why people interact with HCI systems the way they do and master the challenges around contributing to the design and implementation of usable, secure, and privacy-preserving systems. I hope that utilising VR studies for USEC research contributes to Garfinkel and Lipford’s goal of academic research and speeds up the discovery of well-evaluated and impactful HCI systems. Finally, before closing the book, the expert interviews in chapter 3 have inspired me greatly and it is in my interest to conclude this thesis with a final remark: After three years of extensive empirical research, I hope that my PhD work, but above all also the comments by the experts in chapter 3, promote profound reflections within the USEC research community on the when, how, and why human-centred usability and security research is being conducted and how we can best transition research artefacts into practice to generate real-world impact.

COVID-19 Statement

This PhD research was conducted during the COVID-19 pandemic¹ from October 2019 until December 2022. Although there are no explicit signs that the pandemic impacted this PhD research and its findings, it cannot be entirely ruled out that the participants' behaviours and preferences were not impacted due to the COVID-19 pandemic.

The research in this thesis complied with the University of Glasgow College of Science & Engineering ethics committee's COVID-19 policies. Precautions, including wearing face masks and utilising disposable hygiene covers for the VR equipment, were taken across all in-person studies to ensure the participants' (and the researcher's) safety and well-being.

¹<https://www.who.int/emergencies/diseases/novel-coronavirus-2019>, last accessed 22/01/2023

Appendix A

Supplemental Materials

Questionnaires, question sets, and surveys can be found in the appendices that follow this one. For questions on the supplemental materials and additional requests (including PDF copies of contributing publications, slide decks for study introductions and conference talks, implementations, etc.) please contact me at florian-mathis@outlook.com. The following appendices outline the surveys, questionnaires, and participant information sheets that were used as part of this PhD research. The standardised questionnaires that have been used in this PhD research, such as the NASA-TLX questionnaire [193], the IPQ [451], the Security Behavior Intentions Scale (SEBIS) [122], and the Affinity for Technology Interaction (ATI) [142], are retrievable through the corresponding original publications. Personal information of the experimenter (for example, phone number) were redacted from the consent forms and the participant information sheets. Qualtrics surveys were shortened to avoid repetitions of conditions and align them better with the thesis' contributions. Videos were created for some of the studies in this thesis to disseminate the research. The videos are publicly available on various YouTube channels (last accessed: 18/02/2023):

Chapter 4

Teaser, Full Talk, and VR Video Material: RepliCueAuth: Validating the Use of a Lab-Based Virtual Reality Setup for Evaluating [139]

- <https://www.youtube.com/watch?v=oStAeg-DhwE> (source: ACM SIGCHI, YouTube)
- https://www.youtube.com/watch?v=3b6_9N-4iFI (source: ACM SIGCHI, YouTube)
- <https://youtube.com/playlist?list=PLs1tzNuOyzfyDx7Y99JatsN42a5MEIxea> (source: Florian Mathis, YouTube)

VR Video Material: Observing Virtual Avatars: The Impact of Avatars' Fidelity on Identifying Interactions [304]

- https://youtube.com/playlist?list=PLs1tzNuOyzfwdvIpth_3NjDC2jb_T6LpI
(source: Florian Mathis, YouTube)

Chapter 5

Teaser, Full Talk, and VR Video Material: Virtual Reality Observations: Using Virtual Reality to Augment Lab Based Shoulder Surfing Research [303]

- <https://www.youtube.com/watch?v=FjhAn6YKQYs> (source: Florian Mathis, YouTube)
- https://www.youtube.com/watch?v=kclKvbd_zYE (source: Florian Mathis, YouTube)
- <https://youtube.com/playlist?list=PLs1tzNuOyzfyQqqFILbfVi4LNGcBVcQ6X>
(source: Florian Mathis, YouTube)

Teaser and Full Talk: Can I Borrow Your ATM! Using Virtual Reality for Simulated In Situ Authentication Research [305]

- <https://www.youtube.com/watch?v=i41Kzfw6M3g> (source: Florian Mathis, YouTube)
- <https://www.youtube.com/watch?v=H981srgIdtE> (source: Florian Mathis, YouTube)

Chapter 6

Teaser: Stay Home! Conducting Remote Usability Evaluations of Novel Real-world Authentication Systems Using Virtual Reality [302]

- <https://www.youtube.com/watch?v=5BgArSKUFS4> (source: Florian Mathis, YouTube)

Full Talk: Remote XR Studies: The Golden Future of HCI Research? [309]

- <https://www.youtube.com/watch?v=cY1uiOuebvI> (source: Florian Mathis, YouTube)

Thesis Talk at SOUPS 2022

Full Talk: Moving Usable Security and Privacy Research Out of the Lab: Adding Virtual Reality to the Research Arsenal [301]

- <https://www.youtube.com/watch?v=doNs7WC-wjU> (source: USENIX, YouTube)

Appendix B

Appendix for Chapter 3

Participant Information Sheet



School of
Computing Science

Understanding Approaches, Challenges, and Requirements of Usable Security Researchers and Practitioners

Participant Information Sheet

Researcher: Florian Mathis f.mathis.1@research.gla.ac.uk

Supervisor: Dr. Mohamed Khamis, Mohamed.khamis@glasgow.ac.uk

Dr. Kami Vaniea, kvaniea@inf.ed.ac.uk

1. Invitation

You are being invited to take part in a research interview. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

Thank you for reading this.

2. Purpose of the Interview

We would like to reach out to you to conduct a short interview via Skype or any other suitable communication tool. Based on an extensive literature review we noticed your contributions to the human-centred security community and are very interested in learning more about your research. The aim of the project is to get an overview of current challenges of the development and evaluation of security systems and to what extent we can enhance usability and security evaluations in the long run. The participation is voluntary. The analysis of the interviews will be published at top-tier venues such as CHI, the premier international conference of Human-Computer Interaction, and SOUPS, Symposium on Usable Privacy and Security.

3. Conditions and Data Storage

The interview will last for approximately 45 minutes and will be recorded and saved directly in the University of Glasgow cloud to keep it confidential (<https://gla-my.sharepoint.com>). Access to the raw data is restricted to the researcher (Florian Mathis) and his supervisors (Dr. Mohamed Khamis, Dr. Kami Vaniea) only. We will anonymise your data (by default). However, we would appreciate it if we are allowed to use parts of your statements in combination with your authority and your publications. This would be beneficial for our HCI community and would also allow readers to see statements in specific contexts.

Essential statement on confidentiality as required by University Ethics Committee:

Confidentiality may be limited and conditional – and the researcher has a duty of care to report to the relevant authorities possible hard/danger to participant or others.

4. Data Usage

The data will be used within our research and is part of Florian Mathis' Research Phd. We will store the audio recordings in the University of Glasgow cloud (<https://gla-my.sharepoint.com>). Access to the raw data is

restricted to the researcher (Florian Mathis) and his supervisors (Dr. Mohamed Khamis, Dr. Kami Vaniea) only. Based on the request of interviewees the data can be destroyed at any point. The data will be kept until beyond the end of the Research PhD (up to 10 years) and findings of the interviews might be re-used for additional research projects within Florians' Research PhD.

5. Who has reviewed the study?

This study adheres to the BPS ethical guidelines, and has been approved by the College of Science and Engineering ethics committee of The University of Glasgow.

6. Funding and Contact

This research is supported by the University of Edinburgh and the University of Glasgow jointly funded PhD studentships: <https://www.gla.ac.uk/research/ourresearchenvironment/prs/uofguofedinphdstudentships/>

The project has been reviewed and approved by the Research Ethics Committee in the School of Computing Science at the University of Glasgow (number protocol: tba). For further information please feel free to get in touch with the researcher f.mathis.1@research.gla.ac.uk (or via phone: +447402698437).

Whilst you are free to discuss your participation in this study with the researcher, if you would like to speak to someone not involved in the study, you may contact the Ethics Committee at Christoph.Scheepers@glasgow.ac.uk.

For further information, or if you wish to receive a summary of the findings of this experiment at a later date, please contact the researcher or the supervisor of this project, details listed below.

Researcher

Florian Mathis

Email: f.mathis.1@research.gla.ac.uk

Tel: [REDACTED]

Supervisor

Dr. Mohamed Khamis

Email: Mohamed.Khamis@glasgow.ac.uk

Tel: [REDACTED]

Data Protection and Confidentiality

Your data will be processed in accordance with the Data Protection Act 1998 (up until 24th May 2018) and the General Data Protection Regulation 2016 (GDPR) thereafter. All information collected about you will be kept strictly confidential. Unless they are anonymised in our records, your data will be referred to by a unique participant number rather than by name. If you consent to being audio recorded, all recordings will be destroyed once they have been transcribed. Your data will only be viewed by the researcher/research team. All electronic data will be stored on a password-protected computer file within the School of Computing Science. All paper records will be stored in a locked filing cabinet within the School of Computing Science. Your consent information will be kept separately from your responses in order to minimise risk in the event of a data breach.

Data Protection Rights

University of Glasgow is a Data Controller for the information you provide. You have the right to access information held about you. Your right of access can be exercised in accordance with the Data Protection Act 1998 (up until 24th May 2018) and the General Data Protection Regulation thereafter. You also have other rights including rights of correction, erasure, objection, and data portability. For more details, including the right to lodge a complaint with the Information Commissioner's Office, please visit www.ico.org.uk. Questions, comments and requests about your personal data can also be sent to the University Data Protection Officer - dp@gla.ac.uk (<https://www.gla.ac.uk/myglasgow/dpfooffice/contact/>)

Consent Form

Understanding Approaches, Challenges, and Requirements of Usable Security Researchers and Practitioners

CONSENT FORM

Before agreeing to this consent form, you should have been given an information sheet to read, which explains the general purpose of this interview and the tasks it involves. If you did not receive this, please inform the researcher (Florian Mathis, f.mathis.1@research.gla.ac.uk). Throughout the interview, the researcher will explain in detail the different activities that you will be completing, but if you have any questions or require medical attention, please do not hesitate to ask. You can withdraw your given statements at any point.

In this study, you are required answer questions within the context of usable security and within one of your predefined publications. The data will be treated as confidential and kept in secure storage at all times.

The material will be used in research publications, both print and online. The interview will take approximately 45 minutes to complete and will be compensated with an £8 Amazon voucher at the end. You can withdraw from the interview at any time. To take part in the interview, the following criteria must be met:

- I am at least 18 years old.
- I am willing to get interviewed by Florian Mathis f.mathis.1@research.gla.ac.uk within the context of usable security.

If you agree in using parts of your data within the context of your authority please tick the box:

Important: The default (i.e., not ticking the box) is that we fully anonymise your data. If you decide to agree that we can use the data within the context of your authority, we would inform you about drafted examples before submitting to any conference in any cases. For instance, we would outline paragraphs that include your statements in combination with your authority. Such could be, for instance, statements such as: "This is according to Dr. XYZ a major limitation in her/his current [...]. In particular, as outlined in her/his work [X] ...".

Understanding Approaches, Challenges, and Requirements of Usable Security Researchers and Practitioners

1. I confirm that I have read and understand the Participant Information Sheet, and understand my Data Protection Rights under GPDR for the above study, and have had the opportunity to ask questions.
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, and am free to omit answering any particular question, without providing a reason.
3. I give consent for my actions to be recorded during the interview.
4. I understand that all data collected from me will be treated confidentially and anonymized by default and will be seen in its raw form only by the experimenters.
5. By default, if published all data collected is not identifiable as coming from me. However, I have the right to agree that the researchers are allowed to use the data within the context of my authority. I am aware of the fact that this permission is granted by ticking the box above.

Experimenter details: Florian Mathis (f.mathis.1@research.gla.ac.uk)

Supervisor details: Dr. Mohamed Khamis (Mohamed.khamis@glasgow.ac.uk) and

Dr. Kami Vaniea (kvaniea@inf.ed.ac.uk)

This study has been approved by the Ethics Committee.

By signing this form, you have read the conditions stated above and agree to take part in the study.

FULL NAME: _____

SIGNATURE: _____

DATE: _____

EMAIL (contact details): _____



University of Glasgow | School of Computing Science



GIST
GLASGOW INTERACTIVE
SYSTEMS GROUP

Interview Invitation

USEC experts were asked if they are willing to be interviewed about their research. Links to some of their USEC papers that were relevant for the interviews were added as part of the interview request. Figure B.1 shows one of the anonymised requests.

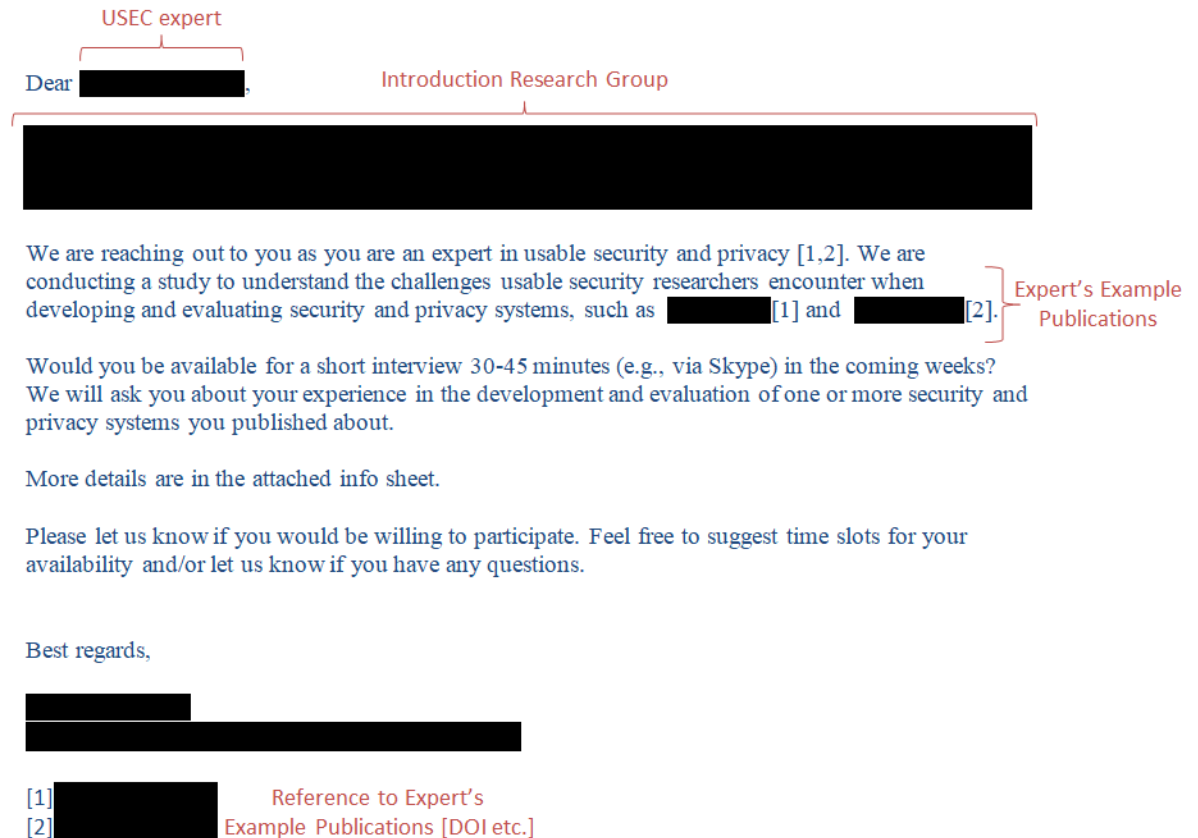


Figure B.1: The interview request included an introduction of the research group, example papers of the expert, and an attached information sheet. Note that some parts in the email are censored for anonymity reasons.

Semi-structured Interview Questions

1. Typical Research Journey from Idea to Publication

- (a) Let us consider a novel security or privacy-preserving system: If we walk along the path, from an initial idea to the final publication, how would these steps look like?
- (b) With a focus on each specific step: What are challenges or limitations that you encountered when designing, implementing, and evaluating such prototype systems?

2. Research Challenges and Limitations

- (a) Were there limitations that you encountered when iteratively designing, implementing, and evaluating prototype systems?
- (b) What were the most challenging parts when developing [*experts' prototype system*]? Were there any limitations or things you would have preferred to do differently but could not do so?
- (c) What are your thoughts regarding the approaches USEC researchers apply to evaluate privacy and security?

3. The Ecological Validity of Current Evaluations

- (a) Do you see *controlled lab studies* as the "way to go" to evaluate security and privacy-aware prototype systems?
- (b) What are your thoughts on the different study types (e.g., lab, online, or in-the-wild studies) USEC researchers currently apply to assess a prototype system's privacy/security and usability?
- (c) What keeps USEC researchers and practitioners away from investigating security and privacy-aware systems in more realistic contexts (e.g., at a public space such as a bus station)?
- (d) Would you prefer to see "*more realistic*" studies, for example, field studies? Can you please outline why or why not you think so?
- (e) Talking about the ecological validity of human-centered evaluations: What conditions have in your opinion a significant influence on the validity of research findings?
- (f) Let's assume you have the time and resources available to re-run parts of your [*papers study*] again. Would there be anything you would like to investigate in addition to the metrics you have already mentioned in your publications?

Appendix C

Appendix for Chapter 4

Participant Information Sheet



School of
Computing Science

VirSec: Comparing Usability and Security of Touch, Mid-Air Gestures, and Gaze gathered in virtual reality to the real world

Participant Information Sheet

Researcher: Florian Mathis f.mathis.1@research.gla.ac.uk

Supervisor: Dr. Mohamed Khamis, Mohamed.khamis@glasgow.ac.uk
(2nd) Dr. Kami Vanica, kvanica@inf.ed.ac.uk

IMPORTANT – Exclusion criteria

In order to take part in this study, you must meet the following requirements:

1. Aged 18 or over
2. No history (personal and family) of epileptic seizures, strokes, or photosensitivity
3. Not a member of any of the following groups
 - a. Pregnant women
 - b. The elderly
 - c. Sufferers of any serious medical conditions i.e. you fall into one of the following categories
 - i. Inpatient care
 - ii. Incapacity
 - iii. Chronic serious health conditions
 - iv. Permanent or long term conditions
 - v. Conditions requiring multiple treatments
 - d. Sleep deprived
 - e. Under the influence of alcohol
 - f. Previously suffered concussion or traumatic brain injury
 - g. Prone to dizziness from immersive virtual experiences
 - h. Sufferers of panic attacks or generalised anxiety disorders which might be provoked by wearing headphones / being unable to hear your surroundings
 - i. Prone to issues with balance or motor function (i.e. you can walk around a room over the course of an hour).
4. Be comfortable with wearing a head-mounted display (HMD) such as a HTC VIVE, Oculus Quest

1. Invitation

You are being invited to take voluntarily part in a research experiment. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

Thank you for reading this.

2. Purpose of the User Study

We would like to reach out to you to participate in a paid study examining the feasibility of virtual reality as a testbed for usability and security evaluations. Your participation is voluntary, and you are free to withdraw at any time, without giving any reason, and you are free to omit answering any particular question, without providing a reason. The analysis of this experiment will be published at top-tier venues such as CHI, the

premier international conference of Human-Computer Interaction, SOUPS, Symposium on Usable Privacy and Security, and IEEE VR. All publications are fully anonymised and findings and specific measurements cannot be traced back to you.

The study is exploring the level of transferability of results gathered in VR and the real world. The objective is to see to what extent results gathered in VR match results gathered in the real world.

3. What will happen to me if I take part?

Changed based on Study:

Study 1)

You will be fitted with a virtual reality headset such that it sits comfortably on your head, and you can hear correctly. This will occur in a room at the School of Computing Science (Sir Alwyn Williams Building or Lilybank Gardens). You are then going to enter a password on a public display in a virtual environment with a) touch gestures, b) mid-air gestures, and c) gaze (e.g., smooth pursuits). We will demonstrate all interaction techniques and guide you in performing these interactions during a training session. During all three inputs we will capture following data:

- Virtual Reality:
 - We are going to record the graphical representation of the entire virtual environment. This includes eye movements (with the integrated Tobii eyetracker), body movements (with VIVE Tracker and HTC Vive Controllers), and static objects within the environment (e.g., the authentication scheme). For all recordings we will use appropriate file extensions (e.g., .csv, .txt) and store them in separate files directly in the University of Glasgow cloud.
- Real World:
 - We are going to record you from the real world during your tasks within the virtual world. This includes the recording of a video from different perspectives and/or taking photographs. You can see an example of this in the picture below. The picture shows a researcher (Florian Mathis) performing a task in virtual reality. Florian is equipped with the HTC VIVE and two HTC VIVE controllers. The way we are going to record your interactions is similar to the picture below.



Study 2

You are going to watch authentications that were recorded in a virtual environment on a desktop computer. This allows you to perform observation attacks on authentications. The experiment will occur in a room at the School of Computing Science (Sir Alwyn Williams Building or Lilybank Gardens).

At the end the researcher (Florian Mathis) will ask some additional questions within the context of the experiment (semi-structured interview). This helps him and his supervisors to better understand the experience you have undertaken. At the end of each 1h session we will hand out the £8 Amazon voucher.

4. Why have I been chosen?

Your participation has been solicited through emails or notice board postings to which you replied. Your participation is voluntary, and you are free to withdraw at any time, without giving any reason, and you are free to omit answering any particular question, without providing a reason.

5. Conditions and Data Storage

Each experiment will last for approximately 1 hour. All gathered data during the session will be stored directly in the University of Glasgow cloud to keep it confidential (<https://gla-my.sharepoint.com>). Access to the raw data is restricted to the researcher (Florian Mathis) and his supervisors (Dr. Mohamed Khamis, Dr. Kami Vaniea) only. Your data is fully anonymised and there is no way to trace it back to you. The results of the study may appear in a number of published studies, in a confidential format where anonymity is preserved. Based on your agreement we will use recordings (and screenshots of those recordings) as video and image material for scientific papers and/or presentations at conferences.

6. Data Usage

The data will be used within our research and is part of Florian Mathis' Research Phd. We will store the raw data in the University of Glasgow cloud (<https://gla-my.sharepoint.com>). Access to the raw data is restricted to the researcher (Florian Mathis) and his supervisors (Dr. Mohamed Khamis, Dr. Kami Vaniea) only. Based on the request of participants the data can be destroyed at any point. The data will be kept until beyond the

end of the Research PhD (up to 10 years) and findings of the experiment might be re-used for additional research projects within Florians' Research PhD.

7. Who has reviewed the study?

This study adheres to the BPS ethical guidelines, and has been approved by the College of Science and Engineering ethics committee of The University of Glasgow.

8. Funding and Contact

This research is supported by the University of Edinburgh and the University of Glasgow jointly funded PhD studentships: <https://www.gla.ac.uk/research/ourresearchenvironment/prs/uofguofedinphdstudentships/>

The project has been reviewed and approved by the Research Ethics Committee in the School of Computing Science at the University of Glasgow (number protocol: tba). For further information please feel free to get in touch with the researcher f.mathis.1@research.gla.ac.uk (or via phone: [REDACTED]).

Whilst you are free to discuss your participation in this study with the researcher, if you would like to speak to someone not involved in the study, you may contact the Ethics Committee at Christoph.Scheepers@glasgow.ac.uk.

For further information, or if you wish to receive a summary of the findings of this experiment at a later date, please contact the researcher or the supervisor of this project, details listed below.

Data Protection and Confidentiality

Your data will be processed in accordance with the Data Protection Act 1998 (up until 24th May 2018) and the General Data Protection Regulation 2016 (GDPR) thereafter. All information collected about you will be kept strictly confidential. Unless they are anonymised in our records, your data will be referred to by a unique participant number rather than by name. If you consent to being audio recorded, all recordings will be destroyed once they have been transcribed. Your data will only be viewed by the researcher/research team. All electronic data will be stored on a password-protected computer file within the School of Computing Science. All paper records will be stored in a locked filing cabinet within the School of Computing Science. Your consent information will be kept separately from your responses in order to minimise risk in the event of a data breach.

Data Protection Rights

University of Glasgow is a Data Controller for the information you provide. You have the right to access information held about you. Your right of access can be exercised in accordance with the Data Protection Act 1998 (up until 24th May 2018) and the General Data Protection Regulation thereafter. You also have other rights including rights of correction, erasure, objection, and data portability. For more details, including the right to lodge a complaint with the Information Commissioner's Office, please visit www.ico.org.uk. Questions, comments and requests about your personal data can also be sent to the University Data Protection Officer - dp@gla.ac.uk (<https://www.gla.ac.uk/myglasgow/dpfooffice/contact/>)

Thank you for volunteering to take part in this study. If there are any questions or issues, or if you wish to receive a summary of the findings of this experiment at a later date, please feel free to get in touch with the researcher at any time.

Researcher

Florian Mathis
Email: f.mathis.1@research.gla.ac.uk
Tel: [REDACTED]

Supervisor

Dr. Mohamed Khamis
Email: Mohamed.Khamis@glasgow.ac.uk
Tel: [REDACTED]

Consent Form



CONSENT FORM

Title of Experiment: VirSec: Comparing Usability and Security of Touch, Mid-Air Gestures, and Gaze gathered in virtual reality to the real world

Experimenter details: Florian Mathis (f.mathis.1@research.gla.ac.uk)

Supervisor details: Dr. Mohamed Khamis (Mohamed.khamis@glasgow.ac.uk) and

Dr. Kami Vaniea (kvaniea@inf.ed.ac.uk)

Before agreeing to this consent form, you should have been given an information sheet to read, which outlines exclusion criteria and explains the general purpose of this experiment and the tasks it involves. If you did not receive this, please inform the researcher (Florian Mathis, f.mathis.1@research.gla.ac.uk).

Please tick the box after each statement to indicate that you have read and understand the statement, and that you agree with it.

1. I confirm that I have read and understand the Participant Information sheet, and understand my Data Protection Rights under GDPR for the above study, and have had the opportunity to ask questions.	
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, and am free to omit answering any particular question, without providing a reason.	
3. I give consent for my actions to be recorded (audio and video) during the study.	
4. I understand that all data collected from me will be treated confidentially and anonymized, will be seen in its raw form only by the experimenters, and if published will not be identifiable as coming from me.	
5. I agree that the researchers can use video recordings for public outreach, for instance, showing parts of the recordings at conference venues and/or use the material in videos to showcase the system.	
6. By agreeing to take part in this study also agree that recordings can be used for follow-up evaluations by researchers in the school of computing science and their collaborators. This includes the investigation of the recordings of my interactions within the virtual environment and the video recordings from the real world while performing tasks.	

7. I agree that the researchers are allowed to archive all recordings taken during the experiment (e.g., video recordings with a camera in the real world) in online repositories such as Enlighten: Research Data: http://researchdata.gla.ac.uk/ . I am aware of the fact that I can get in touch with the researchers at any time to demand the deletion or retrieval of these recordings.	
8. I agree to take part in the above study (VirSec: Comparing Usability and Security of Touch, Mid-Air Gestures, and Gaze gathered in virtual reality to the real world).	

This study has been approved by the Ethics Committee.

By signing this form, you have read the conditions stated above and agree to take part in the study.

FULL NAME: _____

SIGNATURE: _____

DATE and PLACE: _____

Semi-structured Interview Questions

The semi-structured interviews in the usability study were loosely guided by the following questions that were asked for all three input methods: touch, mid-air, and eye gaze.

1. General Questions

- Please tell us how you would feel using this method in public.
- Please tell us (a) what you liked; and (b) what you did not like when using this method.
- Is there anything in particular that you would like to improve in this method?
- Have you used this method previously? If yes, where?
- How did you feel when interacting with the input method? Would you define it as a positive or negative experience?

2. VR-specific Questions [Please consider the situation where you interact with the authentication scheme you have just experienced in the real world.]

- Can you please walk us through the input method and tell us what differences may appear when using this method in the real world rather than in VR as just experienced?
- Do you think the virtual environment affected you in the way you provided input with the method?

At the end, the participants were asked if they have any additional comments or questions.

Participant Information Sheet and Consent Form



THE UNIVERSITY *of* EDINBURGH
informatics

Introduction

Thank you for your interest in participating in this survey about gesture identifications performed in Virtual Reality (VR).

You are being invited to take part in this research study. Before you decide to take part it is important for you to understand why this research is being done and what it will involve.

Please take time to read the following information carefully. If you have any questions regarding this research please contact the lead researcher Florian Mathis (florian.mathis@glasgow.ac.uk, joint PhD at the University of Glasgow and University of Edinburgh).

Purpose of the study

The purpose of this study is to examine the impact of avatars used to represent users in a virtual environment on identifying users' movements. We aim to understand what avatar granularity is required to evaluate security systems in a follow-up step. The analysis of this survey will be used for further developments in Unity 3D and published at top-tier venues (e.g., IEEE VR, ACM CHI). All publications are fully anonymised and findings and specific measurements cannot be traced back to you.

Procedure of the study

In the first section we collect general information from you. This includes your experience with Virtual Reality (VR). Note that this study does not require any experience with VR. You are then going to watch videos of interactions that were **a)** pre-recorded in a virtual environment on a desktop computer or **b)** pre-recorded in the real world. The pre-recorded videos allow you to perform so called "observation attacks". In this study, your task is to identify specific gestures performed by a human. In addition to watching the videos, you have to guess which gestures you have just experienced and fill in your perceived identification performance on a 5-point Likert scale. After performing multiple "observation attacks" you are going to self-report your perceived mental workload with the

NASA TLX [1] questionnaire that we embedded in this survey. We divided the survey into four parts with three breaks in-between. We highly encourage you to take these breaks during the study. We end the survey by asking you to rank the avatars used in the study and provide general feedback.

Why have I been chosen?

Your participation has been solicited through Prolific [2].

What are the possible benefits of taking part?

By completing this survey you will be reimbursed according to the principle of "ethical rewards" on Prolific [2]. You will also be given the opportunity to find out more about this research by contacting the researcher:

florian.mathis@glasgow.ac.uk

Data Storage and Usage

The survey takes approximately 1h (without any breaks). All gathered data during the session will be stored directly in the University of Edinburgh cloud to keep it confidential. Access to the raw data is restricted to the researcher (Florian Mathis) and his supervisors (Dr. Mohamed Khamis, Dr. Kami Vaniea) only. Your data is fully anonymised and there is no way to trace it back to you. The results of the study may appear in a number of published studies, in a confidential format where anonymity is preserved. Based on your agreement we will use findings for scientific papers and/or presentations at conferences.

Who has reviewed this study?

This study adheres to the BPS ethical guidelines and has been approved by the College of Science and Engineering ethics committee of The University of Glasgow (#300190215) and by the Informatics Forum ethics committee of the University of Edinburgh.

Funding and Contact

This research is supported by the University of Edinburgh and the University of Glasgow jointly funded PhD studentships.

For further information please feel free to get in touch with the lead researcher:

florian.mathis@glasgow.ac.uk

For further information, or if you wish to receive a summary of the findings of this experiment at a later date, please contact the lead researcher Florian Mathis (florian.mathis@glasgow.ac.uk).

Consent Form

Please confirm your participation in this study by completing this consent form:

- I confirm that I have read and understand the information above (information sheet), and understand my Data Protection Rights under GDPR for the above study, and know how to contact the researchers to ask questions.
- I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, and am free to omit answering any particular question, without providing a reason.
- I understand that all data collected from me will be treated confidentially and anonymized, will be seen in its raw form only by the experimenters, and if published will not be identifiable as coming from me.
- I agree that the anonymised data can be used for follow-up evaluations by researchers in the school of computing science and their collaborators. Note that this project is part of a joint PhD between the University of Glasgow and the University of Edinburgh.
- I agree that the researchers are allowed to archive aggregated data and findings (anonymised) taken during the experiment in online repositories such as Enlighten: Research Data: <http://researchdata.gla.ac.uk/>. I am aware of the fact that I can get in touch with the researchers at any time to demand the deletion or retrieval of the recorded data.
- I agree to take part in this survey.

By selecting **“I AGREE TO ALL OF THE ABOVE”** I have read the conditions stated above and agree to take part in the study. This also takes me to the first page of the study.

Qualtrics: Thesis Survey for Section 4.7

Start of Block: Prolific ID



Q614 Please enter your Prolific ID here:

End of Block: Prolific ID

Start of Block: Demographics

Q4 Do you have a left-right disorientation? A left-right disorientation means that it is challenging for you to differentiate between the left and right side.

- Yes (1)
- No (2)
- Prefer not to say (8)

Q5 Do you have any vision impairment diagnosis? If yes, please specify otherwise leave it empty.

Q6 Have you heard about the term "Virtual Reality" (or short VR) before?

- Yes (1)
- No (2)
-

Q7 Have you experienced a Virtual Reality System (e.g. HTC Vive, Oculus Quest) before?

- Yes (1)
- No (2)
-

Q8 How many times (approximately) have you experienced VR within the last five years (2015-2020)?

- 0 times. (1)
- 1-5 times, with a short usage duration. (2)
- 1-5 times, with a long usage duration. (3)
- 5-10 times, with a short usage duration. (4)
- 5-10 times, with a long usage duration. (5)
- > 10 times. (6)
-

Q9

With which role(s) do you identify most?

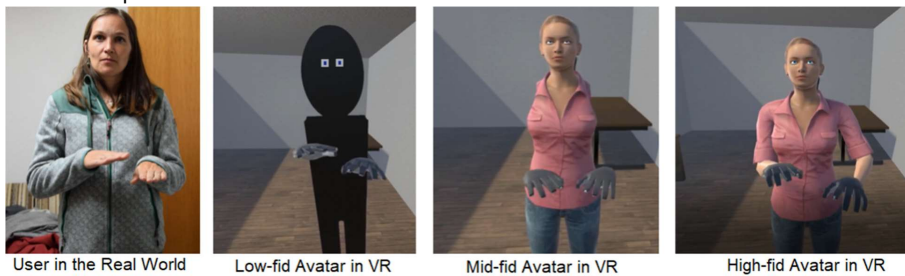
- I consider myself as an end-user and consumer of VR content (e.g., consuming games in VR). (1)
- I consider myself as a VR developer (e.g., developing VR applications). (2)
- I consider myself as a content creator for VR applications. (3)
- I consider myself as a researcher in VR (e.g., doing a VR-focused PhD). (4)
- Other: (5) _____
-

Q10 Timing
 First Click (1)
 Last Click (2)
 Page Submit (3)
 Click Count (4)

End of Block: Demographics

Start of Block: Introduction all,text

Q1 Your task is to watch pre-recorded videos that show a user providing input (more on the input techniques below). We experimented with three different avatars to represent a user in a virtual environment. Below you can see a user in the real world and three different avatars that we use to represent the user in a virtual environment.



Q2

In addition to the different avatars depicted above, we experiment with three different input techniques. Note that with input techniques we refer to the way a user provides input to a system.

The input techniques are **touch** gestures, **mid-air** gestures, and **eye gaze** gestures (smooth pursuits). The descriptions below should give you an idea of the techniques and the gestures. Please read them carefully. We will also show you introduction videos at a later stage.

Touch gestures:

Touch gestures are gestures that are performed on the surface of the input device, in this case on a situated display. The user in the video performs touch gestures in all four directions (left,

right, up, down) directly on the surface. In addition to the gestures in the specific directions, the user also performs a so called "single tap" on the surface without moving their finger towards a specific direction.

This means we distinguish between 5 touch gestures: **left / right / up / down / tap**

Mid-air gestures:

Mid-air gestures are gestures that are performed in the air in front of the input device without touching the input device. The user in the video performs mid-air gestures in all four directions (left, right, up, down) with their right arm. In addition, the user can perform a so called "front gesture" to provide input. This is when the user moves their arm to the front.

This means we distinguish between 5 mid-air gestures: **left / right / up / down / front**

Eye Gaze gestures (Smooth Pursuits):

Smooth pursuits are eye movements that closely follow a moving object. The user in the video performs smooth pursuits by following moving targets on the situated display. In this study, the user performs so called "linear diagonal" movements, clockwise movements, counter-clockwise movements, and zigzag vertical/horizontal movements.

This means we distinguish between 8 smooth pursuit gestures: **linear diagonal (4) / circular (2) / zigzag (2)**

Please proceed to the next page when you are ready to watch the first introduction video to familiarize yourself with the avatars and the gestures.

Q3 Timing
 First Click (1)
 Last Click (2)
 Page Submit (3)
 Click Count (4)

End of Block: Introduction all,text

Start of Block: Introduction Video - Eye Gaze Gestures

JS

Q1

In this part of the study we focus on: Eye Gaze gestures (Smooth Pursuits)

Your **task** is to **identify the gesture** shown in the video **from user's (or avatar's) perspective**.

Please watch the introduction video for the input technique **Eye Gaze** as many times you want. Please make sure that you are familiar with both the **input technique** and the set of **gestures** used in this study before starting with the identification task.

Smooth pursuits are eye movements that closely follow a moving object. The user in the video performs smooth pursuits by following moving targets on the situated display. In this study, the user performs so called "linear diagonal" movements, clockwise movements, counter-clockwise movements, and zigzag vertical/horizontal movements.

This means we distinguish between 8 smooth pursuit gestures: linear diagonal (4) / circular (2) / zigzag (2)

Q2 Please confirm that you have watched the entire video at least one time and fully understand the input technique and the set of gestures by clicking "I have watched the video above at least one time and I understand how the input technique and corresponding gestures work."

It is important to watch and understand the introduction video to be able to answer the following questions. Therefore, we hereby ask you to watch the video again if the input technique and/or the gestures are not clear.

I have watched the video above at least one time and I understand how the input technique and corresponding gestures work. (1)

Q3 Timing

First Click (1)

Last Click (2)

Page Submit (3)

Click Count (4)

End of Block: Introduction Video - Eye Gaze Gestures

Start of Block: Context check - Gaze Gesture

Q562 After watching the introduction video, can you please tell us (1) which input modality we are now investigating and (2) which of the following best describes your task?



Q563 (1) Please tell us, which of the following input modalities best describes what we are now investigating.

- Eye movements to perform Smooth Pursuits (1)
- Hand movements to perform Mid-air Gestures (2)
- Hand movements to perform Touch Gestures (3)
- Head movements to perform Head-based Gestures (4)



Q564 (2) Please tell us, which of the following best describes your next task.

- My task is to identify the gesture shown in the video from user's/avatar's perspective. (1)
- My task is to mimic the gesture I see in the video and report back how I feel about performing them. (2)
- My task is to watch the videos and report back which gesture I like the most. (3)

End of Block: Context check - Gaze Gesture

Start of Block: Participant Note



Q1

You are going to watch multiple videos and each video starts immediately after each page load. Your task is to identify the gesture shown in the video from user's/avatar's perspective. Please note that you can watch following videos only one time. It is not possible to replay the video a second time. Each video starts with a countdown from 3 to 0.

Please proceed to the next page when you are ready.

Q2 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Participant Note

Start of Block: Eye Gaze Gesture Real-World Diagonal_bottom_left_top_right

JS

Q91 Video: Eye Gaze Gesture Interaction

Q146 Please indicate which gesture the user performed (from the user's perspective):

- The user performed clockwise circular eye movements. (1)
 - The user performed counter-clockwise circular eye movements. (2)
 - The user performed linear diagonal eye movements from top left to bottom right. (3)
 - The user performed linear diagonal eye movements from top right to bottom left. (4)
 - The user performed linear diagonal eye movements from bottom left to top right. (5)
 - The user performed linear diagonal eye movements from bottom right to top left. (6)
 - The user performed horizontal zig-zag eye movements along the x-axis. (7)
 - The user performed vertical zig-zag eye movements along the y-axis. (8)
-

Q148 Please indicate how confident you are with your answer.

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
I am confident that I could correctly identify the type of gesture. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It was easy to identify the gesture. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q327 Each video should play only one time. If you viewed the video multiple times, please let us know how many times you viewed it (otherwise leave it empty).

Q226 If the video did not load properly or you faced any other issues that did not let you to watch the video above please report this here. Note that the answer **will not** impact your reimbursement.

The video playback did not work in this case. (1)

- Q152 Timing
- First Click (1)
- Last Click (2)
- Page Submit (3)
- Click Count (4)

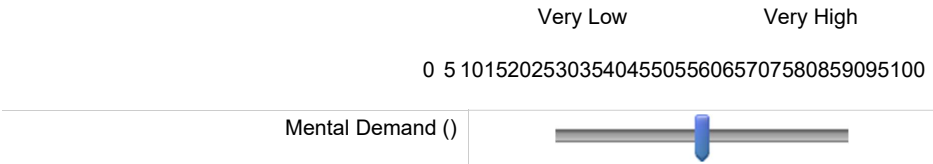
[Repeated for all Eye Gaze Gestures]

Start of Block: Perception and Nasa-TLX#1 [gaze]

Q83 Please indicate your perceived mental workload when running the identification tasks.

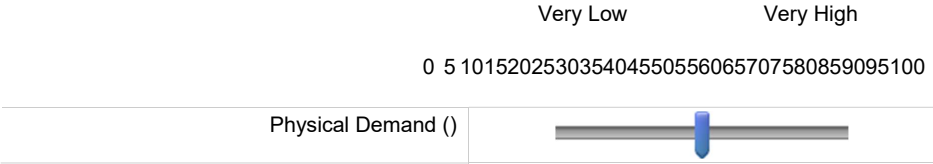
Mental Demand

How much mental and perceptual activity was required? Was the task easy or demanding, simple or complex?



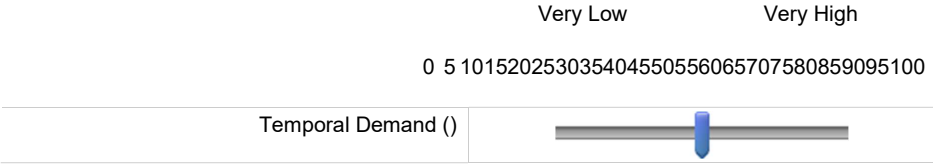
Q84 Physical Demand

How much physical activity was required? Was the task easy or demanding, slack or strenuous?



Q85 Temporal Demand

How much time pressure did you feel due to the pace at which the tasks or task elements occurred? Was the pace slow or rapid?



Q86 Performance

How successful were you in performing the task? How satisfied were you with your performance?

Perfect Failure

0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100



Q87 Effort

How hard did you have to work (mentally and physically) to accomplish your level of performance?

Very Low Very High

0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100



Q88 Frustration Level

How irritated, stressed, and annoyed versus content, relaxed, and complacent did you feel during the task?

Very Low Very High

0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100



[Repeated for Avatars within Eye Gaze]

Start of Block: block_for_break

Q1 We have now reached a point where taking a break is well deserved.

Get some tea, coffee, or fresh air if you would like to before continuing with the remaining part(s) of the survey.

Q2 Timing

First Click (1)

Last Click (2)

Page Submit (3)

Click Count (4)

End of Block: block_for_break

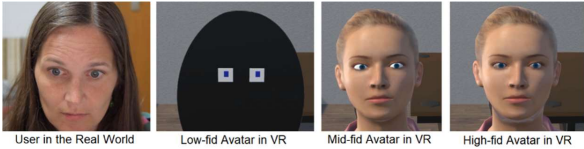
[Repeated for Touch Gestures and Mid-Air Gestures]

Q1

Perceived Identification Ease

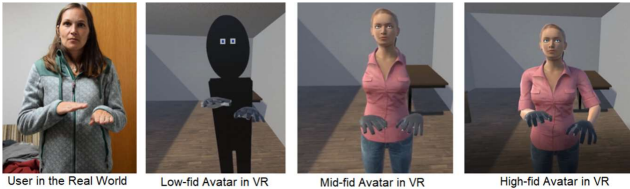
Please answer the following questions by referring back to the avatars displayed below.

Q2 Perceived Ease of Identifying Eye Gaze Gestures (Smooth Pursuits)



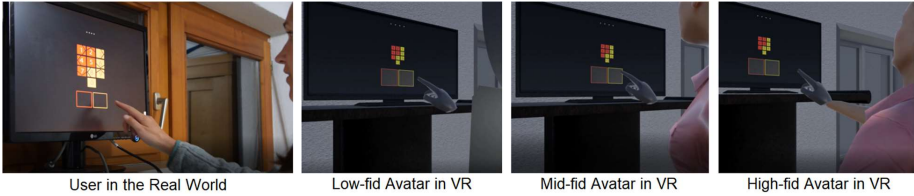
	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
It was easy to identify eye gaze gestures in the real-world recording. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It was easy to identify eye gaze gestures when the low-fidelity avatar was used in the videos. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It was easy to identify eye gaze gestures when the mid-fidelity avatar was used in the videos. (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It was easy to identify eye gaze gestures when the high-fidelity avatar was used in the videos. (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q5 Perceived Ease of Identifying Mid-Air Gestures



	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
It was easy to identify mid-air gestures in the real-world recording. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It was easy to identify mid-air gestures when the low-fidelity avatar was used in the videos. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It was easy to identify mid-air gestures when the mid-fidelity avatar was used in the videos. (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It was easy to identify mid-air gestures when the high-fidelity avatar was used in the videos. (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q6 Perceived Ease of Identifying Touch Gestures



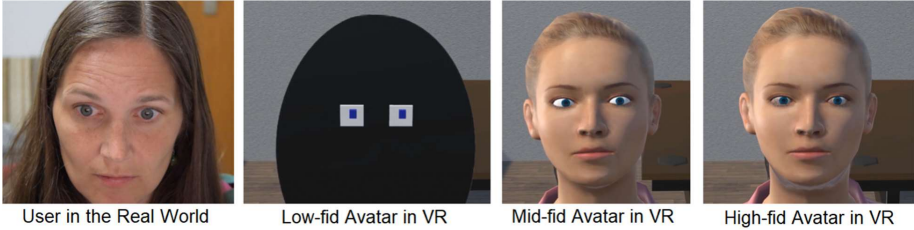
	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
It was easy to identify touch gestures in the real-world recording. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It was easy to identify touch gestures when the low-fidelity avatar was used in the videos. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It was easy to identify touch gestures when the mid-fidelity avatar was used in the videos. (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It was easy to identify touch gestures when the high-fidelity avatar was used in the videos. (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q7

Ranking Please rank the avatars we used to represent a user in a virtual environment. In particular, **which avatar** provided you with information that was the **most helpful**? **Rank 1** should be the one that was the **most helpful representation of the user**, whereas **rank 4** the one that was the **least helpful**. Note that you can give the same rank to two (or more) if you wish to do so. Please also consider the user in the real world. For instance, if one avatar makes the identification task easier than viewing the user in the real world the latter should then be ranked lower (e.g., Rank 2 instead of Rank 1).

Please perform the ranking for **each input technique**: eye gaze, mid-air, and touch gestures.

Q8 Eye Gaze Gestures (Smooth Pursuits)



	Rank 1 (1)	Rank 2 (2)	Rank 3 (3)	Rank 4 (4)
User in the real world (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Low-fidelity Avatar (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mid-fidelity Avatar (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
High-Fidelity Avatar (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q11 Please justify your ranking: Why did you perform the ranking the way you did it? (~2-3 sentences)

Q9 Mid-Air Gestures



User in the Real World



Low-fid Avatar in VR



Mid-fid Avatar in VR

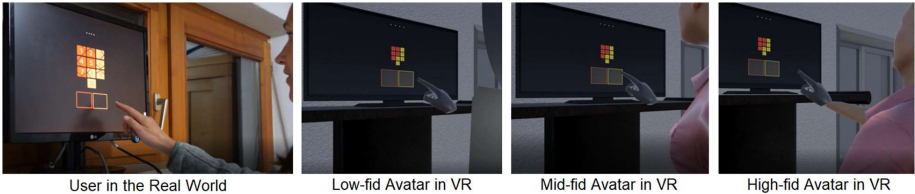


High-fid Avatar in VR

	Rank 1 (1)	Rank 2 (2)	Rank 3 (3)	Rank 4 (4)
User in the real world (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Low-fidelity Avatar (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mid-fidelity Avatar (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
High-Fidelity Avatar (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q4 Please justify your ranking: Why did you perform the ranking the way you did it? (~2-3 sentences)

Q3 Touch Gestures



	Rank 1 (1)	Rank 2 (2)	Rank 3 (3)	Rank 4 (4)
User in the real world (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Low-fidelity Avatar (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mid-fidelity Avatar (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
High-Fidelity Avatar (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q10 Please justify your ranking: Why did you perform the ranking the way you did it? (~2-3 sentences)

Q12 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: End Study Block

Start of Block: Feedback

Q613
Feedback, Suggestions, and Problems Please use the text box below if you faced any issues (e.g., lack of clarity, technical problems) or if you want to give us additional feedback and suggestions for improvement. We highly appreciate your feedback.

End of Block: Feedback

Start of Block: Thank you

Q518 Thank you for participating in this survey! For further information, or if you wish to receive a summary of the findings of this experiment at a later date, please contact the lead researcher Florian Mathis (florian.mathis@glasgow.ac.uk).

Please continue to the next page to return to prolific and to confirm that you participated in this study.

End of Block: Thank you

Participant Information Sheet and Consent Form



THE UNIVERSITY *of* EDINBURGH
informatics

Introduction

Thank you for your interest in participating in this survey about the evaluation of a system's resistance to observation attacks in VR.

You are being invited to take part in this research study. Before you decide to take part it is important for you to understand why this research is being done and what it will involve.

Please take time to read the following information carefully. If you have any questions regarding this research please contact the lead researcher Florian Mathis (florian.mathis@glasgow.ac.uk, joint PhD at the University of Glasgow and University of Edinburgh).

Purpose of the study

We would like to reach out to you to participate in a paid study examining the feasibility of virtual reality as a testbed for usability and security evaluations. Your participation is voluntary, and you are free to withdraw at any time, without giving any reason, and you are free to omit answering any particular question, without providing a reason. The analysis of this experiment will be published at top-tier venues such as CHI, the premier international conference of Human-Computer Interaction, SOUPS, Symposium on Usable Privacy and Security, and IEEE VR. All publications are fully anonymised and findings and specific measurements cannot be traced back to you.

The study is exploring the level of transferability of results gathered in VR to the real world. Specifically, the objective is to see to what extent results from observation attacks on authentications made in VR match with results from observation attacks in the real world.

What will happen to me if I take part?

You are going to watch authentications on your desktop computer that were recorded in a virtual environment. This allows you to perform observation attacks on authentications. You are then asked to provide up to three guesses and report on your confidence. At the end of the study, we will ask some additional study-specific questions.

Why have I been chosen?

Your participation has been solicited through Prolific.

What are the possible benefits of taking part?

By completing this survey you will receive £7.50. Additionally, 1 out of the 22 participants who take part in this survey has the chance to win an additional £7.50 based on their performance. Chances of winning increases with the number of successfully attacked PINs. Partial correct PINs also contribute to the chance of winning additional £7.50. The closer the guesses to the correct PIN the higher the chances.

You will also be given the opportunity to find out more about this research by contacting the researcher: florian.mathis@glasgow.ac.uk

Data Storage and Usage

All gathered data during the session will be stored directly in the University of Edinburgh cloud to keep it confidential. Access to the raw data is restricted to the researcher (Florian Mathis) and his supervisors (Dr. Mohamed Khamis, Dr. Kami Vaniea) only. Your data is fully anonymised and there is no way to trace it back to you. The results of the study may appear in a number of published studies, in a confidential format where anonymity is preserved. Based on your agreement we will use findings for scientific papers and/or presentations at conferences.

Who has reviewed this study?

This study adheres to the BPS ethical guidelines and has been approved by the College of Science and Engineering ethics committee of The University of Glasgow and by the Informatics Forum ethics committee of the University of Edinburgh.

Funding and Contact

This research is supported by the University of Edinburgh and the University of Glasgow jointly funded PhD studentships.

For further information, or if you wish to receive a summary of the findings of this experiment at a later date, please contact the lead researcher Florian Mathis (florian.mathis@glasgow.ac.uk).

Consent Form

Please confirm your participation in this study by completing this consent form:

- I confirm that I have read and understand the information above information sheet, and understand my Data Protection Rights under GDPR for the above study, and know how to contact the researchers to ask questions.
- I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, and am free to not answer any particular question, without providing a reason.
- I understand that all data collected from me will be treated confidentially and anonymized, will be seen in its raw form only by the experimenters, and if published will not be identifiable as coming from me.
- I agree that the anonymised data can be used for follow-up evaluations by researchers in the School of Computing Science and their collaborators. Note that this project is part of a joint PhD between the University of Glasgow and the University of Edinburgh.
- I agree that the researchers are allowed to archive aggregated data and findings (anonymised) taken during the experiment in online repositories such as Enlighten: Research Data: <http://researchdata.gla.ac.uk/>. I am aware of the fact that I can get in touch with the researchers at any time to demand the deletion or retrieval of the recorded data.
- I agree to take part in this survey.

By selecting **“I AGREE TO ALL OF THE ABOVE”** I have read the conditions stated above and agree to take part in the study. This also takes me to the first page of the study.

- I AGREE TO ALL OF THE ABOVE

Qualtrics: Thesis Survey for Section 4.8

Start of Block: Demographics



Q6 Please enter your Prolific ID here:

Start of Block: Study Introduction

Q12

Study Introduction

This study evaluates the security of three novel methods for PIN entry on displays like a game console or an ATM. In this study you will be pretending to be an attacker who is trying to guess the PIN by watching a person in Virtual Reality (VR) enter it.

Today you will be observing the below three types of 4-digit PIN entry. Before each type, we will explain the entry method to you, then show you a sequence of 8 embedded videos where a VR person enters a PIN and then you are asked to guess what the PIN is.

Eye Gaze Mid-Air Touch

Scenario

In the videos, you will be watching a VR person attempting to enter a PIN into a terminal. To make guessing the PIN easier for you, we will be showing you two camera angles that are most likely to help you for that type of entry. For example, we will be showing you a camera angle that shows the eyes and the screen for Eye Gaze and for Touch we show two camera angles of the hands touching the screen. In all videos, you should assume that the VR person is entering the PIN using a similar setup to the picture below.



Attack Videos

There are two common situations where an attacker (you) could observe someone else entering a PIN. The first is where you are present when they enter it and can watch them, but only get to see it entered once. The second is where the entry is video recorded by the attacker or by a nearby camera, such as a surveillance camera.

In this study, we will be showing you two types of videos:

- **single-view**, where you can only view the video once. These videos have a 5-second countdown so you have a moment to prepare. Once the video starts it is not possible to pause or replay the video.
- **repeated-view**, where you can replay or pause the video as often as you like.

For each PIN entry method, we will show you 8 videos alternating between repeated-view and single-view. After each video, you will be asked for your best guess of the PIN. You can also optionally provide two other guesses if you are unsure. Please provide your best guesses even if you are unsure or are only confident about a couple of the observed numbers.

One participant will be selected to win an **additional £7.50** with the odds based on performance. Chances of winning increases with the number of successfully attacked PINs. **Partial correct PINs** also contribute.

Because videos can go by rather fast, we recommend that you have pen and paper handy to

make notes. At the end of the survey, you can optionally upload photos of your notes to help us understand your attack approach. However, this is not mandatory.

End of Block: Study Introduction

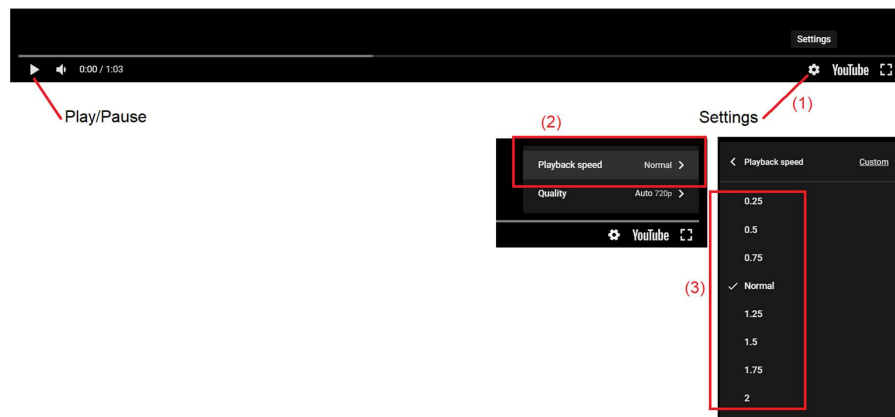
Start of Block: Introduction Youtube Player

Q223

YouTube-Player Introduction

In this study we use embedded Youtube videos with all the same controls as normal YouTube. We disable controls for single-view videos, but for repeated-view videos you are in full control. Below is a short overview of how to **play, pause, replay, speed up, and slow down the videos.**

Through the Youtube player interface (1), you can **change the playback speed** (2) from normal to, for example, 0.5 (3) to slow down the video or to 2 to speed up the video.



Q222 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

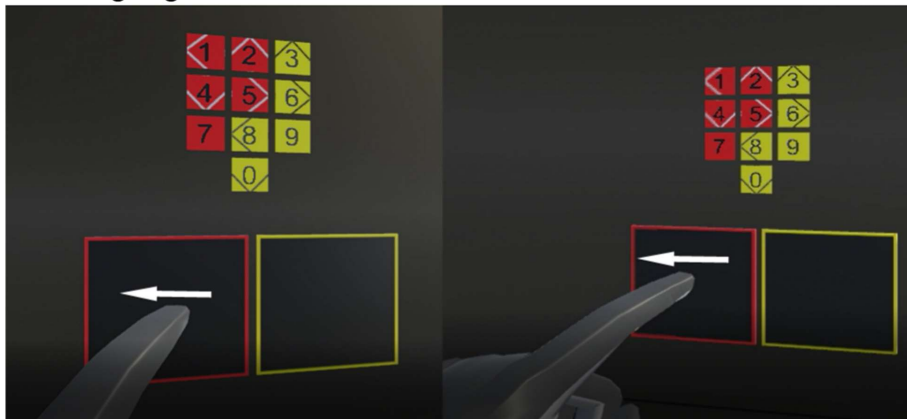
End of Block: Introduction Youtube Player

Start of Block: Introduction Touch

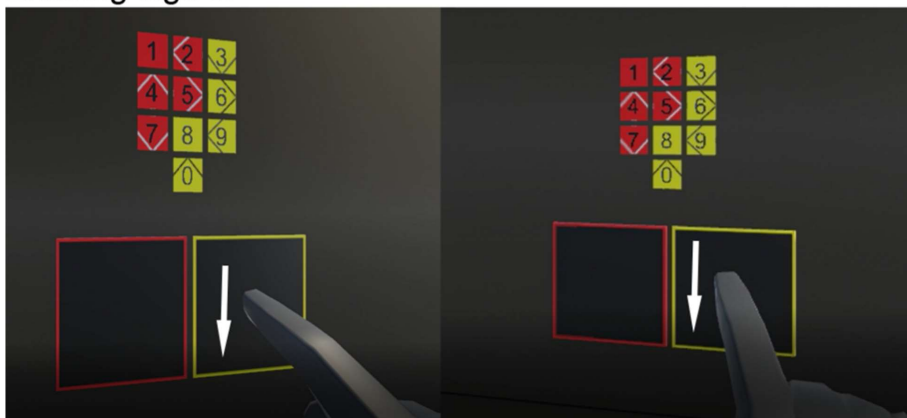
Q16 Touch

In the Touch entry method, a user finds their PIN number on a small PIN pad shown to them on top of the screen and then enters it by making the gesture (up, down, left, right) in the appropriately coloured square at the bottom of the screen. After they enter each number, the PIN pad image changes. So in the screenshots below, the user first uses their left hand to swipe left in the red box to enter a "1" then they use their right hand to swipe down in the yellow box to enter a "3". Note that the screenshot below shows a side-by-side view of the same input from two different angles. The red/yellow colouring of the pin pad is always the same with 1, 2, 4, 5, and 7 always being red and being entered using the left hand. But the swipe directions change after each number is entered.

Entering digit 1:



Entering digit 3:



Possible Gestures

A user can make the following touch gestures:

Arrow to the right -> gesture to the right. **Arrow to the left** -> gesture to the left.

Arrow to the top -> gesture up. **Arrow to the bottom** -> gesture down. **No**

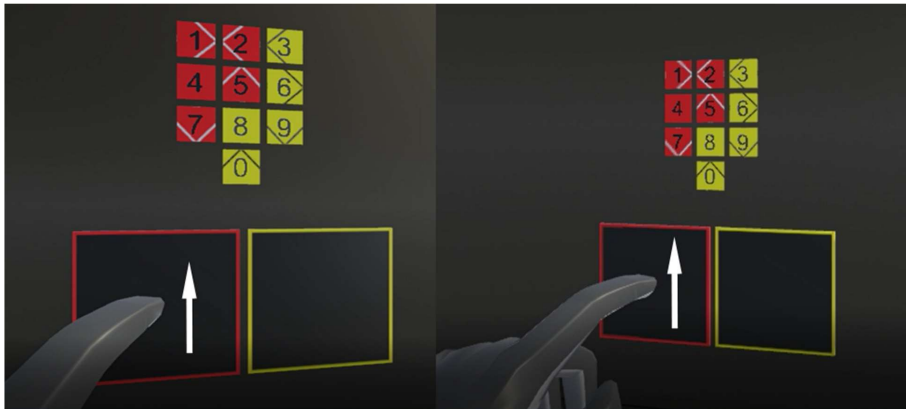
arrow -> a single tap. This is identical to a single touch in the coloured box.

Below you can find an introduction video that explains the input method and your task: attacking/guessing the entered PIN.

JS *

Q17

Please enter the digit that the user enters according to the picture below.



Q642 **Remember:** The survey automatically alternates between single and repeated view videos. Videos in single-view start after a short countdown and are only played once. Repeated-view videos have no time limitations and allow full control over the playback (e.g., pause, play, rewind, change speed).

Please continue only if you fully understood the input described above. We'll start with a repeated-view video. Be ready, the single-view videos are short!

End of Block: Introduction Touch

Start of Block: Touch: Video 8

Q18

Please provide your best guess of the correct PIN in the box(es) below. You are required to provide at least one guess. If you can only observe parts of the PIN in the video then just provide your best guess for the other numbers. All PINs consist of four digits and guesses are required to be this length.



Q19 Guess 1

Q21 Guess 2

Q22 Guess 3

Q23 Please indicate how easy it was to attack this PIN and how confident you are with your guess.

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
Attacking this PIN was easy. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am confident that my guess is correct. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q24 Which method did you use to attack the PIN? Was there anything special about this PIN? For example, particularly easy/difficult? Can you tell why?

Q20 If the video did not load properly or you faced any other issues that did not let you to watch the video above please report this here and say what happened. Note that the answer will not impact your reimbursement.

-
- Q28 Timing
 - First Click (1)
 - Last Click (2)
 - Page Submit (3)
 - Click Count (4)

End of Block: Touch: Video 8

Start of Block: Touch: Video 7 [SINGLE]

Q240

Please provide your best guess of the correct PIN in the box(es) below. You are required to provide at least one guess. If you can only observe parts of the PIN in the video then just provide your best guess for the other numbers. All PINs consist of four digits and guesses are required to be this length.



Q241 Guess 1

Q242 Guess 2

Q243 Guess 3

Q244 Please indicate how easy it was to attack this PIN and how confident you are with your guess.

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
Attacking this PIN was easy. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am confident that my guess is correct. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q245 Which method did you use to attack the PIN? Was there anything special about this PIN? For example, particularly easy/difficult? Can you tell why?

Q246 Please let us know if you watched this video multiple times. If the video did not load properly or you faced any other issues that did not let you to watch the video above please also report this here and say what happened. Note that the answer will not impact your reimbursement.

Q247 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Touch: Video 7 [SINGLE]

Start of Block: Touch: Video 5

Q248

Please provide your best guess of the correct PIN in the box(es) below. You are required to provide at least one guess. If you can only observe parts of the PIN in the video then just provide your best guess for the other numbers. All PINs consist of four digits and guesses are required to be this length.

Q249 Guess 1

Q250 Guess 2

Q251 Guess 3

Q252 Please indicate how easy it was to attack this PIN and how confident you are with your guess.

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
Attacking this PIN was easy. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am confident that my guess is correct. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q253 Which method did you use to attack the PIN? Was there anything special about this PIN? For example, particularly easy/difficult? Can you tell why?

Q254 If the video did not load properly or you faced any other issues that did not let you to watch the video above please report this here and say what happened. Note that the answer will not impact your reimbursement.

Q255 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Touch: Video 5

Start of Block: Touch: Video 1 [SINGLE]

Q256

Please provide your best guess of the correct PIN in the box(es) below. You are required to provide at least one guess. If you can only observe parts of the PIN in the video then just provide your best guess for the other numbers. All PINs consist of four digits and guesses are required to be this length.



Q257 Guess 1

Q258 Guess 2

Q259 Guess 3

Q260 Please indicate how easy it was to attack this PIN and how confident you are with your guess.

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
Attacking this PIN was easy. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am confident that my guess is correct. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q261 Which method did you use to attack the PIN? Was there anything special about this PIN? For example, particularly easy/difficult? Can you tell why?

Q262 Please let us know if you watched this video multiple times. If the video did not load properly or you faced any other issues that did not let you to watch the video above please also report this here and say what happened. Note that the answer will not impact your reimbursement.

Q263 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Touch: Video 1 [SINGLE]

Start of Block: Touch: Video 6

Q264

Please provide your best guess of the correct PIN in the box(es) below. You are required to provide at least one guess. If you can only observe parts of the PIN in the video then just provide your best guess for the other numbers. All PINs consist of four digits and guesses are required to be this length.



Q265 Guess 1

Q266 Guess 2

Q267 Guess 3

Q268 Please indicate how easy it was to attack this PIN and how confident you are with your guess.

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
Attacking this PIN was easy. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am confident that my guess is correct. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Q269 Which method did you use to attack the PIN? Was there anything special about this PIN? For example, particularly easy/difficult? Can you tell why?



Q270 If the video did not load properly or you faced any other issues that did not let you to watch the video above please report this here and say what happened. Note that the answer will not impact your reimbursement.



- Q271 Timing
- First Click (1)
- Last Click (2)
- Page Submit (3)
- Click Count (4)

End of Block: Touch: Video 6

Start of Block: Touch: Video 4 [SINGLE]

Q272

Please provide your best guess of the correct PIN in the box(es) below. You are required to provide at least one guess. If you can only observe parts of the PIN in the video then just provide your best guess for the other numbers. All PINs consist of four digits and guesses are required to be this length.

Q273 Guess 1

Q274 Guess 2

Q275 Guess 3

Q276 Please indicate how easy it was to attack this PIN and how confident you are with your guess.

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
Attacking this PIN was easy. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am confident that my guess is correct. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q277 Which method did you use to attack the PIN? Was there anything special about this PIN? For example, particularly easy/difficult? Can you tell why?

Q278 Please let us know if you watched this video multiple times. If the video did not load properly or you faced any other issues that did not let you to watch the video above please also report this here and say what happened. Note that the answer will not impact your reimbursement.

Q279 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Touch: Video 4 [SINGLE]

Start of Block: Touch: Video 3

Q280

Please provide your best guess of the correct PIN in the box(es) below. You are required to provide at least one guess. If you can only observe parts of the PIN in the video then just provide your best guess for the other numbers. All PINs consist of four digits and guesses are required to be this length.

Q281 Guess 1

Q282 Guess 2

Q283 Guess 3

Q284 Please indicate how easy it was to attack this PIN and how confident you are with your guess.

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
Attacking this PIN was easy. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am confident that my guess is correct. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Q285 Which method did you use to attack the PIN? Was there anything special about this PIN? For example, particularly easy/difficult? Can you tell why?



Q286 If the video did not load properly or you faced any other issues that did not let you to watch the video above please report this here and say what happened. Note that the answer will not impact your reimbursement.



Q287 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Touch: Video 3

Start of Block: Touch: Video 2 [SINGLE]

Q288

Please provide your best guess of the correct PIN in the box(es) below. You are required to provide at least one guess. If you can only observe parts of the PIN in the video then just provide your best guess for the other numbers. All PINs consist of four digits and guesses are required to be this length.

*

Q289 Guess 1

Q290 Guess 2

Q291 Guess 3

Q292 Please indicate how easy it was to attack this PIN and how confident you are with your guess.

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
Attacking this PIN was easy. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am confident that my guess is correct. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q293 Which method did you use to attack the PIN? Was there anything special about this PIN? For example, particularly easy/difficult? Can you tell why?

Q294 Please let us know if you watched this video multiple times. If the video did not load properly or you faced any other issues that did not let you to watch the video above please also report this here and say what happened. Note that the answer will not impact your reimbursement.

Q295 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Touch: Video 2 [SINGLE]

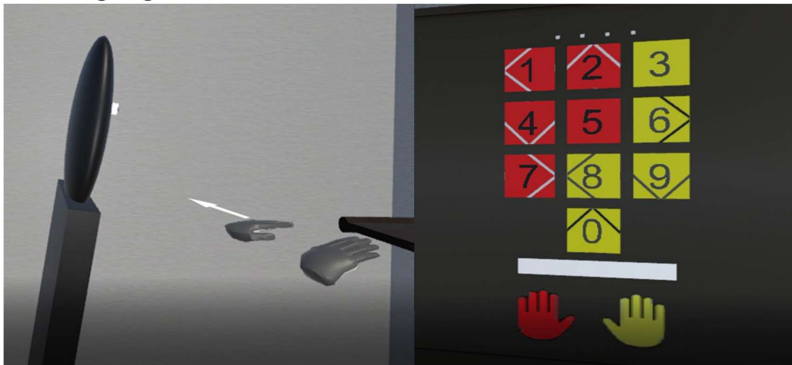
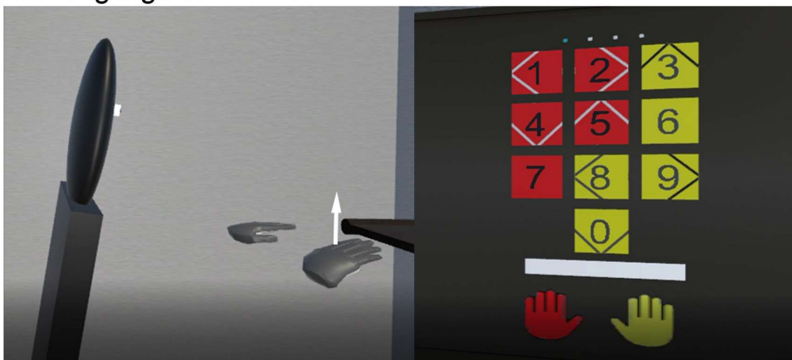
[Repeated for Mid-air and Eye Gaze]

[Below are the Introduction Sections for Mid-Air and Eye Gaze]

Start of Block: Introduction Mid-Air**Q14 Mid-Air**

In the Mid-Air entry method, a user finds their PIN number on a small PIN pad shown to them on top of the screen and then enters it by making gestures (up, down, left, right, front) in the mid-air.

After they enter each number, the PIN pad image changes. So in the screenshots below, the user first uses their left hand to make a mid-air gesture to the left to enter a "1" then they use their right hand to make a mid-air gesture upward to enter a "3". Note that the screenshot below shows a side-by-side view of the same input from two different angles. The red/yellow colouring of the PIN pad is always the same with 1,2,4,5, and 7 always being red and entered using the left hand. But the mid-air directions change after each number entered.

Entering digit 1:**Entering digit 3:**

Possible Gestures

A user can make the following mid-air gestures:

Arrow to the right -> gesture to the right. **Arrow to the left** -> gesture to the left.

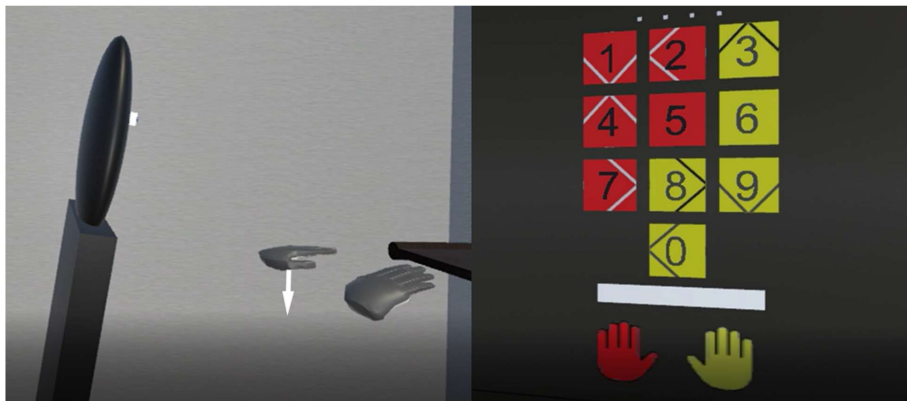
Arrow to the top -> gesture up. **Arrow to the bottom** -> gesture down. **No arrow**: a gesture to the front.

Below you can find an introduction video that explains the input method and your task: attacking/guessing the entered PIN.



Q638

Please enter the digit that the user enters according to the picture below.



Q641 Remember: The survey automatically alternates between single and repeated view videos. Videos in single-view start after a short countdown and are only played once. Repeated-view videos have no time limitations and allow full control over the playback (e.g., pause, play, rewind, change speed).

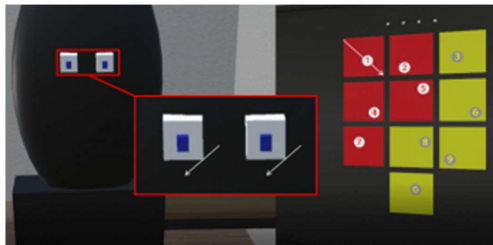
Please continue only if you fully understood the input described above. We'll start with a repeated-view video. Be ready, the single-view videos are short!

End of Block: Introduction Mid-Air

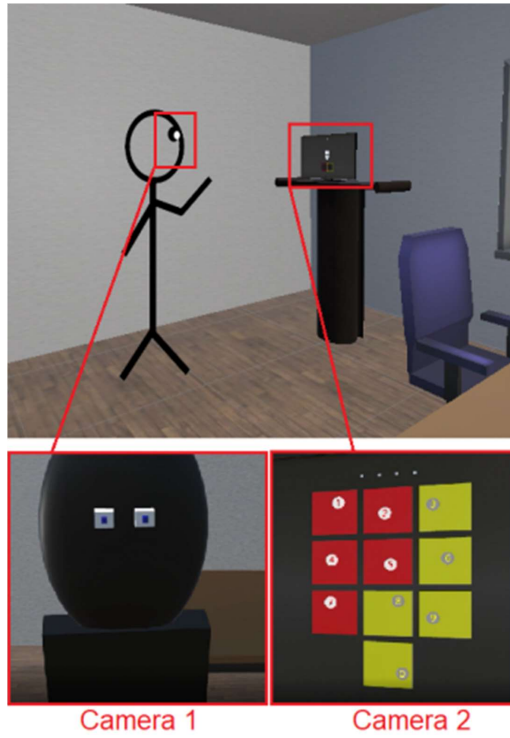
Start of Block: Introduction Gaze

Q12 Eye Gaze

In the eye gaze entry method, a PIN pad is shown to the user where each square on the pad contains a number that is moving around in the square. So for example, the 1 might be moving diagonally from top left to the bottom right. The user enters a number by looking at it and following its motions with their eyes. So in the example of the 1, they would focus their eyes on the moving 1 while it goes through its diagonal motion. The movement directions of the numbers change after each successful number entry. The red/yellow square colours are not used in this entry method and can be ignored.



The video shows views from two angles. The left view shows the VR person from the perspective of the display as if they were shown by a webcam so that you (the attacker) can see their eyes clearly. The right view shows the PIN pad from the perspective of the VR person. **As a result, the images are effectively mirrored. So when entering a 1 that is moving from the top left to the bottom right of the square, the eyes of the VR person will look like they are moving from the top right to the bottom left.**



Possible Movements

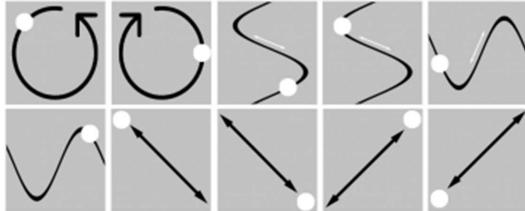
A user can make the following eye gaze movements:

Clockwise and **counter-clockwise circular** eye movements.

Vertical and **horizontal zigzag** eye movements (along both axis and both directions).

Diagonal eye movements (in all four directions).

The screenshot below shows all possible movements of the digits:

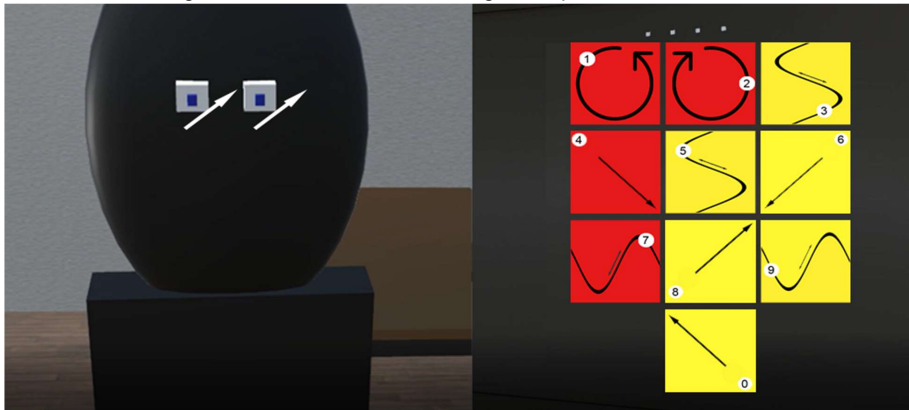


Below is a video showing a VR person entering the PIN "1234" first in slow motion, then at regular speed.



Q639

Please enter the digit that the user enters according to the picture below.



Q640 Remember: The survey automatically alternates between single and repeated view videos. Videos in single-view start after a short countdown and are only played once. Repeated-view videos have no time limitations and allow full control over the playback (e.g., pause, play, rewind, change speed).

Please continue only if you fully understood the input described above. We'll start with a repeated-view video. Be ready, the single-view videos are short!

End of Block: Introduction Gaze

[After observing all input methods and PINs]

Start of Block: General Questions and File Upload

Q643

Thank you for attacking the 24 PINs. Below, we have some additional general questions about the input methods.

Perception of the Security

We would like to know your thoughts about each input method in terms of **security**. Please answer each question below with 2-3 sentences.

Q646 What do you think about the **touch** input method in terms of **security**?

Q647 What do you think about the **mid-air** input method in terms of **security**?

Q644 What do you think about the **eye gaze** input method in terms of **security**?

Q649 Which input method is the **most secure** one in your opinion?

- Touch (1)
- Mid-Air (2)
- Eye Gaze (3)

Q654 Which input method is the **least secure** one in your opinion?

- Touch (1)
- Mid-Air (2)
- Eye Gaze (3)

Q650 **Perception of the Input Methods**

Imagine that you are using these methods (touch, mid-air, eye gaze) to enter your PIN. Please answer each question below with 2-3 sentences.

Q651 How do you feel about using the **touch** input method? What do you think about using this method to enter your PIN?

Q652 How do you feel about using the **mid-air** input method? What do you think about using this method to enter your PIN?

Q653 How do you feel about using the **eye gaze** input method? What do you think about using this method to enter your PIN?

Q629

Did you take any notes during the observation tasks (e.g., on a paper or in a digital form)?

- No, I did not take any notes. (6)
- Yes, I mostly noted down the PIN numbers. (1)
- Yes, I mostly drew figures to help me sort out the PIN. (4)
- Yes, I draw a sketch of the virtual environment. (7)
- Yes, I drew the screen or the person. (5)
- Other (2) _____

Display This Question:

If Did you take any notes during the observation tasks (e.g., on a paper or in a digital form)? = Yes, I mostly noted down the PIN numbers.

Or Did you take any notes during the observation tasks (e.g., on a paper or in a digital form)? = Yes, I mostly drew figures to help me sort out the PIN.

Or Did you take any notes during the observation tasks (e.g., on a paper or in a digital form)? = Yes, I draw a sketch of the virtual environment.

Or Did you take any notes during the observation tasks (e.g., on a paper or in a digital form)? = Yes, I drew the screen or the person.

Or Did you take any notes during the observation tasks (e.g., on a paper or in a digital form)? = Other

Q26

File Upload (optional)

Please feel free to upload a screenshot of any notes you took (e.g., on a paper or in a digital form). **Note that if you upload or send us your notes you will get £0.5 in addition to the basic compensation.**

This is optional and not required, but would help us to better understand your approach when observing the authentications.

Please also note that this decision does not **impact your reimbursement in a negative way**.

If you would like to upload multiple files/pictures please .zip them before uploading or send them to florian.mathis@glasgow.ac.uk.

End of Block: General Questions and File Upload

Start of Block: Feedback

Q29

Feedback, Suggestions, and Problems: Please use the text box below if you faced any issues (e.g., lack of clarity, technical problems) or if you want to give us additional feedback and suggestions for improvement. Please also let us know about any other comments. We highly appreciate your feedback.

Q30 Thank you for participating in this survey! For further information, or if you wish to receive a summary of the findings of this experiment at a later date, please contact the lead researcher Florian Mathis: florian.mathis@glasgow.ac.uk.

Please continue to the next page to return to prolific and confirm that you participated in this study.

End of Block: Feedback

Appendix D

Appendix for Chapter 5

Appendix for the Study in Section 5.3

Participant Information Sheet



School of
Computing Science

VirSec: Exploring the Value of 3D Observations on Authentication Schemes in Virtual Reality

Participant Information Sheet

Researcher: Florian Mathis f.mathis.1@research.gla.ac.uk

Supervisor: Dr. Mohamed Khamis, Mohamed.khamis@glasgow.ac.uk
(2nd) Dr. Kami Vanica, kvanica@inf.ed.ac.uk

IMPORTANT – Exclusion criteria

In order to take part in this study, you must meet the following requirements:

1. Aged 18 or over
2. No history (personal and family) of epileptic seizures, strokes, or photosensitivity
3. Not a member of any of the following groups
 - a. Pregnant women
 - b. The elderly
 - c. Sufferers of any serious medical conditions i.e. you fall into one of the following categories
 - i. Inpatient care
 - ii. Incapacity
 - iii. Chronic serious health conditions
 - iv. Permanent or long term conditions
 - v. Conditions requiring multiple treatments
 - d. Sleep deprived
 - e. Under the influence of alcohol
 - f. Previously suffered concussion or traumatic brain injury
 - g. Prone to dizziness from immersive virtual experiences
 - h. Sufferers of panic attacks or generalised anxiety disorders which might be provoked by wearing headphones / being unable to hear your surroundings
 - i. Prone to issues with balance or motor function (i.e. you can walk around a room over the course of an hour).
4. Be comfortable with wearing a head-mounted display (HMD) such as a HTC VIVE or Oculus Quest and be comfortable with filling in a survey and watching short videos where a user performs PIN entries in a virtual environment.

1. Invitation

You are being invited to take voluntarily part in a research experiment. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

Thank you for reading this.

2. Purpose of the User Study

We would like to reach out to you to participate in a paid study examining the benefits of 3D recordings to assess a system's resistance to observations. Your participation is voluntary, and you are free to withdraw at any time, without giving any reason, and you are free to omit answering any particular question, without providing a reason. The analysis of this experiment will be published at top-tier venues such as CHI, the

premier international conference of Human-Computer Interaction, SOUPS, Symposium on Usable Privacy and Security, and IEEE VR. All publications are fully anonymised and findings and specific measurements cannot be traced back to you.

3. What will happen to me if I take part?

You will be fitted with a virtual reality headset such that it sits comfortably on your head, and you can hear correctly. This will occur in a room at the School of Computing Science (Sir Alwyn Williams Building or Lilybank Gardens) or in the office of the lead researcher (Austria). You are then going to observe different PIN entries directly in VR. We will demonstrate how such an observation attack could look like and guide you in performing these attacks during a training session. During all observations we will capture following data:

- Virtual Reality:
 - We are going to record the graphical representation of the entire virtual environment. This includes static objects within the environment (e.g., the authentication scheme). For all recordings we will use appropriate file extensions (e.g., .csv, .txt) and store them in separate files on our local machine and then upload it anonymised (through participant ID's) to the University of Glasgow cloud.
- Real World:
 - We are going to record you from the real world during your tasks within the virtual world. This includes the recording of a video from different perspectives and/or taking photographs. You can see an example of this in the picture below. The picture shows a researcher (Florian Mathis) performing a task in virtual reality. Florian is equipped with the HTC VIVE and two HTC VIVE controllers. The way we are going to record your interactions is similar to the picture below.



You are also going to watch authentications that were recorded in a virtual environment on a desktop computer. This allows you to perform observation attacks on authentications. At the end of the study, there will be some additional questions within the context of the experiment. This will be in the form of semi-structured interviews and helps the research team to better understand the experience you have undertaken. At the end of each ~1h session we will hand out the £8/10€ per hour in cash.

4. Why have I been chosen?

Your participation has been solicited through emails, social media postings, word-of-mouth, notice board postings, or a call on Prolific to which you replied. Your participation is voluntary, and you are free to withdraw at any time, without giving any reason, and you are free to omit answering any particular question, without providing a reason.

5. Conditions and Data Storage

Each experiment will last for approximately 1 hour. All gathered data during the session will be stored directly in the University of Glasgow cloud to keep it confidential (<https://gla-my.sharepoint.com>). Access to the raw data is restricted to the researcher (Florian Mathis) and his supervisors (Dr. Mohamed Khamis, Dr. Kami Vaniea) only. Your data is fully anonymised and there is no way to trace it back to you. The results of the study may appear in a number of published studies, in a confidential format where anonymity is preserved. Based on your agreement we will use the data (e.g., screenshots, drawings) as video and image material for scientific papers and/or presentations at conferences.

6. Data Usage

The data will be used within our research and is part of Florian Mathis' Research PhD. We will store the raw data in the University of Glasgow cloud (<https://gla-my.sharepoint.com>). Access to the raw data is restricted to the researcher (Florian Mathis) and his supervisors (Dr. Mohamed Khamis, Dr. Kami Vaniea) only. Based on the request of participants the data can be destroyed at any point. The data will be kept until beyond the end of the Research PhD (up to 10 years) and findings of the experiment might be re-used for additional research projects within Florian's Research PhD.

7. Who has reviewed the study?

This study adheres to the BPS ethical guidelines, and has been approved by the College of Science and Engineering ethics committee of The University of Glasgow. The approved study was also submitted to the University of Edinburgh ethics committee as the lead researcher Florian Mathis is part of both universities.

8. Funding and Contact

This research is supported by the University of Edinburgh and the University of Glasgow jointly funded PhD studentships: <https://www.gla.ac.uk/research/ourresearchenvironment/prs/uofguofedinphdstudentships/>

The project has been reviewed and approved by the Research Ethics Committee in the School of Computing Science at the University of Glasgow (number protocol: tba). For further information please feel free to get in touch with the researcher f.mathis.1@research.gla.ac.uk (or via phone: [REDACTED]).

Whilst you are free to discuss your participation in this study with the researcher, if you would like to speak to someone not involved in the study, you may contact the Ethics Committee at Christoph.Scheepers@glasgow.ac.uk.

For further information, or if you wish to receive a summary of the findings of this experiment at a later date, please contact the researcher or the supervisor of this project, details listed below.

Data Protection and Confidentiality

Your data will be processed in accordance with the Data Protection Act 1998 (up until 24th May 2018) and the General Data Protection Regulation 2016 (GDPR) thereafter. All information collected about you will be kept strictly confidential. Unless they are anonymised in our records, your data will be referred to by a unique participant number rather than by name. If you consent to being audio recorded, all recordings will be destroyed once they have been transcribed. Your data will only be viewed by the researcher/research team. All electronic data will be stored on a password-protected computer file within the School of Computing Science. All paper records will be stored in a locked filing cabinet within the School of Computing Science. Your consent information will be kept separately from your responses in order to minimise risk in the event of a data breach.

Data Protection Rights

University of Glasgow is a Data Controller for the information you provide. You have the right to access information held about you. Your right of access can be exercised in accordance with the Data Protection Act 1998 (up until 24th May 2018) and the General Data Protection Regulation thereafter. You also have other rights including rights of correction, erasure, objection, and data portability. For more details, including the right to lodge a complaint with the Information Commissioner's Office, please visit www.ico.org.uk. Questions, comments and requests about your personal data can also be sent to the University Data Protection Officer - dp@ gla.ac.uk (<https://www.gla.ac.uk/myglasgow/dpfooffice/contact/>)

Thank you for volunteering to take part in this study. If there are any questions or issues, or if you wish to receive a summary of the findings of this experiment at a later date, please feel free to get in touch with the researcher at any time.

Researcher

Florian Mathis

Email: f.mathis.1@research.gla.ac.uk

Tel: [REDACTED]

1st Supervisor (University of Glasgow)

Dr. Mohamed Khamis

Email: Mohamed.Khamis@glasgow.ac.uk

Tel: [REDACTED]

Consent Form



CONSENT FORM

Title of Experiment: Exploring the Value of 3D Observations on Authentication Schemes in Virtual Reality

Experimenter details: Florian Mathis (f.mathis.1@research.gla.ac.uk, [REDACTED])

Supervisor details: Dr. Mohamed Khamis (Mohamed.khamis@glasgow.ac.uk)
Dr. Kami Vaniea (kvaniea@inf.ed.ac.uk)

Before agreeing to this consent form, you should have been given an information sheet to read, which outlines exclusion criteria and explains the general purpose of this experiment and the tasks it involves. If you did not receive this, please inform the researcher (Florian Mathis, f.mathis.1@research.gla.ac.uk). Please tick the box after each statement to indicate that you have read and understand the statement, and that you agree with it.

1. I confirm that I have read and understand the Participant Information sheet, and understand my Data Protection Rights under GPDR for the above study, and have had the opportunity to ask questions.	
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, and am free to omit answering any particular question, without providing a reason.	
3. I give consent for my actions to be recorded (audio and video) during the study.	
4. I give consent for my actions to be recorded (user inputs in specific text-fields) during the study.	
5. I understand that all data collected from me will be treated confidentially and anonymized, will be seen in its raw form only by the experimenters, and if published will not be identifiable as coming from me.	
6. I agree that the researchers can use video recordings for public outreach, for instance, showing parts of the recordings at conference venues and/or use the material in videos and publications to showcase the system.	
7. By agreeing to take part in this study I also agree that recordings and data can be used for follow-up evaluations by researchers in the school of computing science and their collaborators.	
8. I agree that the researchers are allowed to archive all data taken during the experiment (e.g., video recordings with a camera in the real world; time spent on a page/provided data in the survey) in online repositories such as Enlighten: Research Data: http://researchdata.gla.ac.uk/ . I am aware of the fact that I can get in touch with the researchers at any time to demand the deletion or retrieval of these recordings.	
9. I agree to take part in the above study (Exploring the Value of 3D Observations on Authentication Schemes in Virtual Reality).	

This study has been approved by the Ethics Committee.

By signing this form, you have read the conditions stated above and agree to take part in the study.

FULL NAME: _____

SIGNATURE: _____

DATE and PLACE: _____

Qualtrics: Thesis Survey



THE UNIVERSITY *of* EDINBURGH
informatics

3D Virtual Reality Observations

Start of Block: Study ID Only

Q4 Please enter your study ID here:

End of Block: Study ID Only

Start of Block: Introduction to Tasks

Q9

Study Introduction

This study evaluates the resistance to observations of PIN and pattern entries. The authentications are performed on (1) an ATM in a public space and (2) a smartphone at a bus station. You will be pretending to be an attacker who is trying to guess the PIN/pattern by watching a person in virtual reality (VR) enter it. Note that each PIN or pattern has a length of 4 (e.g., "1234"). An example of a PIN entry on an ATM ("1234") and a pattern entry on a smartphone ("1234") is depicted below. Note that you will see example videos at a later stage.

PIN entry on ATM ("1234")



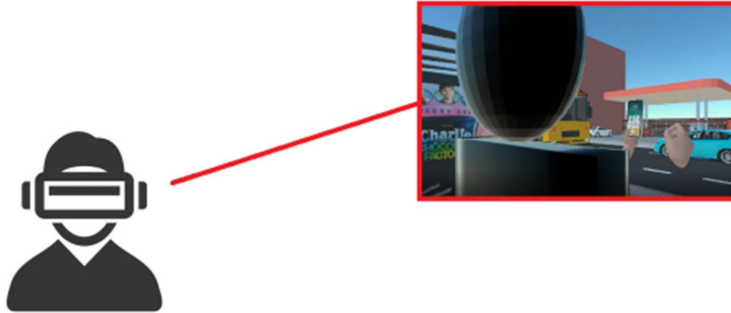
Pattern entry on Smartphone ("1234")



You will perform observations on PIN and Pattern entries with three different types of observations. Before each type, we will explain the observation technique to you, then show you a sequence of 4 recordings where a VR person enters a PIN (or pattern). You are then asked to guess what the PIN/Pattern is.

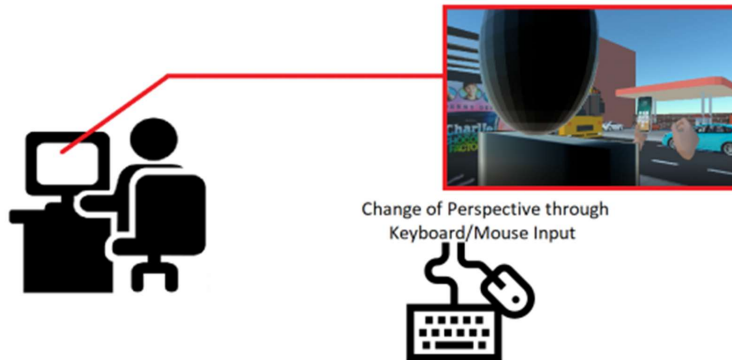
The three observation techniques are:

Immersive Virtual Reality Observation



You will be fitted with a virtual reality headset such that it sits comfortably on your head. You can freely move around to explore the virtual environment. Please note that the experimenter will observe you during this task to avoid any uncertain situations such as bumping into physical obstacles in the real world that you are not aware of.

3D Screen Observation



Here, you will have full control of the 3D environment on a computer screen. You can navigate within the virtual environment using the keyboard keys ``W,A,S,D``. You can also rotate the view using a right mouse-click and moving the mouse.

Static 2D Screen Observation



In this observation technique, you are presented with a recording of the authentication on a computer screen that you cannot manipulate. This means you cannot change your observation angle and are asked to observe the authentication the way it has been recorded.

Note that we will run introduction observations where you are going to watch a person in virtual reality enter "1234" in advance of each observation type.

Scenario

In the three conditions above, you will be watching a VR person attempting to enter a PIN on a) a smartphone or b) on an ATM. In all videos, you should assume that the VR person is entering the PIN/Pattern using a similar setup to the pictures below.

Smartphone Authentication at a Bus Station



ATM Authentication in a Public Space



Attack Videos

There are two common situations where an attacker (you) could observe someone else entering a PIN. The first is where you are present when they enter it and can watch them, but only get to see it entered once. The second is where the entry is video recorded by the attacker or by a nearby camera, such as a surveillance camera.

In addition to the three different observation techniques, we will be showing you two types of authentication recordings:

- **single-view**, where you can only view the authentication once. Once the video starts it is not possible to replay the authentication recording.
- **repeated-view**, where you can replay the video as often as you like.

For each observation technique, we will show you authentications alternating between repeated-view and single-view observations. After each observation, you will be asked for your best guess of the PIN. You can also optionally provide two additional guesses if you are unsure. Please provide your best guess even if you are unsure or are only confident about a couple of the observed numbers.

One participant will be selected to win an **additional €15** with the odds based on performance. Chances of winning increases with the number of successfully attacked PINs. **Partial correct PINs** also contribute.

End of Block: Introduction to Tasks

Start of Block: Observation Type: Static 2D Video Observation

Q13 In the **Static 2D Video Observation** method, you are presented with a recording of the authentication that you cannot manipulate. This means you cannot change your observation angle and are asked to observe the authentication the way it has been recorded. We will then play the recording of the authentication that you are supposed to observe. At the end of the observation you are required to provide your best guess and your level of confidence.

We now start with the training phase.

We will introduce you to the virtual environment and run an example observation on a person in VR entering "1234" (PIN), "1234" (pattern) on a smartphone, and "1234" (PIN) on an ATM.

End of Block: Observation Type: Static 2D Video Observation

Start of Block: INPUT_1_repeated

Q73 Please provide your best guess of the correct PIN in the box(es) below. You are required to provide at least one guess. If you can only observe parts of the PIN in the video then just provide your best guess for the other numbers. All **PINs consist of four digits** and guesses are required to be this length.

Q74 Guess 1:

Q75 Guess 2:

Q76 Guess 3:

Q77 Please indicate how confident you are with your guess and how easy it was to observe the authentication.

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
I am confident that my guess is correct. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It was easy to observe the authentication. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q78 Did you observe the authentication more than once?

- Yes, I watched the authentication more than once. (1)
- No, I only watched the authentication once. (2)

Q80 Was there anything special about this authentication? For example, particularly easy/difficult? Can you tell why?

Q81 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: INPUT_1_repeated

Start of Block: Next Authentication

Q447 We now continue with the next PIN/pattern entry. Please watch the authentication before proceeding to the next page.

End of Block: Next Authentication

Start of Block: INPUT_2_single

Q82 Please provide your best guess of the correct PIN in the box(es) below. You are required to provide at least one guess. If you can only observe parts of the PIN in the video then just provide your best guess for the other numbers. All **PINs consist of four digits** and guesses are required to be this length.

*

Q83 Guess 1:

*

Q84 Guess 2:

*

Q85 Guess 3:

Q86

Please indicate how confident you are with your guess, how easy it was to observe the authentication, and if the situation depicted a scenario that could occur in the real world.

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
I am confident that my guess is correct. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It was easy to observe the authentication. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q87 Was there anything special about this authentication? For example, particularly easy/difficult? Can you tell why?

Q88 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: INPUT_2_single

Start of Block: INPUT_3_repeated

Q89 Please provide your best guess of the correct PIN in the box(es) below. You are required to provide at least one guess. If you can only observe parts of the PIN in the video then just provide your best guess for the other numbers. All **PINs consist of four digits** and guesses are required to be this length.

*

Q90 Guess 1:

*

Q91 Guess 2:

*

Q92 Guess 3:

Q93 Please indicate how confident you are with your guess and how easy it was to observe the authentication.

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
I am confident that my guess is correct. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It was easy to observe the authentication. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q94 Did you observe the authentication more than once?

- Yes, I watched the authentication more than once. (1)
- No, I only watched the authentication once. (2)

Q96 Was there anything special about this authentication? For example, particularly easy/difficult? Can you tell why?

Q97 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: INPUT_3_repeated

Start of Block: INPUT_4_single

Q98 Please provide your best guess of the correct PIN in the box(es) below. You are required to provide at least one guess. If you can only observe parts of the PIN in the video then just provide your best guess for the other numbers. All **PINs consist of four digits** and guesses are required to be this length.

*

Q99 Guess 1:

*

Q100 Guess 2:

*

Q101 Guess 3:

Q102

Please indicate how confident you are with your guess, how easy it was to observe the authentication, and if the situation depicted a scenario that could occur in the real world.

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
I am confident that my guess is correct. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It was easy to observe the authentication. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q103 Was there anything special about this authentication? For example, particularly easy/difficult? Can you tell why?

- Q104 Timing
- First Click (1)
- Last Click (2)
- Page Submit (3)
- Click Count (4)

End of Block: INPUT_4_single

Start of Block: INPUT_5_repeated

Q105 Please provide your best guess of the correct PIN in the box(es) below. You are required to provide at least one guess. If you can only observe parts of the PIN in the video then just provide your best guess for the other numbers. All **PINs consist of four digits** and guesses are required to be this length.



Q106 Guess 1:



Q107 Guess 2:



Q108 Guess 3:

Q109 Please indicate how confident you are with your guess and how easy it was to observe the authentication.

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
I am confident that my guess is correct. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It was easy to observe the authentication. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q110 Did you observe the authentication more than once?

- Yes, I watched the authentication more than once. (1)
 - No, I only watched the authentication once. (2)
-

Q112 Was there anything special about this authentication? For example, particularly easy/difficult? Can you tell why?

Q113 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: INPUT_5_repeated

Start of Block: INPUT_6_single

Q114 Please provide your best guess of the correct PIN in the box(es) below. You are required to provide at least one guess. If you can only observe parts of the PIN in the video then just provide your best guess for the other numbers. All **PINs consist of four digits** and guesses are required to be this length.



Q115 Guess 1:



Q116 Guess 2:



Q117 Guess 3:

Q118

Please indicate how confident you are with your guess, how easy it was to observe the authentication, and if the situation depicted a scenario that could occur in the real world.

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
I am confident that my guess is correct. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It was easy to observe the authentication. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q119 Was there anything special about this authentication? For example, particularly easy/difficult? Can you tell why?

Q120 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: INPUT_6_single

Start of Block: INPUT_7_repeated

Q121 Please provide your best guess of the correct PIN in the box(es) below. You are required to provide at least one guess. If you can only observe parts of the PIN in the video then just provide your best guess for the other numbers. All **PINs consist of four digits** and guesses are required to be this length.

*

Q122 Guess 1:

*

Q123 Guess 2:



Q124 Guess 3:

Q125 Please indicate how confident you are with your guess and how easy it was to observe the authentication.

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
I am confident that my guess is correct. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It was easy to observe the authentication. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q126 Did you observe the authentication more than once?

- Yes, I watched the authentication more than once. (1)
- No, I only watched the authentication once. (2)

Q128 Was there anything special about this authentication? For example, particularly easy/difficult? Can you tell why?

Q129 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: INPUT_7_repeated

Start of Block: INPUT_8_single

Q130 Please provide your best guess of the correct PIN in the box(es) below. You are required to provide at least one guess. If you can only observe parts of the PIN in the video then just provide your best guess for the other numbers. All **PINs consist of four digits** and guesses are required to be this length.



Q131 Guess 1:



Q132 Guess 2:



Q133 Guess 3:

Q134

Please indicate how confident you are with your guess, how easy it was to observe the authentication, and if the situation depicted a scenario that could occur in the real world.

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
I am confident that my guess is correct. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It was easy to observe the authentication. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Q135 Was there anything special about this authentication? For example, particularly easy/difficult? Can you tell why?



- Q136 Timing
- First Click (1)
- Last Click (2)
- Page Submit (3)
- Click Count (4)

End of Block: INPUT_8_single

Start of Block: INPUT_9_repeated

Q137 Please provide your best guess of the correct PIN in the box(es) below. You are required to provide at least one guess. If you can only observe parts of the PIN in the video then just provide your best guess for the other numbers. All **PINs consist of four digits** and guesses are required to be this length.



Q138 Guess 1:



Q139 Guess 2:



Q140 Guess 3:

Q141 Please indicate how confident you are with your guess and how easy it was to observe the authentication.

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
I am confident that my guess is correct. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It was easy to observe the authentication. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q142 Did you observe the authentication more than once?

- Yes, I watched the authentication more than once. (1)
- No, I only watched the authentication once. (2)

Q144 Was there anything special about this authentication? For example, particularly easy/difficult? Can you tell why?

Q145 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: INPUT_9_repeated

Start of Block: INPUT_10_single

Q146 Please provide your best guess of the correct PIN in the box(es) below. You are required to provide at least one guess. If you can only observe parts of the PIN in the video then just provide your best guess for the other numbers. All **PINs consist of four digits** and guesses are required to be this length.

*

Q147 Guess 1:

*

Q148 Guess 2:

*

Q149 Guess 3:

Q150

Please indicate how confident you are with your guess, how easy it was to observe the authentication, and if the situation depicted a scenario that could occur in the real world.

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
I am confident that my guess is correct. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It was easy to observe the authentication. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q151 Was there anything special about this authentication? For example, particularly easy/difficult? Can you tell why?

Q152 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: INPUT_10_single

Start of Block: INPUT_11_repeated

Q153 Please provide your best guess of the correct PIN in the box(es) below. You are required to provide at least one guess. If you can only observe parts of the PIN in the video then just provide your best guess for the other numbers. All **PINs consist of four digits** and guesses are required to be this length.



Q154 Guess 1:



Q155 Guess 2:



Q156 Guess 3:

Q157 Please indicate how confident you are with your guess and how easy it was to observe the authentication.

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
I am confident that my guess is correct. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It was easy to observe the authentication. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q158 Did you observe the authentication more than once?

- Yes, I watched the authentication more than once. (1)
 - No, I only watched the authentication once. (2)
-

Q160 Was there anything special about this authentication? For example, particularly easy/difficult? Can you tell why?

Q161 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: INPUT_11_repeated

Start of Block: INPUT_12_single

Q270 Please provide your best guess of the correct PIN in the box(es) below. You are required to provide at least one guess. If you can only observe parts of the PIN in the video then just provide your best guess for the other numbers. All **PINs consist of four digits** and guesses are required to be this length.



Q271 Guess 1:



Q272 Guess 2:



Q273 Guess 3:

Q274

Please indicate how confident you are with your guess, how easy it was to observe the authentication, and if the situation depicted a scenario that could occur in the real world.

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
I am confident that my guess is correct. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It was easy to observe the authentication. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q275 Was there anything special about this authentication? For example, particularly easy/difficult? Can you tell why?

Four horizontal lines for text input.

Q276 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: INPUT_12_single

Start of Block: NASA-TLX [first method]

Q361
Please rate the following scales based on the observation method you have just experienced. The scales are part of the **NASA Task Load Index (NASA-TLX)**: A widely used, subjective, multidimensional assessment tool that rates perceived workload in order to assess a task, system, or team's effectiveness or other aspects of performance.

Q368 **NASA Task Load Index (NASA-TLX)**

How mentally demanding was the task?

Very Low Very High

0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100



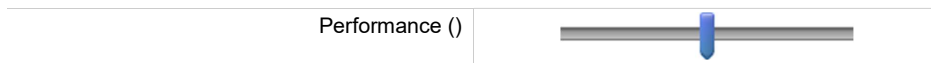
Q369
How physically demanding was the task?
Very Low Very High
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100



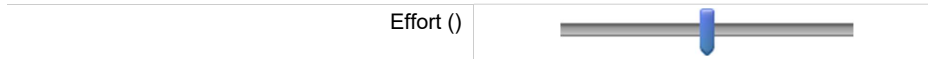
Q370
How hurried or rushed was the pace of the task?
Very Low Very High
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100



Q371
How successful were you in accomplishing what you were asked to do?
Perfect Failure
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100



Q372
How hard did you have to work to accomplish your level of performance?
Very Low Very high
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100



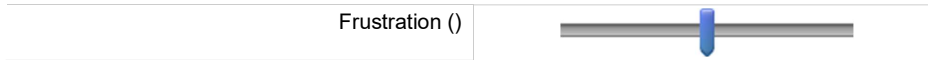
Q373

How insecure, discouraged, imitated, stressed, and annoyed were you?

Very Low

Very High

0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100



Q367 Timing

First Click (1)

Last Click (2)

Page Submit (3)

Click Count (4)

End of Block: NASA-TLX [first method]

Start of Block: IPQ [first method]

Q375

Please rate the following scales based on the observation method you have just experienced. The scales are part of the **Igroup Presence Questionnaire (IPQ)**: A scale for measuring the sense of presence experienced in a virtual environment (VE). We understand the *sense of presence* as the *subjective* sense of being in a virtual environment.

Q383 Igroup Presence Questionnaire (IPQ)

	Not at All (1)	(2)	(3)	(4)	(5)	(6)	Very Much (7)
In the computer generated world I had a sense of "being there". (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q385

	Fully Disagree (1)	(2)	(3)	(4)	(5)	(6)	Fully Agree (7)
Somehow I felt that the virtual world surrounded me. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q386

	Fully Disagree (1)	(2)	(3)	(4)	(5)	(6)	Fully Agree (7)
I felt like I was just perceiving pictures. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q387

	Did not Feel (1)	(2)	(3)	(4)	(5)	(6)	Felt Present (7)
I did not feel present in the virtual space. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q388

	Fully Disagree (1)	(2)	(3)	(4)	(5)	(6)	Fully Agree (7)
I had a sense of acting in the virtual space, rather than operating something from outside. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q389

	Fully Disagree (1)	(2)	(3)	(4)	(5)	(6)	Fully Agree (7)
I felt present in the virtual space. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q390

	Extremely Aware (1)	(2)	(3)	Moderately Aware (4)	(5)	(6)	Not Aware at All (7)
How aware were you of the real world surrounding while navigating in the virtual world? (i.e. sounds, room temperature, other people, etc.)? (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q392

	Fully Disagree (1)	(2)	(3)	(4)	(5)	(6)	Fully Agree (7)
I was not aware of my real environment. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q393

	Fully Disagree (1)	(2)	(3)	(4)	(5)	(6)	Fully Agree (7)
I still paid attention to the real environment. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q394

	Fully Disagree (1)	(2)	(3)	(4)	(5)	(6)	Fully Agree (7)
I was completely captivated by the virtual world. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q395

	Completely Real (1)	(2)	(3)	(4)	(5)	(6)	Not Real at All (7)
How real did the virtual world seem to you? (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q396

	Not Consistent (1)	(2)	(3)	Moderately Consistent (4)	(5)	(6)	Very Consistent (7)
How much did your experience in the virtual environment seem consistent with your real world experience? (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q397

	About as real as an imagined world (1)	(2)	(3)	(4)	(5)	(6)	Indistinguishable from the real world (7)
How real did the virtual world seem to you? (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q398

	Fully Disagree (1)	(2)	(3)	(4)	(5)	(6)	Fully Agree (7)
The virtual world seemed more realistic than the real world. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q382 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: IPQ [first method]

[Repeated for 3D On-Screen Observation and Immersive Virtual Reality Observation]

Start of Block: Study: End [A few Questions on Scales/Items + Interview]

Q50 Real-world Experience

Please answer following questions about your **experience of observations on PIN/Pattern entries in the real world** when considering all three roles: being the victim, the attacker, or a bystander who observes another person performing such an attack.

Q51 Have you experienced the scenario where someone (you or a third person) authenticated on an ATM and was observed by an additional person (you or a third person)? Please tick all situations (i.e. as victim, attacker, bystander) you experienced in the past.

- Yes, as the victim. (1)
 - Yes, as additional bystander. (2)
 - Yes, as the attacker. (3)
 - No, I have never experienced such a situation in the real world. (4)
-

Q69 Have you experienced the scenario where someone (you or a third person) authenticated on a smartphone and was observed by an additional person (you or a third person)? Please tick all situations (i.e. as victim, attacker, bystander) you experienced in the past.

- Yes, as the victim. (1)
 - Yes, as additional bystander. (2)
 - Yes, as the attacker. (3)
 - No, I have never experienced such a situation in the real world. (4)
-

Q70 Please consider the two settings: Entering a PIN on an ATM and entering a PIN/Pattern on a smartphone while waiting at the bus station.

Please rate to what extent you agree/disagree with the following two statements:

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
The ATM situation studied in this study could occur in the real world in a similar way. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The bus station situation studied in this study could occur in the real world in a similar way. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Q54

Perception of the Observation Types

Which observation experience did you perceive as most similar to **observations in the real world**?

_____ **Immersive Virtual Reality Observation** (You have been fitted with a virtual reality headset and could position yourself in the virtual environment.) (1)

_____ **3D On-Screen Observation** (You had full control of the 3D environment on a computer screen and could navigate within the virtual environment using keyboard and mouse input.) (2)

_____ **Static 2D Observation** (You were presented with video recordings of authentications that you could not manipulate.) (3)

End of Block: Study: End [A few Questions on Scales/Items + Interview]

Semi-structured Interviews

The semi-structured interviews at the end of the study were loosely guided by the following questions. Questions were asked for all three observation methods: for 2D Video Observations (i.e., 2DVO), 3D Observations (i.e., 3DVO), and VR Observations (i.e., VRO).

1. Which observation method did you perceive as most similar to observations in the real world? Why?
2. How did you feel about using *[observation method]*? What was easy and/or challenging?
3. In which observation method do you think have you been most successful in terms of correctly guessed PINs/patterns? Please explain why.
4. Could you please tell us why (or why not) you felt being part of the environment where the authentication happened?
5. Could you please tell us why (or why not) you had a sense of acting in the virtual space rather than operating something from the outside?
6. Could you please tell us how the real-world surrounding impacted you while performing the observation task?
7. Please consider the experienced environment and a real-world environment where you are standing next to a real person and perform the same observation task. What would be different to what you have just experienced in our study?
8. In which authentication context was it easier for you to perform the observation? Please explain why.
9. Did you change your observation strategy when attacking PINs vs patterns on the smartphone? Please explain why.
10. Did you experience any difficulties when attacking PINs compared to patterns (or vice versa)?
11. Did you change your observation strategy between the smartphone and the ATM? Please explain why.

Participants' Observation Positions in Smartphone PIN/Pattern

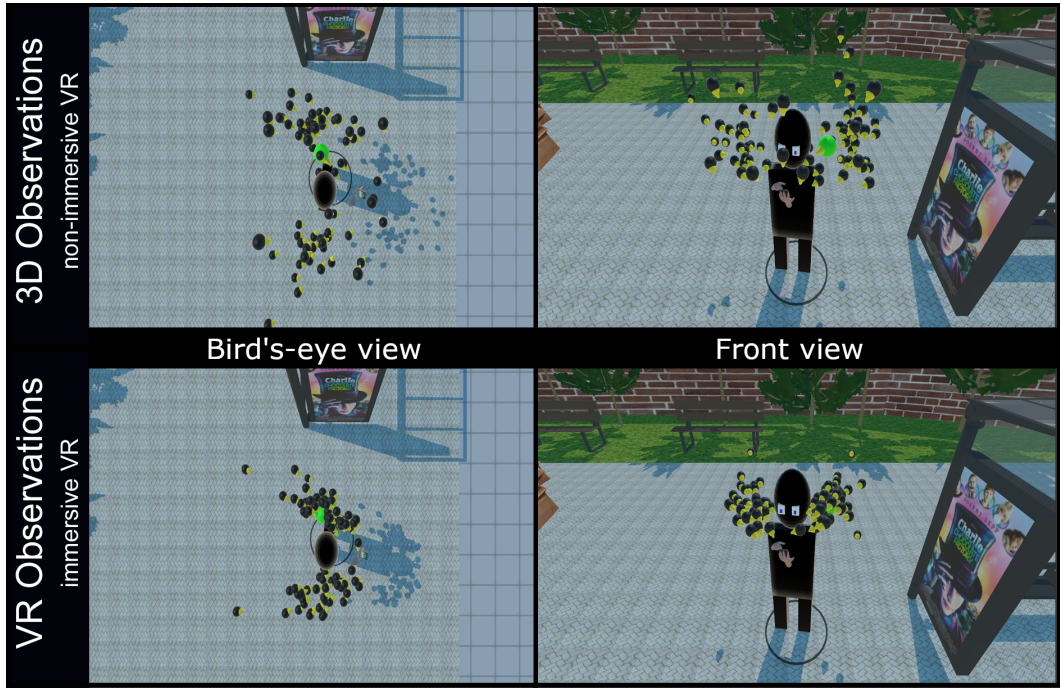


Figure D.1: Participants' self-selected positions when observing smartphone PIN authentications. The green observation position represents the expert-defined observation position in 2DVO as a reference.

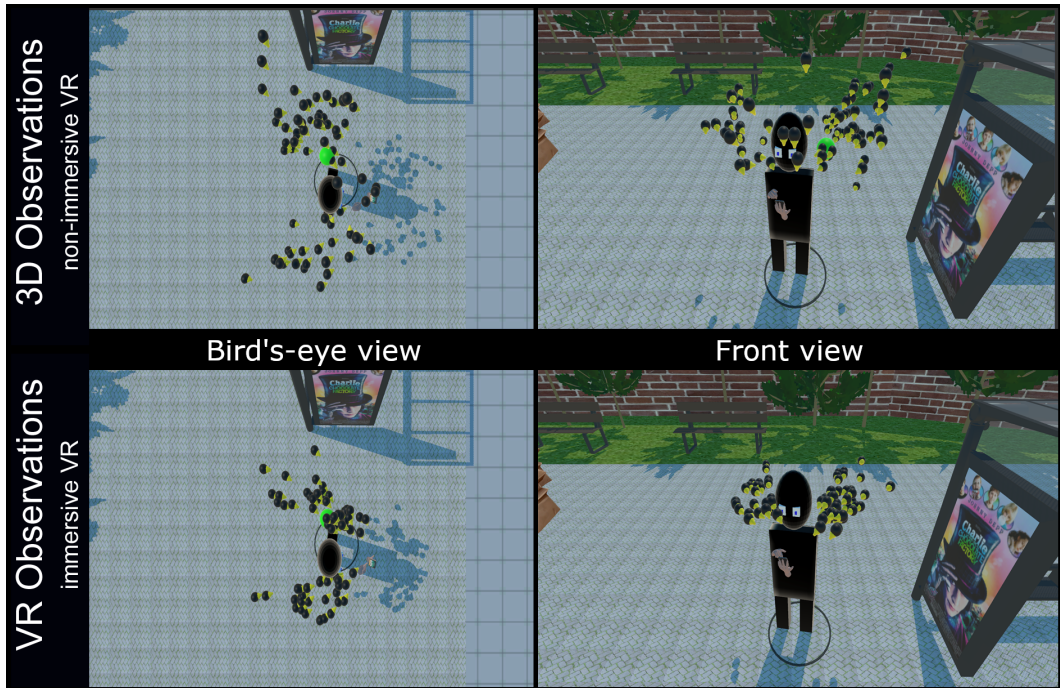


Figure D.2: Participants' self-selected positions when observing smartphone pattern authentications. The green observation position represents the expert-defined observation position in 2DVO as a reference.

Appendix for the Study in Section 5.4

Participant Information Sheet



School of
Computing Science

VirSec: In-Situ Evaluation of Authentication Schemes Using Virtual Reality

Please note for participants we slightly modified the title because of the blinded research approach.

Participant Information Sheet

Researcher: Florian Mathis f.mathis.1@research.gla.ac.uk

Supervisor: Dr. Mohamed Khamis, Mohamed.khamis@glasgow.ac.uk

(2nd) Dr. Kami Vanica, kvanica@inf.ed.ac.uk

IMPORTANT – Exclusion criteria

In order to take part in this study, you must meet the following requirements:

1. Aged 16 or over
2. No history (personal and family) of epileptic seizures, strokes, or photosensitivity
3. Not a member of any of the following groups
 - a. Pregnant women
 - b. The elderly
 - c. Sufferers of any serious medical conditions i.e. you fall into one of the following categories
 - i. Inpatient care
 - ii. Incapacity
 - iii. Chronic serious health conditions
 - iv. Permanent or long term conditions
 - v. Conditions requiring multiple treatments
 - d. Sleep deprived
 - e. Under the influence of alcohol
 - f. Previously suffered concussion or traumatic brain injury
 - g. Prone to dizziness from immersive virtual experiences
 - h. Sufferers of panic attacks or generalised anxiety disorders which might be provoked by wearing headphones / being unable to hear your surroundings
 - i. Prone to issues with balance or motor function (i.e. you can walk around a room over the course of an hour).
4. Be comfortable with wearing a head-mounted display (HMD) such as a Oculus Quest 2.

1. Invitation

You are being invited to take voluntarily part in a research experiment. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

Thank you for reading this.

2. Purpose of the User Study

We would like to reach out to you to participate in a paid study examining how users interact with an automated-teller-machine (ATM) in the lab and through the help of virtual reality. Your participation is voluntary, and you are free to withdraw at any time, without giving any reason, and you are free to omit answering any particular question, without providing a reason. The analysis of this experiment will be

published at top-tier venues such as CHI, the premier international conference of Human-Computer Interaction, SOUPS, Symposium on Usable Privacy and Security, and IEEE VR/ACM VRST. All publications are fully anonymised and findings and specific measurements cannot be traced back to you.

3. What will happen to me if I take part?

You will be fitted with a virtual reality headset such that it sits comfortably on your head, and you can hear correctly. This will occur in front of the experimenters' office (Austria). Note that the entire experiment takes place outside and the place is protected from rain. You are tasked to interact with an ATM in the real world and in virtual reality. We will introduce you to the different interaction scenarios and to the virtual environment that we use for this study. You will then go through the different scenarios and are asked to fill in a set of standardised questionnaires which measure your perceived workload, your sense of presence, and the system's usability. We will also add some individual questions to collect additional qualitative data about your experience and perception when using the system.

During the study, we will capture the following data:

Virtual Reality:

- We are going to record the graphical representation of the entire virtual environment. This includes static objects within the environment (e.g., the ATM) and your movements. For all recordings we will use appropriate file extensions (e.g., .csv, .txt, .mp4) and store them in separate files on our local machine. We will upload the anonymised data (through participant ID's) to the University of Glasgow cloud.

Real World:

- We are also going to record you in the real world at times where you interact with our system. This may include the recording of a video from different perspectives and/or taking photographs. You can see an example of this in the picture below. The picture shows a researcher (Florian Mathis) performing a task in virtual reality. Florian is equipped with a VR headset (i.e., HTC VIVE and two HTC VIVE controllers). The way we are going to record your interactions is similar to the picture below. For this study, we are going to use the Oculus Quest 2 without any additional controllers..



At the end of the study, there will be some additional questions within the context of the experiment. This will be in the form of semi-structured interviews. This helps the research team to better understand the experience you have undertaken. At the end of each 1h session we will reimburse you with 10€ per hour in cash.

4. Why have I been chosen?

Your participation has been solicited through emails, social media postings, word-of-mouth, notice board postings, or a call on Prolific to which you replied. Your participation is voluntary, and you are free to withdraw at any time, without giving any reason, and you are free to omit answering any particular question, without providing a reason.

5. Conditions and Data Storage

Each experiment will last for approximately 1 hour. All gathered data during the sessions will be stored directly in the University of Glasgow cloud to keep it confidential (<https://gla-my.sharepoint.com>) or locally. In the case where data is stored on your local device (e.g., as .csv or .txt) we ask you at the end of the session to send us the corresponding files. Access to the raw data is restricted to the researcher (Florian Mathis) and his supervisors (Dr. Mohamed Khamis, Dr. Kami Vaniea) only. Your data is fully anonymised and there is no way to trace it back to you. The results of the study may appear in a number of published studies, in a confidential format where anonymity is preserved. Based on your agreement we will use the data (e.g., screenshots, drawings) as video and image material for scientific papers and/or presentations at conferences.

6. Data Usage

The data will be used within our research and is part of Florian Mathis' Research Phd. We will store the raw data in the University of Glasgow cloud (<https://gla-my.sharepoint.com>). Access to the raw data is restricted to the researcher (Florian Mathis) and his supervisors (Dr. Mohamed Khamis, Dr. Kami Vaniea) only. Based on the request of participants the data can be destroyed at any point. The data will be kept until beyond the end of the Research PhD (up to 10 years) and findings of the experiment might be re-used for additional research projects within Florians' Research PhD.

7. Who has reviewed the study?

This study adheres to the BPS ethical guidelines, and has been approved by the College of Science and Engineering ethics committee of The University of Glasgow. The approved study was also submitted to the University of Edinburgh ethics committee as the lead researcher Florian Mathis is part of both universities.

8. Funding and Contact

This research is supported by the University of Edinburgh and the University of Glasgow jointly funded PhD studentships: <https://www.gla.ac.uk/research/ourresearchenvironment/prs/uofguofedinphdstudentships/>

The project has been reviewed and approved by the Research Ethics Committee in the School of Computing Science at the University of Glasgow. For further information please feel free to get in touch with the researcher f.mathis.1@research.gla.ac.uk (or via phone: [REDACTED]).

Whilst you are free to discuss your participation in this study with the researcher, if you would like to speak to someone not involved in the study, you may contact the Ethics Committee at Christoph.Scheepers@glasgow.ac.uk.

For further information, or if you wish to receive a summary of the findings of this experiment at a later date, please contact the researcher or the supervisor of this project, details listed below.

Data Protection and Confidentiality

Your data will be processed in accordance with the Data Protection Act 1998 (up until 24th May 2018) and the General Data Protection Regulation 2016 (GDPR) thereafter. All information collected about you will be kept strictly confidential. Unless they are anonymised in our records, your data will be referred to by a unique participant number rather than by name. If you consent to being audio recorded, all recordings will be destroyed once they have been transcribed. Your data will only be viewed by the researcher/research team. All electronic data will be stored on a password-protected computer file within the School of Computing Science. All paper records will be stored in a locked filing cabinet within the School of Computing Science. Your consent information will be kept separately from your responses in order to minimise risk in the event of a data breach.

Data Protection Rights

University of Glasgow is a Data Controller for the information you provide. You have the right to access information held about you. Your right of access can be exercised in accordance with the Data Protection Act 1998 (up until 24th May 2018) and the General Data Protection Regulation thereafter. You also have other rights including rights of correction, erasure, objection, and data portability. For more details, including the right to lodge a complaint with the Information Commissioner's Office, please visit www.ico.org.uk. Questions, comments and requests about your personal data can also be sent to the University Data Protection Officer - dp@gla.ac.uk (<https://www.gla.ac.uk/myglasgow/dpfooffice/contact/>)

Thank you for volunteering to take part in this study. If there are any questions or issues, or if you wish to receive a summary of the findings of this experiment at a later date, please feel free to get in touch with the researcher at any time.

Researcher

Florian Mathis

Email: f.mathis.1@research.gla.ac.uk**Tel:** [REDACTED]**1st Supervisor (University of Glasgow)**

Dr. Mohamed Khamis

Email: Mohamed.Khamis@glasgow.ac.uk**Tel:** [REDACTED]

Consent Form



CONSENT FORM

Title of Experiment: In-Situ Evaluation of Authentication Schemes Using Virtual Reality

Experimenter details: Florian Mathis (f.mathis.1@research.gla.ac.uk, [REDACTED])

Supervisor details: Dr. Mohamed Khamis (Mohamed.khamis@glasgow.ac.uk)
Dr. Kami Vaniea (kvaniea@inf.ed.ac.uk)

Before agreeing to this consent form, you should have been given an information sheet to read, which outlines exclusion criteria and explains the general purpose of this experiment and the tasks it involves. If you did not receive this, please inform the researcher (Florian Mathis, f.mathis.1@research.gla.ac.uk). Please tick the box after each statement to indicate that you have read and understand the statement, and that you agree with it.

	If you agree with the statement on the left please tick the box below.
1. I confirm that I have read and understand the Participant Information sheet, and understand my Data Protection Rights under GDPR for the above study, and have had the opportunity to ask questions.	<input type="checkbox"/>
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, and am free to omit answering any particular question, without providing a reason.	<input type="checkbox"/>
3. I give consent for my actions to be recorded (audio and video) during the study.	<input type="checkbox"/>
4. I understand that all data collected from me will be treated confidentially and anonymized, will be seen in its raw form only by the experimenters, and if published will not be identifiable as coming from me.	<input type="checkbox"/>
5. I agree that the researchers can use video recordings for public outreach, for instance, showing parts of the recordings at conference venues and/or use the material in videos and publications to showcase the system.	<input type="checkbox"/>
6. By agreeing to take part in this study I also agree that recordings and data can be used for follow-up evaluations by researchers in the school of computing science and their collaborators.	<input type="checkbox"/>
7. I agree that the researchers are allowed to archive all data taken during the experiment (e.g., video recordings with a camera in the real world) in online repositories such as Enlighten: Research Data: http://researchdata.gla.ac.uk/ . I am aware of the fact that I can get in	<input type="checkbox"/>



University of Glasgow | School of Computing Science



GIST
GLASGOW INTERACTIVE
SYSTEMS GROUP

touch with the researchers at any time to demand the deletion or retrieval of these recordings.	
8. I agree to take part in the above study (In-Situ Evaluation of Authentication Schemes Through Virtual Reality).	<input type="checkbox"/>

This study has been approved by the Ethics Committee.

<p>By signing this form, you have read the conditions stated above and agree to take part in the study.</p> <p>FULL NAME: _____</p> <p>SIGNATURE: _____</p> <p>DATE and PLACE: _____</p>
--

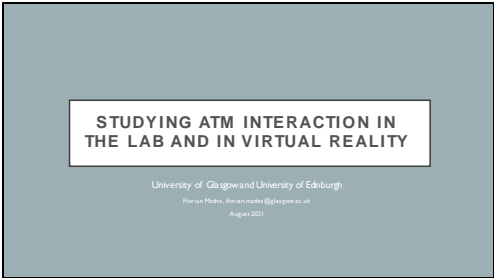
Ranking

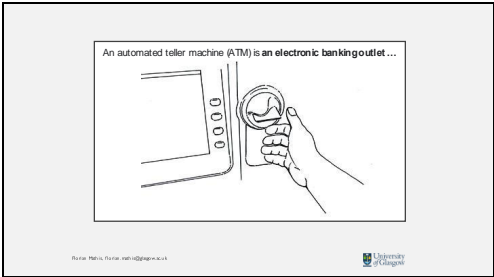
Please rank the ATM experiences according to their realism. The most realistic scenario should be ranked first (1) whereas the least realistic scenario should be ranked last (5).

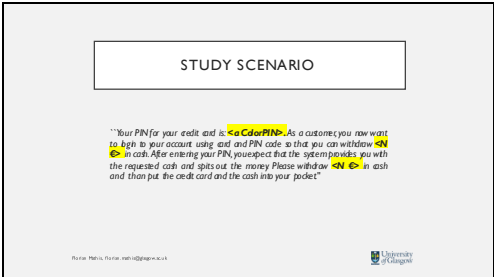
SCENARIO	RANK:
LAB-BASED REAL WORLD ATM	Rank:
LAB-BASED REAL WORLD LAB	Rank:
VIRTUAL REALITY LAB	Rank:
VIRTUAL REALITY ATM (SAME ENVIRONMENT AS IN THE REAL WORLD)	Rank:
VIRTUAL REALITY ATM PUBLIC (WITH BYSTANDERS)	Rank:

Slide Deck for User Study Introduction

07.10.2022

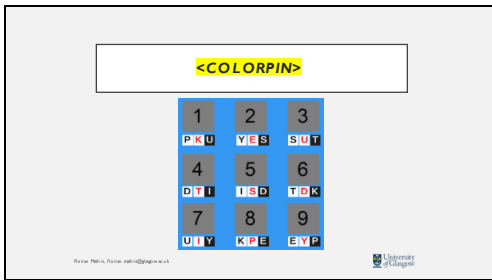


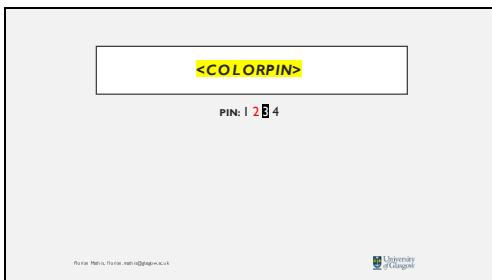




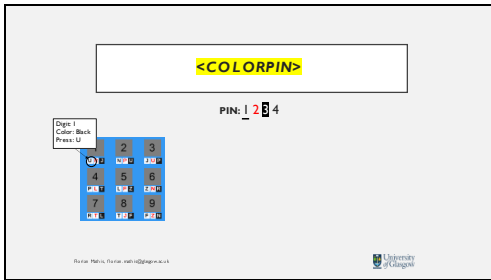
07.10.2022

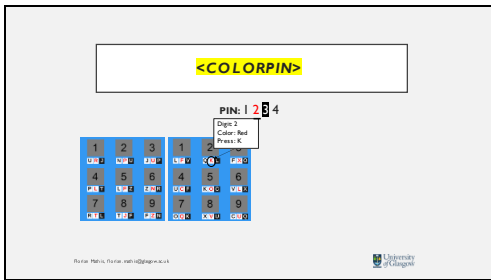


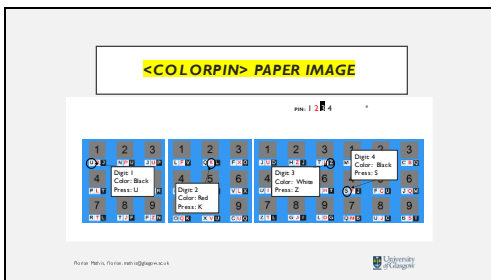




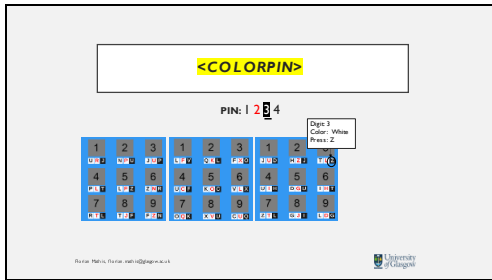
07.10.2022

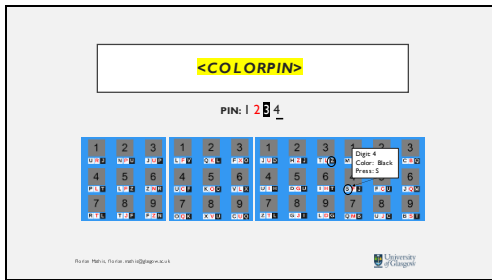






07.10.2022








07.10.2022

EXAMPLE: <COLORPIN>

3 2 8

Digs: 3
Color: Black
Press: J

1	2	3
4	5	6
7	8	9
0	*	#

Rohan Patel, rohan.patel@open.ac.uk 

EXAMPLE: <COLORPIN>


3 2 8

Digs: 3
Color: Black
Press: J

Digs: 2
Color: Red
Press: K

1	2	3
4	5	6
7	8	9
0	*	#

1	2	3
4	5	6
7	8	9
0	*	#

Rohan Patel, rohan.patel@open.ac.uk 

EXAMPLE: <COLORPIN>

3 2 8

Digs: 3
Color: Black
Press: J


Digs: 2
Color: Red
Press: K

Digs: 9
Color: White
Press: G

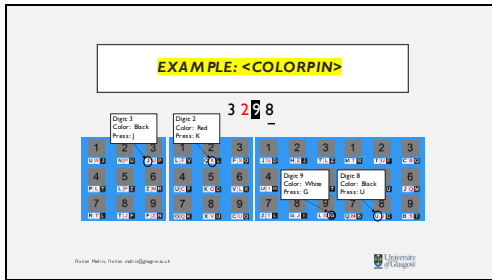
1	2	3
4	5	6
7	8	9
0	*	#

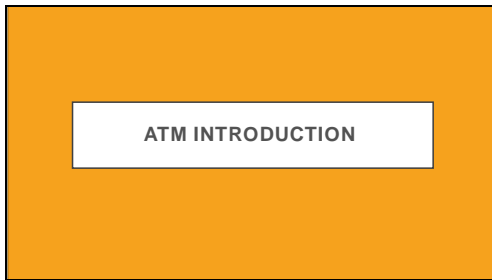
1	2	3
4	5	6
7	8	9
0	*	#

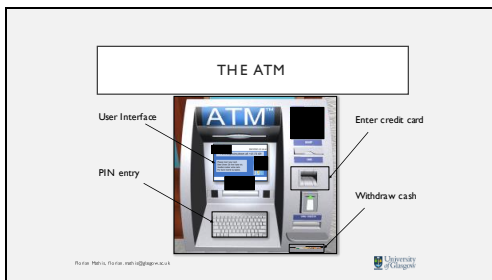
1	2	3
4	5	6
7	8	9
0	*	#

Rohan Patel, rohan.patel@open.ac.uk 

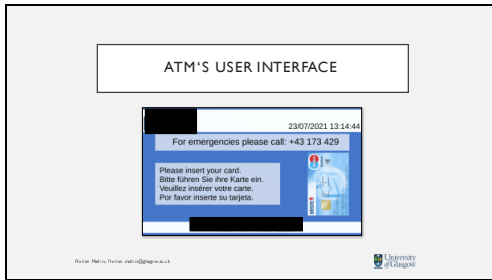
07.10.2022

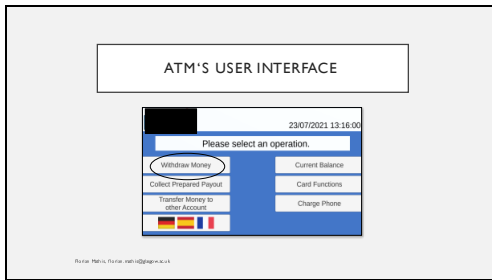


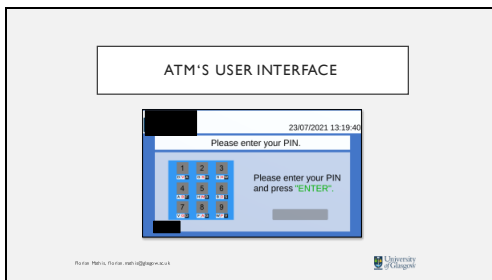




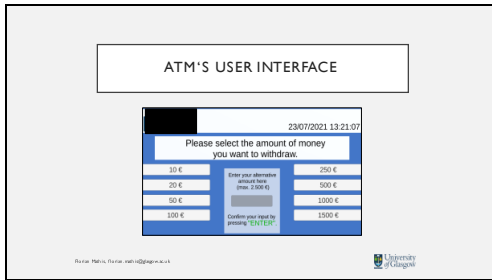
07.10.2022

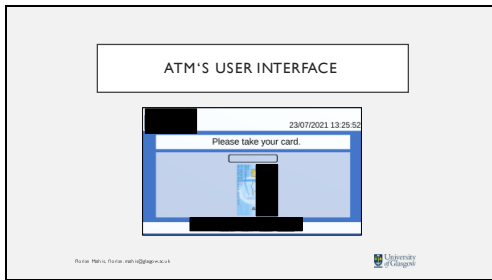


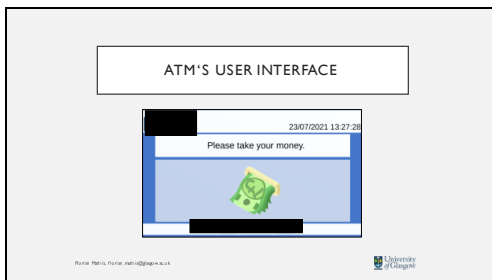




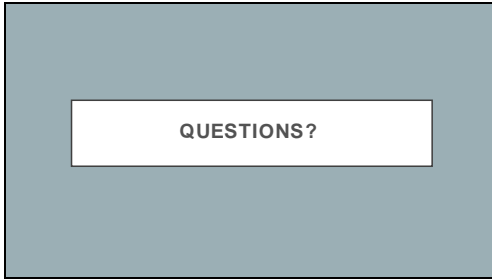
07.10.2022







07.10.2022



Structured Interview: Authentication Scenario

Statements (*) are answered on 5-point Likert scales (*Strongly Disagree – Strongly Agree*).

1. Please walk us, in detail, through the task you have just experienced.
2. What was your main goal? Please explain why.
3. What were the tasks that were required to achieve that goal?
4. What (if any) is the difference between withdrawing cash at a real-world bank ATM and what you have just experienced?
5. *If participants' cash withdrawal was not successful (e.g., wrong PIN):* What were the main difficulties when trying to withdraw the amount of cash we asked you to withdraw?
6. "While completing the task, I felt I was part of a laboratory study." *
7. "I was aware of the experimenter during the task." *
8. "The experimenter's presence impacted my performance negatively." *
9. "The experimenter's presence impacted my behaviour." *
10. "I found that recalling the PIN made it more challenging to complete the other cash withdrawal steps." *
11. "I found that the other cash withdrawal steps made it more challenging to recall the correct PIN." *

Semi-structured Interview

We used a semi-structured interview approach at the end of the study. The following questions were used to roughly ask the same questions to all participants but due to the nature of a semi-structured interview approach the questions differed across the participants.

1. Could you please walk us through your ranking on: "Which experience did you perceive as most similar to using an ATM in the real world?"
2. How did you feel about interacting with the ATM in the real world? What was easy and/or challenging?
3. How did you feel about interacting with the ATM in virtual reality? What was easy and/or challenging?
4. Please consider the experienced environment and a real-world environment where you are standing in front of an ATM. What would be different to what you have just experienced in:

- (a) our real-world part of the study?
- (b) our VR part of the study?
5. Did the amount of cash you had to withdraw impact your authentication behaviour?
If so, how?
6. Do you regularly shield your PIN entry when using an ATM in the real world?
7. *If yes to 6):* How do you shield your PIN entry?
8. Did you shield your PIN entry in the study? Why? Why not?
9. *If yes to 8):* How did you shield your PIN entry in the study?
10. What do you think this study is about?

Raw NASA-TLX Scores for Each Subdimension

Table D.1: The individual dimensions of the NASA-TLX scores. No post-hoc tests on the level of each dimension were performed due to the lack of significance of the overall mean raw NASA-TLX scores. Scores represent the means and the standard deviations.

NASA-TLX	(1) <i>RW Lab</i>	(2) <i>RW ATM</i>	(3) <i>VR Lab</i>	(4) <i>VR ATM</i>	(5) <i>VR ATM Public</i>		
Mental (ColorPIN only)	44.00 (29.18)	59.50 (28.85)	41.75 (28.38)	58.25 (26.38)	61.50(25.70)	No significant main effects on the overall NASA-TLX scores.	
Mental (ATM + ColorPIN)	n/a	39.00 (28.31)	n/a	50.00 (32.71)	58.25 (27.81)		
Physical (ColorPIN only)	9.25 (9.91)	10.25 (11.34)	19.00 (20.10)	13.25 (19.38)	19.00 (18.68)		
Physical (ATM + ColorPIN)	n/a	15.75 (18.05)	n/a	22.25 (21.12)	23.50 (18.38)		
Temporal (ColorPIN only)	33.25 (31.32)	24.50 (24.89)	25.25 (26.90)	25.00 (27.88)	30.75 (27.31)		
Temporal (ATM + ColorPIN)	n/a	25.50 (24.59)	n/a	22.75 (26.05)	39.50 (22.13)		
Performance (ColorPIN only)	34.75 (39.48)	31.25 (35.53)	27.25 (33.30)	23.00 (32.65)	33.00 (33.44)		
Performance (ATM + ColorPIN)	n/a	37.25 (35.62)	n/a	31.50 (35.11)	33.25 (34.14)		
Effort (ColorPIN only)	42.25 (27.36)	44.00 (32.58)	43.00 (30.47)	44.25 (26.80)	55.25 (24.47)		
Effort (ATM + ColorPIN)	n/a	44.75 (25.57)	n/a	50.00 (25.45)	51.25 (24.02)		
Frustration (ColorPIN only)	27.25 (21.24)	36.50 (28.86)	34.00 (28.09)	34.5 (27.88)	40.75 (25.85)		
Frustration (ATM + ColorPIN)	n/a	37.00 (26.29)	n/a	34.50 (25.59)	39.50 (22.63)		
Overall Workload Score (ColorPIN only)	31.79 (30.48)	34.33 (32.03)	31.71 (29.49)	33.04 (30.91)	40.04 (30.03)		p> 0.05
Overall Workload Score (ATM + ColorPIN)	n/a	33.21 (28.60)	n/a	35.17 (30.29)	40.88 (27.78)		p> 0.05

Appendix E

Appendix for Chapter 6

Participant Information Sheet



Participant Information Sheet

Title of Experiment: In Situ Evaluation of Authentication Schemes Using Virtual Reality

Experimenter details: Florian Mathis (f.mathis.1@research.gla.ac.uk, [REDACTED])

Supervisor details: Dr. Mohamed Khamis (Mohamed.khamis@glasgow.ac.uk)

Dr. Kami Vaniea (kvaniea@inf.ed.ac.uk)

IMPORTANT – Exclusion criteria

In order to take part in this study, you must meet the following requirements:

1. Aged 16 or over
2. Access to an Oculus Quest 1 or Quest 2 and at least 1.5m free walking space in all four directions (left, right, front, back)
3. No history (personal and family) of epileptic seizures, strokes, or photosensitivity
4. Not a member of any of the following groups
 - a. Pregnant women
 - b. The elderly
 - c. Sufferers of any serious medical conditions i.e. you fall into one of the following categories
 - i. Inpatient care
 - ii. Incapacity
 - iii. Chronic serious health conditions
 - iv. Permanent or long term conditions
 - v. Conditions requiring multiple treatments
 - d. Sleep deprived
 - e. Under the influence of alcohol
 - f. Previously suffered concussion or traumatic brain injury
 - g. Prone to dizziness from immersive virtual experiences
 - h. Sufferers of panic attacks or generalised anxiety disorders which might be provoked by wearing headphones / being unable to hear your surroundings
 - i. Prone to issues with balance or motor function (i.e. you can walk around a room over the course of an hour).
5. Be comfortable with wearing your own Oculus Quest 1/2 for several minutes.
6. Be comfortable with filling in a survey and interacting with a VR environment. Note that the interaction with the virtual environment is without any additional controllers (hand tracking only).
7. Be comfortable with installing a virtual reality application (.apk) on your Oculus Quest 1 or Quest 2
8. Be comfortable with being on a Zoom/Skype call (<https://zoom.us/>, <https://www.skype.com/en/>) with a researcher from the University of Glasgow for the duration of the user study.

Florian Mathis
florian.mathis@glasgow.ac.uk



1. Invitation

Thank you for your interest in participating in this study about automated-teller-machine (ATM) authentication and virtual reality (VR). You are being invited to take part in this research study. Before you decide to take part it is important for you to understand why this research is being done and what it will involve. Please take time to read the following information carefully. If you have any questions regarding this research please feel free to ask the experimenter Florian Mathis (florian.mathis@glasgow.ac.uk).

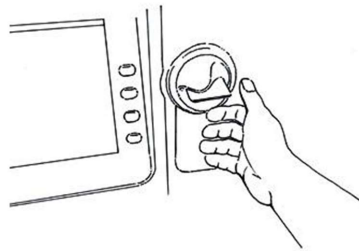


Fig. The figure shows a user entering their credit card into an automated-teller-machine (ATM) to withdraw cash.

2. Purpose of the User Study

We would like to reach out to you to participate in a paid study. Your participation is voluntary, and you are free to withdraw at any time, without giving any reason, and you are free to omit answering any particular question, without providing a reason. The analysis of this experiment is part of Florian's PhD and will be published at top-tier venues such as CHI, the premier international conference of Human-Computer Interaction, SOUPS, Symposium on Usable Privacy and Security, and IEEE VR. All publications are fully anonymised and findings and specific measurements cannot be traced back to you.

3. What will happen to me if I take part?

You are going to interact with four different authentication schemes in a virtual environment. You will be asked to roleplay an ATM cash withdrawal scenario and use the four authentication systems to authenticate. They are a) traditional 4-digit PIN Authentication, b) Glass Unlock Authentication, an authentication system that shows the key-pad layout on the visual private channel of augmented reality glasses, c) Hand Menu Authentication, an authentication system that leverages augmented reality glasses for mid-air touch input next to the user's palm, and d) Tap Authentication, an authentication system that leverages augmented reality glasses to map the digits (0-9) to the user's fingers.

For the user study, you will be fitted with a virtual reality headset (your own Oculus Quest 1/2) such that it sits comfortably on your head. After each task (e.g., authenticating using one of the four authentication systems), you are asked a few questions that are part of standardised in-VR questionnaires. We conclude the study with a semi-structured interview and some follow-up questions on Qualtrics.

4. Why have I been chosen?

Your participation has been solicited through social media, mailing lists, and word-of-mouth. By completing this study you will receive an Amazon voucher € 15 (in your local currency). You will also be given the opportunity to find out more about this research at the end of the study. Please note that to receive the reimbursement we ask you to copy the collected data from your VR headset to your PC and send us the data by uploading it to a shared folder that is stored on the University of Glasgow cloud.

Florian Mathis
florian.mathis@glasgow.ac.uk

5. Conditions and Data Storage

Each experiment will last for approximately 1 hour (+ some additional time for the preparation of the headset in advance of the actual user study). All gathered data during the session will be stored in the University of Glasgow and University of Edinburgh cloud to keep it confidential. Access to the raw data is restricted to the researcher (Florian Mathis) and his supervisors (Dr. Mohamed Khamis, Dr. Kami Vaniea) only. Data is fully anonymised and there is no way to trace it back to you. The results of the study may appear in a number of published studies, in a confidential format where anonymity is preserved. Based on your agreement we will use findings for scientific papers and/or presentations at conferences. Based on the request of participants the data can be destroyed at any point. The data will be kept until beyond the end of the Research PhD (up to 10 years) and findings of the experiment might be re-used for additional research projects within Florian`s research.

6. Who has reviewed the study?

This study adheres to the BPS ethical guidelines, and has been approved by the College of Science and Engineering ethics committee of The University of Glasgow.

7. Funding and Contact

This research is supported by the University of Edinburgh and the University of Glasgow jointly funded PhD studentships: <https://www.gla.ac.uk/research/ourresearchenvironment/prs/uofguofedinhpdstudentships/>

The project has been reviewed and approved by the Research Ethics Committee in the School of Computing Science at the University of Glasgow. For further information please feel free to get in touch with the researcher f.mathis.1@research.gla.ac.uk (or via phone: [REDACTED])

Whilst you are free to discuss your participation in this study with the researcher, if you would like to speak to someone not involved in the study, you may contact the Ethics Committee at Christoph.Scheepers@glasgow.ac.uk.

For further information, or if you wish to receive a summary of the findings of this experiment at a later date, please contact the researcher or the supervisor of this project, details listed below.

Data Protection and Confidentiality

Your data will be processed in accordance with the Data Protection Act 1998 (up until 24th May 2018) and the General Data Protection Regulation 2016 (GDPR) thereafter. All information collected about you will be kept strictly confidential. Unless they are anonymised in our records, your data will be referred to by a unique participant number rather than by name. If you consent to being audio recorded, all recordings will be destroyed once they have been transcribed. Your data will only be viewed by the researcher/research team. All electronic data will be stored on a password-protected computer file within the School of Computing Science. All paper records will be stored in a locked filing cabinet within the School of Computing Science. Your consent information will be kept separately from your responses in order to minimise risk in the event of a data breach.

Data Protection Rights

University of Glasgow is a Data Controller for the information you provide. You have the right to access information held about you. Your right of access can be exercised in accordance with the Data Protection Act 1998 (up until 24th May 2018) and the General Data Protection Regulation thereafter. You also have other rights including rights of correction, erasure, objection, and data portability. For more details, including the right to lodge a complaint with the Information Commissioner's Office, please visit www.ico.org.uk. Questions, comments and requests about your personal data can also be sent to the University Data Protection Officer - dp@gla.ac.uk (<https://www.gla.ac.uk/myglasgow/dpfooffice/contact/>)

Thank you for volunteering to take part in this study. If there are any questions or issues, or if you wish to receive a summary of the findings of this experiment at a later date, please feel free to get in touch with the researcher at any time.

Researcher

Florian Mathis

Email: florian.mathis@glasgow.ac.uk

Tel: [REDACTED]

1st Supervisor (University of Glasgow)

Dr. Mohamed Khamis

Email: Mohamed.Khamis@glasgow.ac.uk

Tel: [REDACTED]

Florian Mathis
florian.mathis@glasgow.ac.uk



Consent Form



CONSENT FORM

Note that this consent form will be distributed to participants who have already agreed to participate in the study in advance of the study. This means that participants provided their consent to take part in the study prior to the actual user study and will have the chance to get in touch with the lead researcher if there are any further questions.

Title of Experiment: In-Situ Evaluation of Authentication Schemes Using Virtual Reality

Experimenter details: Florian Mathis (f.mathis.1@research.gla.ac.uk, [REDACTED])

Supervisor details: Dr. Mohamed Khamis (Mohamed.khamis@glasgow.ac.uk)
Dr. Kami Vaniea (kvaniea@inf.ed.ac.uk)

Before agreeing to this consent form, you should have been given an information sheet to read, which outlines exclusion criteria and explains the general purpose of this experiment and the tasks it involves. If you did not receive this, please inform the researcher (Florian Mathis, f.mathis.1@research.gla.ac.uk). Please tick the box after each statement to indicate that you have read and understand the statement, and that you agree with it.

	If you agree with the statement on the left please tick the box below.
1. I confirm that I have read and understand the Participant Information sheet, and understand my Data Protection Rights under GDPR for the above study, and have had the opportunity to ask questions.	<input type="checkbox"/>
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, and am free to omit answering any particular question, without providing a reason.	<input type="checkbox"/>
3. I give consent for my actions to be recorded (user inputs in specific text-fields, audio recording) during the study. Note that the recording is two-fold: (1) We will record the audio of the Zoom/Skype call for follow up analysis, and (2) we ask you to record your virtual reality view locally on your headset (using the built-in recording: https://www.roadtovr.com/increase-video-capture-resolution-quest-2-recording-quality/).	<input type="checkbox"/>
4. I understand that all data collected from me will be treated confidentially and anonymized, will be seen in its raw form only by the experimenters, and if published will not be identifiable as coming from me.	<input type="checkbox"/>
5. I agree that the researchers can use the virtual reality recordings (no audio) for public outreach. For example, to show parts of the recordings at conference venues and/or use the material in videos and publications to showcase the system.	<input type="checkbox"/>

6. By agreeing to take part in this study I also agree that recordings and data can be used for follow-up evaluations by researchers in the school of computing science and their collaborators (e.g., Dr. Kami Vaniea from the University of Edinburgh).	<input type="checkbox"/>
7. I agree that the researchers are allowed to archive all data taken during the experiment in online repositories such as Enlighten: Research Data: http://researchdata.gla.ac.uk/ . I am aware of the fact that I can get in touch with the researchers at any time to demand the deletion or retrieval of these recordings.	<input type="checkbox"/>
8. I agree to take part in the above online virtual reality study (In-Situ Evaluation of Authentication Schemes Through Virtual Reality).	<input type="checkbox"/>

This study has been approved by the Ethics Committee.

By signing this form, you have read the conditions stated above and agree to take part in the study.	
FULL NAME:	_____
SIGNATURE:	_____
DATE and PLACE:	_____

Ranking

[Post-Study Questionnaire] Ranking

Start of Block: Study ID

Q4 Please enter your study ID here.

End of Block: Study ID

Start of Block: Ranking of the Authentication Systems



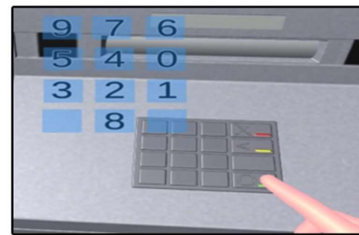
Q499 Please rank the authentication systems based on your **perceived usability**. The most usable authentication method should be ranked 1 (top) and the least usable authentication method should be ranked 4 (bottom).

Authentication Systems

Traditional 4-digit PIN Authentication



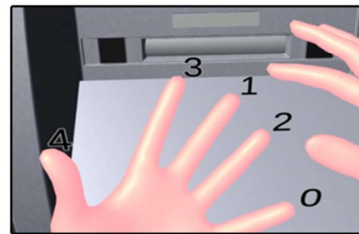
Glass Unlock Authentication



Hand Menu Authentication



Tap Authentication



- _____ Traditional 4-digit PIN Authentication (1)
 _____ Glass Unlock Authentication (2)
 _____ Hand Menu Authentication (3)
 _____ Tap Authentication (4)



Q500 Please rank the authentication systems based on your **perceived security**. The most secure authentication method should be ranked 1 (top) and the least secure authentication method should be ranked 4 (bottom). Note that with the term security we refer to the system's resistance against shoulder surfing, i.e. a bystander observes you authenticating in front of the cash machine (see below).

Authentication Systems

Traditional 4-digit PIN Authentication



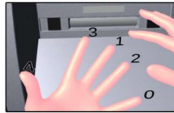
Glass Unlock Authentication



Hand Menu Authentication



Tap Authentication



- _____ Traditional 4-digit PIN Authentication (1)
 _____ Glass Unlock Authentication (2)
 _____ Hand Menu Authentication (3)
 _____ Tap Authentication (4)



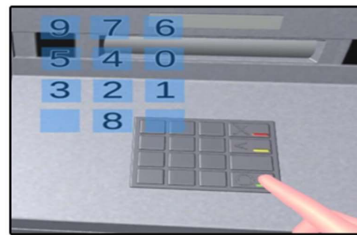
Q501 Please rank the authentication systems based on your **combined perceived usability and security**. Note that this means that the authentication method you would like to use when authenticating in the public (e.g. on a cash machine) should be ranked 1 (top) and the authentication method you disliked should be on rank 4 (bottom).

Authentication Systems

Traditional 4-digit PIN Authentication



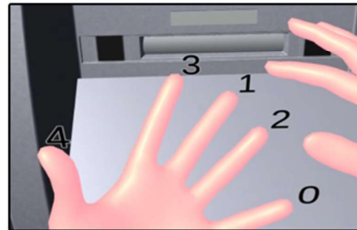
Glass Unlock Authentication



Hand Menu Authentication



Tap Authentication



- _____ Traditional 4-digit PIN Authentication (1)
 _____ Glass Unlock Authentication (2)
 _____ Hand Menu Authentication (3)
 _____ Tap Authentication (4)

End of Block: Ranking of the Authentication Systems

Study Material (Email, Instructions, Slide Deck, etc.)

Dear Participant,

Thanks a lot for showing interesting in our remote virtual reality (VR) user study. We scheduled a zoom session for your selected time slot on [REDACTED]. You can join via Zoom using the following link: [REDACTED]

Todo *before* the Zoom session:

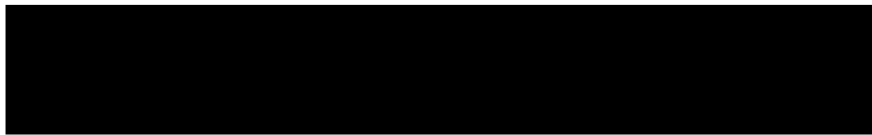
You can find a participant information sheet and a consent form attached to this document. It would be great if you can return the signed consent form and the filled in pre-study questionnaire in advance of the user study (at least 1 day prior to the user study). Your unique ID is: [REDACTED]

1. Return the signed consent form (consent_form.pdf)
2. Fill in the pre-study questionnaire (ID: [REDACTED])
[REDACTED]
3. Make sure that you have **at least 250MB storage space** available on your Oculus Quest. **Ideally, you would have 3 GB storage space available** to be able to record your VR view for the duration of the user study. However, the .apk requires at least 250 MB and the additional storage space is optional if casting your Oculus Quest view to the browser works (see point 8 and the instructions in installation.pdf).
4. Sideload/Install our study .apk on your Oculus Quest 1 / 2. For instructions please refer to installation.pdf.
2. Download Link of our study application: [REDACTED]
[REDACTED]
6. Enable Oculus` hand tracking. For instructions, please refer to installation.pdf.
7. Make sure that you have enough room space during the user study (1m to the left, right, front back). You are not required to walk around, but we want to make sure that there is enough space around you to interact with our systems and to ensure your safety. If you have used your VR headset before while standing this should be no problem. You can re-center your view during the study as described in safety.pdf.
8. If you have successfully installed our .apk on your Quest please launch the application for a few seconds to see if your Quest is correctly setup. **Please close the application after a few seconds.** See the example view (example.jpg) that shows how the environment should look like in terms of size and your position. If your height is off, please make sure your Quests` floor level is correctly set up (see <https://support.oculus.com/guardian/> or go to Settings > Guardian > Set Floor Level).
9. You can call the menu by looking at your palm (right hand) at eye level, then hold your thumb and index finger together until the Oculus icon fills up, then release. See installation.pdf for further instructions or <https://support.oculus.com/articles/headsets-and-accessories/controllers-and-hand-tracking/hand-tracking-quest-2>.
10. Check if casting your Oculus Quest view to the browser works. For instructions, please refer to installation.pdf
11. If you experience any issues installing the application on your VR headset please get in touch with the lead researcher, florian.mathis@glasgow.ac.uk.

12. Make sure that your VR headset is charged and that the room you use for the user study is well-lit. Please block out any sun rays coming in from the outside as this can have an impact on the accuracy of your headset's hand tracking.

Todo *during* the Zoom session:

Join the Zoom session on your PC/laptop on the scheduled time slot. Please make sure that you have your Oculus Quest (as configured according to the instructions) next to you – ideally fully charged.



Todo *after* the Zoom session:

Please make sure that you upload all files (see PostStudy_Datatransfer.pdf) at the end of the study to our internal University of Glasgow cloud:



We are looking forward to talking to you! Thanks a lot already for your help.


Best regards,

Florian

Studying Novel ATM Authentication Systems in Virtual Reality

How to: Install Study Application (.apk) on your device

Step 0) Preparing the Headset, your PC, and your Oculus Account.

Make sure that you have a USB cable to connect your Quest to your computer. The charging cable that ships with the Quest is sufficient here (USB-C connector on both ends). If your computer has a USB-C port, you can use that. Otherwise you need a USB-C to USB-A adapter (default USB slot). Please make sure that you have **at least 250MB storage space** available on your Oculus Quest. **Ideally, you would have 3 GB storage space available** to be able to record your VR view for the duration of the user study. However, this is optional as we plan to use Oculus casting via Zoom's screen sharing option. To check how much space is left on your Quest 2 or Quest: 1) Put on your headset. 2) Select  Settings from the bottom toolbar menu. 3) Select Storage.

Enable Developer Mode

1) To use your Quest in developer mode, which is required for sideloading our app, you will need to register as a developer organization first. Note this is completely free.

Visit <https://dashboard.oculus.com/organizations/create/> and make sure you are logged into the same Oculus account that your Quest is using. Enter a new organization name and tick "I understand" box to agree to the Oculus Terms of Service.

- 2) Turn on your quest and go to the Oculus app on your phone that is linked to your Oculus Quest.
 - a) Open the oculus app and tap "Settings" in the bottom right.
 - b) Locate your Oculus Quest listed in the Settings tab, and make sure it reads 'Connected. (If the app can not connect to your Quest, you may need to tap on the Quest in settings to try and manually make the app connect. If it still can't connect, make sure your Quest is turned on, and your phone has Bluetooth and WiFi turned on as well)
 - c) Tap on the arrow button next to your device, to reveal more options
 - d) Tap the 'More Settings' button
 - e) Tap on 'Developer Mode'
 - f) Flick the switch to On instead of Off
 - g) Reboot your Quest—hold down the power button on the side and select 'Power Off' or 'Restart'.

After rebooting, your Quest should be in Developer Mode.

STEP 1) Install APK

If you are not experienced with installing third-party applications on your Quest 1/2, please find some quick steps here. If you prefer to sideload our .apk using SideQuest please have a look at the procedure here: <https://uploadvr.com/sideload-quest-how-to/>.

Step 1.1) Turn on the Oculus

Step 1.2) Open the Oculus app and go to 'Settings'

Step 1.3) Connect to your device and go to 'More Settings'

Step 1.4) Enable the 'Developer Mode'

Step 1.5) Install ADB (Android Debug Bridge), for more info and instruction see <https://developer.android.com/studio/releases/platform-tools> and <https://developer.android.com/studio/command-line/adb>

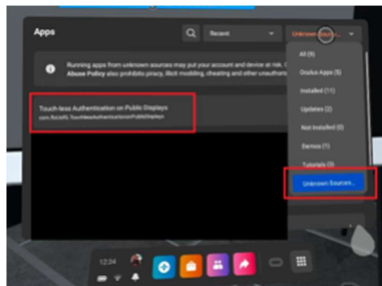
Step 1.6) Download the APK file (link in email)

Step 1.7) Open the CMD/Terminal and navigate to the folder

Step 1.8) Connect your device with the USB cable and allow permission in Oculus, when asked.

Step 1.9) Check that the device is connected/listed with the command adb devices

Step 1.10) Install the .apk file with adb install <apk-path>




Please verify that the installation was successful by heading to Apps > "Unknown Sources" on your VR headset and see if the study apk shows up there.


Please do not launch the application until the user study session.

STEP 2) Configure Oculus Quest

We use Oculus` integrated hand tracking for the duration of our study. If you have never used hand tracking before, please make sure that you enabled it on your Quest

Step 2.1) Press  on your right Touch controller to pull up your universal menu.



Step 2.2) Hover over the clock on the left-hand side of the universal menu. When Quick Settings appears, select it to open the Quick Settings panel.

Step 2.3) Select  *Settings* in the top-right corner.

Step 2.4) Select *Device* from the left menu, then select *Hands and Controllers*.

Step 2.5) Select the toggle next to *Hand Tracking* to turn it on or off.

Once you have enabled Oculus` Hand tracking there are following gestures possible:

Gestures	What it is used for	How to do it	
Point and pinch	To select something	When the cursor appears, point your hand at what you want to select. Then pinch your thumb and finger together to select it.	 Source: Oculus Hand Tracking
Palm pinch	Brings you back to your Oculus Home menu.	Look at your palm at eye level, then hold your thumb and index finger together until the Oculus icon fills up, then release.	 Source: Oculus Hand Tracking

Please see Oculus Quest 2 Hand Tracking Gestures for more information. Source of the images:

<https://support.oculus.com/articles/headsets-and-accessories/controllers-and-hand-tracking/hand-tracking-quest-2>
(updated link: <https://www.meta.com/en-gb/help/quest/articles/headsets-and-accessories/controllers-and-hand-tracking/hand-tracking-quest-2/>)

STEP 3) Cast your Oculus Quest view

Please make sure this works in advance of the user study. You can cast your VR view to your PC for a few seconds to see if the setup works.

Step 3.1) Navigate to <https://www.oculus.com/casting> and login using your Oculus Quest 1 / 2 account.

Step 3.2) Slip on your Quest/Quest 2 headset and select the “Share” option located in the Oculus universal menu.

Step 3.3) Click on “Cast” and select the desired PC (= computer) on your list of available devices.



source: <https://www.techpages.com/vr-oculus-quest-2-how-to-cast-to-a-mobile> and <https://www.techadvisor.com/article/744793/how-to-take-a-screenshot-on-the-oculus-quest-2.html>

After the Study



In order to reset your system to the default settings you need to do the following steps:

1) Uninstall our .apk:



- 1.1) Head to Apps > Unknown Sources
- 1.2) Hover over the ... Menu next to our .apk name.
- 1.3) Remove the application from your VR device by pressing “Uninstall”.

2) Disable Hand tracking:

- 2.1) Press  on your right Touch controller to pull up your universal menu.
- 2.2) Hover over the clock on the left-hand side of the universal menu. When Quick Settings appears, select it to open the Quick Settings panel.
- 2.3) Select  Settings in the top-right corner.
- 2.4) Select Device from the left menu, then select Hands and Controllers.
- 2.5) Select the toggle next to Hand Tracking to turn it off.

3) Disable Developer Mode:

Turn on your Quest headset and open the **Oculus app** on the Android or iOS device you used to set up your Quest.

Follow these steps to enable Developer mode on your Quest:

1. Tap Settings (bottom-right)
2. Select your connected Quest from the device list and connect to it
3. Tap More Settings which appears below your Quest in the device list
4. Tap Developer Mode
5. Tap the switch to disable developer mode
6. Exit Settings on the app & reboot your Quest using the right-side power button

If there are any other questions please get in touch with the lead researcher: florian.mathis@glasgow.ac.uk

Studying Novel ATM Authentication Systems in Virtual Reality

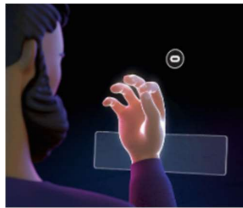
Safety Instructions

Please make sure that you have a play area by at least 1 by 1 metre (3 by 3 feet). We also recommend enabling the Guardian due to safety reasons. Note that if you use your Oculus Quest for Games etc. your settings should already be sufficient.


For further safety instructions please refer to: <https://www.oculus.com/safety-center/quest/> and <https://support.oculus.com/guardian/>

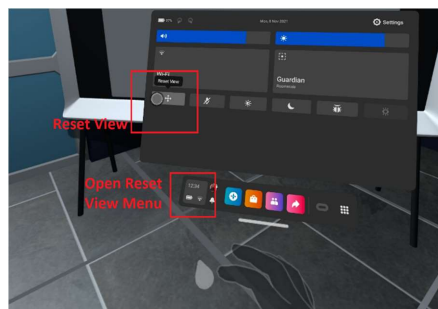
Please note that our study does not require you to move around, but we want to make sure that there is enough space for you to interact with our authentication systems in a safe way.

How to Reset the Direction You're Facing



Source: Oculus Hand Tracking

To reset your view, look straight ahead in the direction you want to be the center of your view. Then make a "Palm pinch gesture". In your headset, you'll see the Oculus logo appear with a quickly filling white ring around it. Once the ring reaches all the way around the Oculus logo, which will take a couple of seconds, you will see the universal menu. You can then press  *Reset View* with a default pinch gesture.



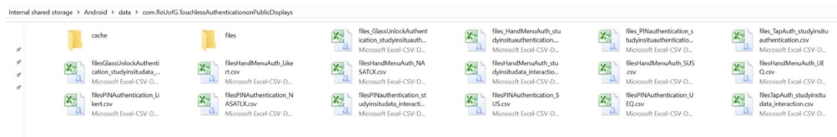
Studying Novel ATM Authentication Systems in Virtual Reality

Data Transfer

Thanks a lot for participating in our user study. For the final step, and to receive the reimbursement, we would like to ask you to send us the collected data that we stored on your Oculus Quest 1 / 2.

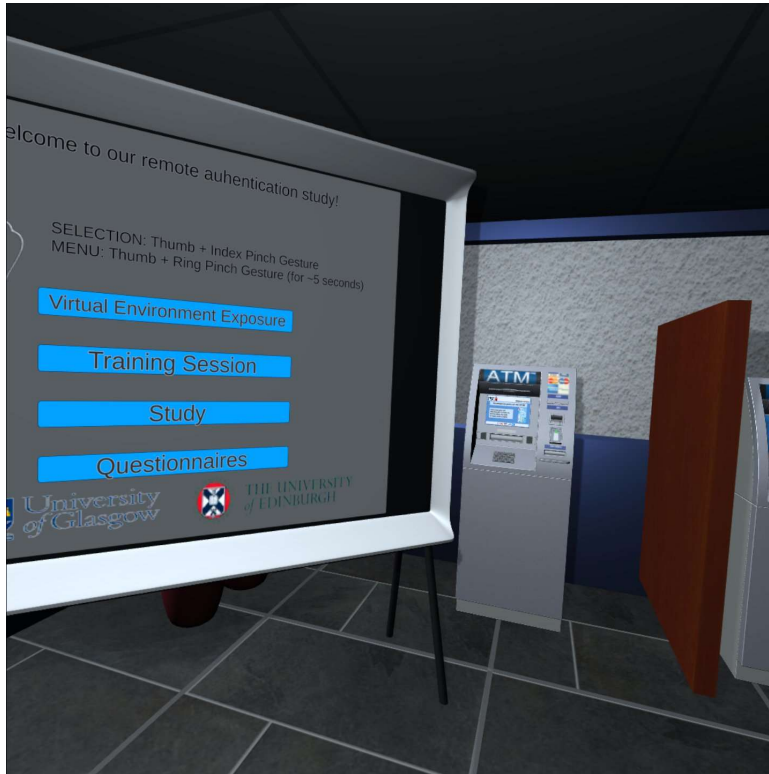
- 1) Start your Oculus Quest 1 / 2.
- 2) Plug your USB-C link cable into a USB-C port on your PC, then plug the other end into your headset.
- 3) A message is displayed in VR on your headset prompting you to **Allow access to data**. Select **Allow**.
- 4) Navigate to your PC and select your Quest: Quest 2 > Internal shared storage > Android > data
- 5) Open the “com.floUofG.TouchlessAuthenticationonPublicDisplays” folder.
- 6) **Make a copy of all .csv files** and store them directly in the University of Glasgow cloud. You should have received the upload link at the end of the user study and can also find it in the initial email sent by the experimenter (Subject: “Authentication on Public Displays: User Study”). Once you have uploaded all files, please get in touch with the experimenter: florian.mathis@glasgow.ac.uk. Please do not delete the data from your Quest until the experimenter checked that all files were fully uploaded to our internal system. Note that this check will happen immediately after the experimenter received your email.

Example files: Select all files and move them to the shared folder.



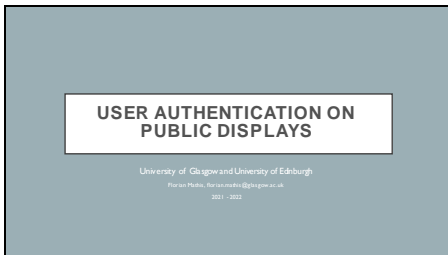
- 6) **Make a copy of the .mp4 video files for the recording (if you recorded the view on your Quest).** Navigate to Quest 2 > Internal shared storage > Oculus > VideoShots and select the most recent recording. If you had to restart the recording at some point during the study please make sure that you copy all video files.

example.jpg for setting up the application.



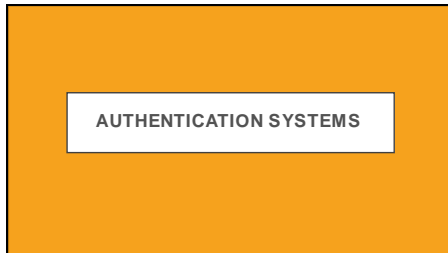
07.10.2022

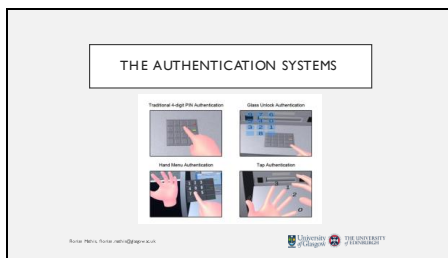


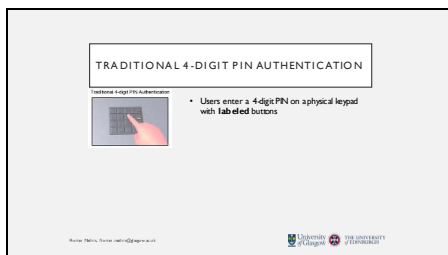




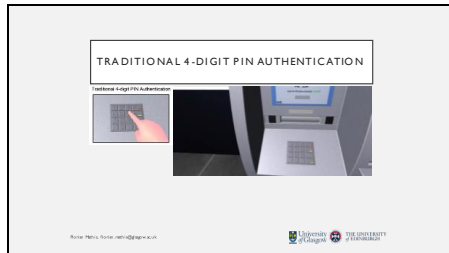
07.10.2022



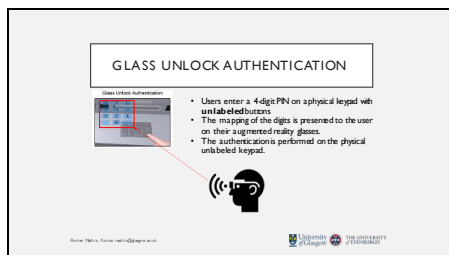




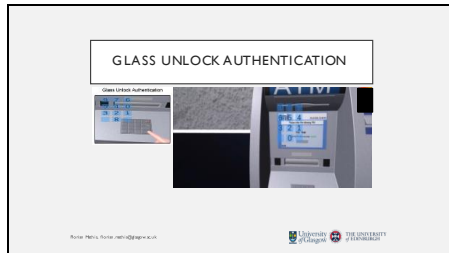
07.10.2022

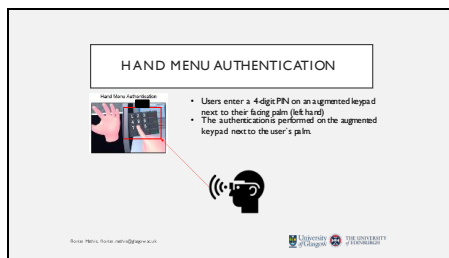


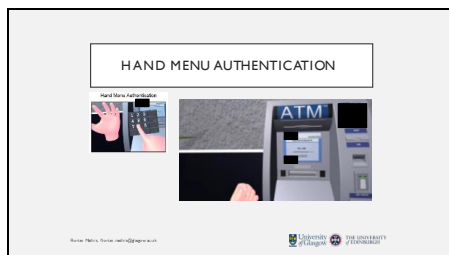




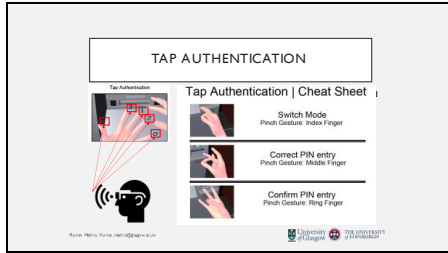
07.10.2022

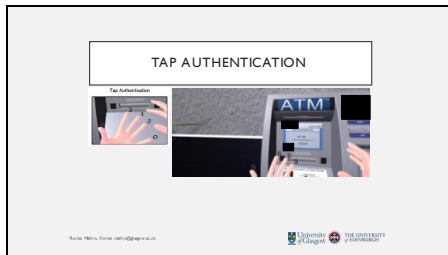


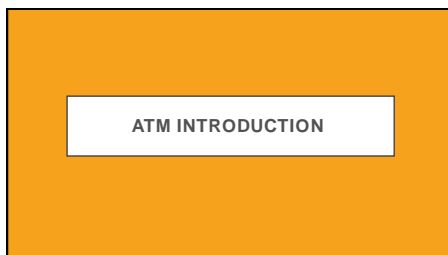




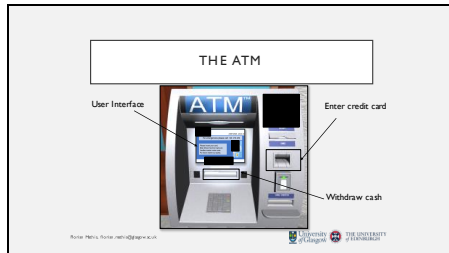
07.10.2022

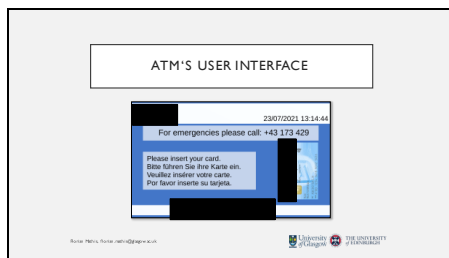


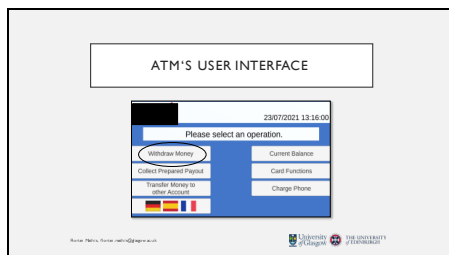




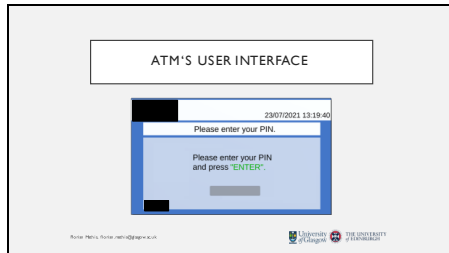
07.10.2022

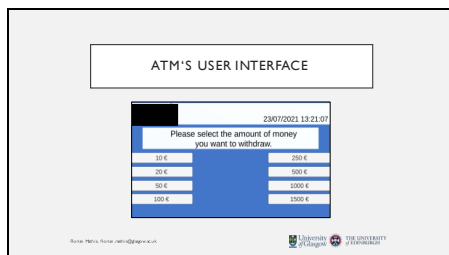


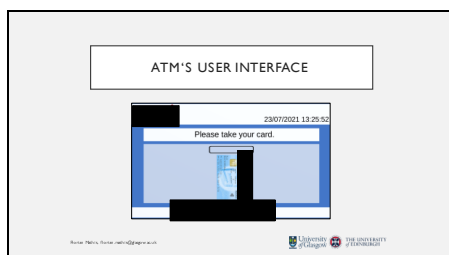




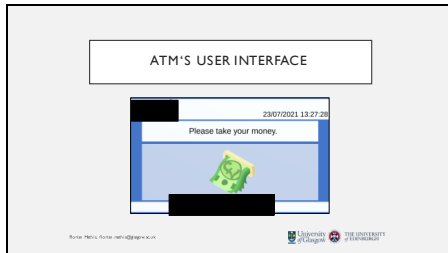
07.10.2022

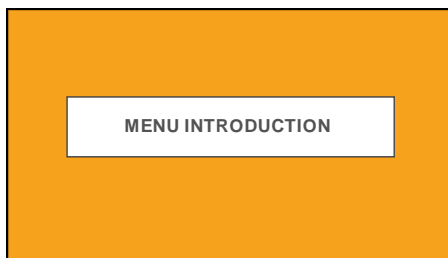


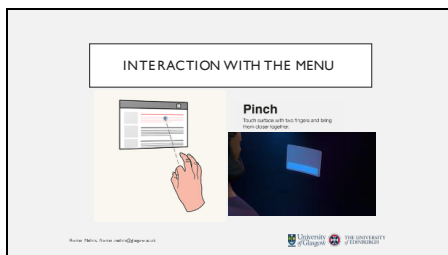




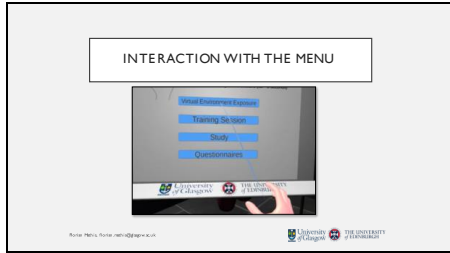
07.10.2022







07.10.2022







07.10.2022

SEMI-STRUCTURED INTERVIEW

Traditional 4-digit PIN Authentication
Smart ID card Authentication
Hand-based Authentication
Two-factor Authentication

Source: <https://www.northerntrust.com>

University of Glasgow
1792-1847
1847-1969
1969-2014
2014-Present

Semi-structured Interview Questions

1. Virtual Reality and Perceived Presence and Realism

- (a) Could you please walk us through the different authentication methods and tell us what differences may appear when using the methods in the real world rather than in VR as just experienced?
- (b) Do you think the virtual environment impacted your behaviour when providing input with the corresponding authentication method? If so, how?
- (c) Could you please tell us why (or why not) you felt being part of the environment where the authentication happened?
- (d) What (if any) is the difference between withdrawing cash at a real-world bank ATM and what you have just experienced?
- (e) How did the virtual environment (+ virtual bystanders) impact your ATM interaction behaviour?
- (f) Please think about your last ATM withdrawal in the real world. What was different to what you have just experienced?
- (g) Could you please tell us how realistic the ATM experience was for you? Please briefly justify your response.

2. Perceived Usability and Ranking of the Prototypes

- (a) Please justify the your ranking of the methods in terms of (a) usability, (b) security, and (c) usability + security.
- (b) Have you used such an authentication method previously in any other context?
- (c) Please tell us how you felt using this authentication method to withdraw cash on an ATM.
- (d) Please tell us (a) what you particularly liked, and (b) what you did not like when using this method to authenticate on an ATM.
- (e) (*only for Glass Unlock*) Did you constantly switch between the private near-eye display and the keypad on the ATM or rather stayed on either of them?

3. Perceived Security of the Prototypes

- (a) How secure do you think is this authentication method against observations where a bystanders observes your authentication?
- (b) Can you think of any attacks that could break the security of this authentication system?

- (c) Consider you want to attack a user's ATM authentication when using this method. How would you try to access their PIN?

4. Enhancements of the Prototypes

- (a) Is there anything in particular that you would like to improve in this authentication method?
- (b) Do you have any other ideas on how authentication in front of public displays like ATMs could look like?

5. Impact of the Real-world Environment and the Experimenter

- (a) Could you please describe your real-world surrounding and how it looks like? Please note that we do not expect a detailed description of your personal space, but it would be great if you could give a rough overview of the room you are currently in.
- (b) Could you please tell us to what extent the real-world surrounding impacted you while performing the authentications?
- (c) Could you please tell us how the experimenter on the Zoom call impacted you while performing the authentications?

Bibliography

- [1] U. 3D. (2021) User manual. <https://docs.unity3d.com/Manual/class-LineRenderer.html>, accessed 18 February 2023.
- [2] J. Abbate, “Getting small: a short history of the personal computer,” *Proceedings of the IEEE*, vol. 87, no. 9, pp. 1695–1698, 1999. [Online]. Available: <https://doi.org/10.1109/5.784256>
- [3] 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*, ser. CHI ’17. New York, NY, USA: Association for Computing Machinery, 2017, p. 3751–3763. [Online]. Available: <https://doi.org/10.1145/3025453.3025461>
- [4] Y. Abdelrahman, F. Mathis, P. Knierim, A. Kettler, F. Alt, and M. Khamis, “Cuevr: Studying the usability of cue-based authentication for virtual reality,” in *Proceedings of the 2022 International Conference on Advanced Visual Interfaces*, ser. AVI 2022. New York, NY, USA: Association for Computing Machinery, 2022. [Online]. Available: <https://doi.org/10.1145/3531073.3531092>
- [5] Y. Abdrabou, M. Khamis, R. M. Eisa, S. Ismael, and A. Elmougy, “Engage: Resisting shoulder surfing using novel gaze gestures authentication,” in *Proc. of the 17th International Conf. on Mobile and Ubiquitous Multimedia*. New York, NY, USA: ACM, 2018, p. 469–473. [Online]. Available: <https://doi.org/10.1145/3282894.3289741>
- [6] Y. Abdrabou, M. Khamis, R. M. 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*, ser. ETRA ’19. New York, NY, USA: Association for Computing Machinery, 2019. [Online]. Available: <https://doi.org/10.1145/3314111.3319837>
- [7] Y. Acar, M. Backes, S. Fahl, S. Garfinkel, D. Kim, M. L. Mazurek, and C. Stransky, “Comparing the usability of cryptographic apis,” in *2017 IEEE Symposium on*

- Security and Privacy (SP)*. IEEE, 2017, pp. 154–171. [Online]. Available: <https://doi.org/10.1109/SP.2017.52>
- [8] A. Adams and A. L. Cox, *Questionnaires, in-depth interviews and focus groups*. Cambridge University Press, 2008, <http://oro.open.ac.uk/11909/>, accessed 18 February 2023.
- [9] A. Adams and M. A. Sasse, “Users Are Not the Enemy,” *Communications of the ACM*, vol. 42, no. 12, p. 40–46, Dec. 1999. [Online]. Available: <https://doi.org/10.1145/322796.322806>
- [10] E. Adar, D. S. Tan, and J. Teevan, “Benevolent deception in human computer interaction,” in *Proceedings of Human Factors in Computing Systems*, ser. CHI '13. New York, NY, USA: ACM, 2013. [Online]. Available: <https://doi.org/10.1145/2470654.2466246>
- [11] Adobe. (2021) Mixamo: Animated 3d characters. <https://www.mixamo.com/>, accessed 18 February 2023.
- [12] S. I. Ahmed, M. R. Haque, S. Guha, M. R. Rifat, and N. Dell, “Privacy, security, and surveillance in the global south: A study of biometric mobile sim registration in bangladesh,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 906–918. [Online]. Available: <https://doi.org/10.1145/3025453.3025961>
- [13] D. Akhawe and A. P. Felt, “Alice in warningland: A {Large-Scale} field study of browser security warning effectiveness,” in *22nd USENIX Security Symposium (USENIX Security 13)*, 2013, pp. 257–272, https://www.usenix.org/system/files/conference/usenixsecurity13/sec13-paper_akhawe.pdf, accessed 18 February 2023.
- [14] S. Albakry, K. Vaniea, and M. K. Wolters, “What is this url’s destination? empirical evaluation of users’ url reading,” in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1–12. [Online]. Available: <https://doi.org/10.1145/3313831.3376168>
- [15] N. Alotaibi, J. Williamson, and M. Khamis, “Thermosecure: Investigating the effectiveness of ai-driven thermal attacks on commonly used computer keyboards,” *ACM Trans. Priv. Secur.*, sep 2022, just Accepted. [Online]. Available: <https://doi.org/10.1145/3563693>

- [16] F. Alt, “Out-of-the-lab research in usable security and privacy,” in *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*. New York, NY, USA: ACM, 2021, p. 363–365. [Online]. Available: <https://doi.org/10.1145/3450614.3464468>
- [17] F. Alt and E. von Zezschwitz, “Emerging trends in usable security and privacy,” *Journal of Interactive Media*, vol. 18, no. 3, pp. 189–195, 2019. [Online]. Available: <https://doi.org/10.1515/icom-2019-0019>
- [18] K. Althobaiti, K. Vaniea, and S. Zheng, “Faheem: Explaining urls to people using a slack bot,” in *Symposium on Digital Behaviour Intervention for Cyber Security*, ser. AISB ’18, April 2018, https://www.research.ed.ac.uk/files/57206667/Althobaiti_2018_Faheem.Explaining_URLs.pdf, accessed 18 February 2023.
- [19] T. Amano, S. Kajita, H. Yamaguchi, T. Higashino, and M. Takai, “Smartphone applications testbed using virtual reality,” in *Proceedings of the 15th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, ser. MobiQuitous ’18. New York, NY, USA: ACM, 2018, p. 422–431. [Online]. Available: <https://doi.org/10.1145/3286978.3287028>
- [20] Amazon. (2022) Amazon mechanical turk: Access a global, on-demand, 24x7 workforce. <https://www.mturk.com/>, accessed 18 February 2023.
- [21] P. Andrew, L. Chris, and S. Flinn, “Workshop on human-computer interaction and security systems,” in *CHI 2003*. Fort Lauderdale, Florida, USA: Association for Computing Machinery, April 2003, <http://www.andrewpatrick.ca/CHI2003/HCISEC/>, accessed 18 February 2023.
- [22] P. Andriotis, T. Tryfonas, and G. Oikonomou, “Complexity metrics and user strength perceptions of the pattern-lock graphical authentication method,” in *International conference on human aspects of information security, privacy, and trust*. Springer, 2014, pp. 115–126. [Online]. Available: https://doi.org/10.1007/978-3-319-07620-1_11
- [23] Apple. (2023) Apple vision pro reality headset. <https://www.apple.com/apple-vision-pro/>, accessed 22 June 2023.
- [24] M. P. Ashby and A. Thorpe, “Self-guardianship at automated teller machines,” *Crime Prevention and Community Safety*, vol. 19, no. 1, pp. 1–16, 2017. [Online]. Available: <https://doi.org/10.1057/s41300-016-0010-3>
- [25] I. Aslan, A. Uhl, A. Meschtscherjakov, and M. Tscheligi, “Mid-air authentication gestures: An exploration of authentication based on palm and finger motions,” in

- Proceedings of the 16th International Conference on Multimodal Interaction*, ser. ICMI '14. New York, NY, USA: ACM, 2014, p. 311–318. [Online]. Available: <https://doi.org/10.1145/2663204.2663246>
- [26] I. Aslan, A. Uhl, A. Meschtscherjakov, and M. Tscheligi, “Design and exploration of mid-air authentication gestures,” vol. 6, no. 3. New York, NY, USA: Association for Computing Machinery, sep 2016. [Online]. Available: <https://doi.org/10.1145/2832919>
- [27] A. J. Aviv, J. T. Davin, F. Wolf, and R. Kuber, “Towards baselines for shoulder surfing on mobile authentication,” in *Proceedings of the 33rd Annual Computer Security Applications Conference*, ser. ACSAC 2017. New York, NY, USA: ACM, 2017, p. 486–498. [Online]. Available: <https://doi.org/10.1145/3134600.3134609>
- [28] 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*, ser. WOOT'10. USA: USENIX Association, 2010, p. 1–7. [Online]. Available: <https://dl.acm.org/doi/10.5555/1925004.1925009>
- [29] A. J. Aviv, F. Wolf, and R. Kuber, “Comparing video based shoulder surfing with live simulation,” in *Proceedings of the 34th Annual Computer Security Applications Conference*, ser. ACSAC '18. New York, NY, USA: ACM, 2018, p. 453–466. [Online]. Available: <https://doi.org/10.1145/3274694.3274702>
- [30] J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis, “Equilibrium theory revisited: Mutual gaze and personal space in virtual environments,” *Presence*, vol. 10, no. 6, 2001. [Online]. Available: <https://doi.org/10.1162/105474601753272844>
- [31] M. Baker, “Over half of psychology studies fail reproducibility test,” *Nature*, vol. 27, pp. 1–3, 2015.
- [32] M. Baldauf, S. Steiner, M. Khamis, and S.-K. Thiel, “Investigating the user experience of smartphone authentication schemes-the role of the mobile context,” in *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019. [Online]. Available: <http://dx.doi.org/10.24251/HICSS.2019.579>
- [33] R. Balebako, J. Jung, W. Lu, L. F. Cranor, and C. Nguyen, ““little brothers watching you”: Raising awareness of data leaks on smartphones,” in *Proceedings of the Ninth Symposium on Usable Privacy and Security*, ser. SOUPS '13. New York, NY, USA: ACM, 2013. [Online]. Available: <https://doi.org/10.1145/2501604.2501616>
- [34] A. Bangor, P. Kortum, and J. Miller, “Determining what individual sus scores mean: Adding an adjective rating scale.” Citeseer, 2009, pp. 114–123, <https://uxpajournal.org/>

determining-what-individual-sus-scores-mean-adding-an-adjective-rating-scale/, accessed 18 February 2023.

- [35] L. Barkhuus and J. A. Rode, “From mice to men - 24 years of evaluation in chi,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '07. New York, NY, USA: ACM, 2007. [Online]. Available: <https://doi.org/10.1145/1240624.2180963>
- [36] H. Benko, C. Holz, M. Sinclair, and E. Ofek, “Normaltouch and texturetouch: High-fidelity 3d haptic shape rendering on handheld virtual reality controllers,” in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, ser. UIST '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 717–728. [Online]. Available: <https://doi.org/10.1145/2984511.2984526>
- [37] B. Berki, “Experiencing the sense of presence within an educational desktop virtual reality,” *Acta Polytechnica Hungarica*, vol. 17, no. 2, pp. 255–265, 2020. [Online]. Available: <http://dx.doi.org/10.12700/APH.17.2.2020.2.14>
- [38] R. Bhagavatula, B. Ur, K. Iacovino, S. M. Kywe, L. F. Cranor, and M. Savvides, “Biometric authentication on iphone and android: Usability, perceptions, and influences on adoption,” 2015, https://www.ndss-symposium.org/wp-content/uploads/2017/09/01_3_3.pdf, accessed 18 February 2023.
- [39] A. Bianchi, “Authentication on public terminals with private devices,” in *Proceedings of the Fifth International Conference on Tangible, Embedded, and Embodied Interaction*, ser. TEI '11. New York, NY, USA: ACM, 2010. [Online]. Available: <https://doi.org/10.1145/1935701.1935815>
- [40] A. Bianchi and I. Oakley, “Multiplexed input to protect against casual observers,” in *Proceedings of HCI Korea*, Seoul, KOR, 2014, <https://dl.acm.org/doi/10.5555/2729485.2729487>, accessed 18 February 2023.
- [41] A. Bianchi, I. Oakley, V. Kostakos, and D. S. Kwon, “The Phone Lock: Audio and Haptic Shoulder-Surfing Resistant PIN Entry Methods for Mobile Devices,” in *Proceedings of the Fifth International Conference on Tangible, Embedded, and Embodied Interaction*, ser. TEI '11. New York, NY, USA: ACM, 2010, p. 197–200. [Online]. Available: <https://doi.org/10.1145/1935701.1935740>
- [42] A. Bianchi, I. Oakley, and D. S. Kwon, “The secure haptic keypad: A tactile password system,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '10. New York, NY, USA: Association for Computing Machinery, 2010, p. 1089–1092. [Online]. Available: <https://doi.org/10.1145/1753326.1753488>

- [43] A. Bianchi, I. Oakley, and D. S. Kwon, "Spinlock: A single-cue haptic and audio pin input technique for authentication," in *Haptic and Audio Interaction Design*, E. W. Cooper, V. V. Kryssanov, H. Ogawa, and S. Brewster, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 81–90. [Online]. Available: https://doi.org/10.1007/978-3-642-22950-3_9
- [44] A. Bianchi, I. Oakley, and D. S. Kwon, "Using mobile device screens for authentication," in *Proceedings of the 23rd Australian Computer-Human Interaction Conference*, ser. OzCHI '11. New York, NY, USA: Association for Computing Machinery, 2011, p. 50–53. [Online]. Available: <https://doi.org/10.1145/2071536.2071542>
- [45] B. Biguer, M. Jeannerod, and C. Prablanc, "The coordination of eye, head, and arm movements during reaching at a single visual target," *Experimental Brain Research*, vol. 46, no. 2, pp. 301–304, May 1982. [Online]. Available: <https://doi.org/10.1007/BF00237188>
- [46] M. Bikos, Y. Itoh, G. Klinker, and K. Moustakas, "An interactive augmented reality chess game using bare-hand pinch gestures," in *2015 International Conference on Cyberworlds (CW)*, 2015, pp. 355–358. [Online]. Available: <https://doi.org/10.1109/CW.2015.15>
- [47] F. Biocca, C. Harms, and J. Gregg, "The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity," in *4th annual international workshop on presence, Philadelphia, PA*, 2001, pp. 1–9, <http://matthewlombard.com/ISPR/Proceedings/2001/Biocca2.pdf>, accessed 18 February 2023.
- [48] M. H. Blackmon, P. G. Polson, M. Kitajima, and C. Lewis, "Cognitive walkthrough for the web," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '02. New York, NY, USA: Association for Computing Machinery, 2002, p. 463–470. [Online]. Available: <https://doi.org/10.1145/503376.503459>
- [49] M. J. Blanca Mena, R. Alarcón Postigo, J. Arnau Gras, R. Bono Cabré, R. Bendayan *et al.*, "Non-normal data: Is anova still a valid option?" *Psicothema*, 2017. [Online]. Available: <https://doi.org/10.7334/psicothema2016.383>
- [50] J. Blascovich, J. Loomis, A. C. Beall, K. R. Swinth, C. L. Hoyt, and J. N. Bailenson, "Immersive virtual environment technology as a methodological tool for social psychology," *Psychological Inquiry*, vol. 13, no. 2, pp. 103–124, 2002. [Online]. Available: https://doi.org/10.1207/S15327965PLI1302_01
- [51] S. Bødker, "When second wave hci meets third wave challenges," in *Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles*, ser.

- NordiCHI '06. New York, NY, USA: Association for Computing Machinery, 2006, p. 1–8. [Online]. Available: <https://doi.org/10.1145/1182475.1182476>
- [52] C. J. Bohil, B. Alicea, and F. A. Biocca, “Virtual reality in neuroscience research and therapy,” *Nature reviews neuroscience*, vol. 12, no. 12, pp. 752–762, 2011. [Online]. Available: <https://doi.org/10.1038/nrn3122>
- [53] J. Bonneau, C. Herley, P. C. v. Oorschot, and F. Stajano, “The quest to replace passwords: A framework for comparative evaluation of web authentication schemes,” in *2012 IEEE Symposium on Security and Privacy*, 2012, pp. 553–567. [Online]. Available: <https://doi.org/10.1109/SP.2012.44>
- [54] M. Bostock, V. Ogievetsky, and J. Heer, “D3: Data-Driven Documents,” *IEEE Transactions on Visualization and Computer Graphics*, 2011, <http://vis.stanford.edu/papers/d3>, accessed 18 February 2023.
- [55] L. Bošnjak and B. Brumen, “Shoulder surfing: From an experimental study to a comparative framework,” *International Journal of Human-Computer Studies*, vol. 130, pp. 1–20, 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1071581918305366>
- [56] L. Bošnjak and B. Brumen, “Shoulder surfing experiments: A systematic literature review,” *Computers & Security*, 2020. [Online]. Available: <https://doi.org/10.1016/j.cose.2020.102023>
- [57] V. Braun and V. Clarke, *Thematic analysis*. American Psychological Association, 2012.
- [58] C. Bravo-Lillo, L. Cranor, J. Downs, S. Komanduri, S. Schechter, and M. Sleeper, “Operating system framed in case of mistaken identity: Measuring the success of web-based spoofing attacks on os password-entry dialogs,” in *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, ser. CCS '12. New York, NY, USA: Association for Computing Machinery, 2012, p. 365–377. [Online]. Available: <https://doi.org/10.1145/2382196.2382237>
- [59] H. Brignull and Y. Rogers, “Enticing people to interact with large public displays in public spaces.” in *Interact*, vol. 3, 2003, pp. 17–24, <https://rauterberg.employee.id.tue.nl/conferences/INTERACT2003/INTERACT2003-p17.pdf>, accessed 18 February 2023.
- [60] J. Brooke *et al.*, “Sus-a quick and dirty usability scale,” *Usability evaluation in industry*, pp. 4–7, 1996. [Online]. Available: <https://doi.org/10.1201/9781498710411-35>

- [61] 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*, ser. PerDis '14. New York, NY, USA: ACM, 2014, p. 1–6. [Online]. Available: <https://doi.org/10.1145/2611009.2611028>
- [62] F. Bruno and M. Muzzupappa, “Product interface design: A participatory approach based on virtual reality,” *International journal of human-computer studies*, vol. 68, no. 5, pp. 254–269, 2010. [Online]. Available: <https://doi.org/10.1016/j.ijhcs.2009.12.004>
- [63] D. Buschek, A. De Luca, and F. Alt, “Improving accuracy, applicability and usability of keystroke biometrics on mobile touchscreen devices,” in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 1393–1402. [Online]. Available: <https://doi.org/10.1145/2702123.2702252>
- [64] V. Bush. (1945) As we may think. <https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/>, accessed 18 February 2023.
- [65] F. Buttussi and L. Chittaro, “Effects of different types of virtual reality display on presence and learning in a safety training scenario,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 2, pp. 1063–1076, 2018. [Online]. Available: <https://doi.org/10.1109/TVCG.2017.2653117>
- [66] K. Caine, “Local standards for sample size at chi,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 981–992. [Online]. Available: <https://doi.org/10.1145/2858036.2858498>
- [67] J. Cambre, J. Colnago, J. Maddock, J. Tsai, and J. Kaye, “Choice of voices: A large-scale evaluation of text-to-speech voice quality for long-form content,” in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1–13. [Online]. Available: <https://doi.org/10.1145/3313831.3376789>
- [68] G. Canova, M. Volkamer, C. Bergmann, and R. Borza, “NoPhish: an anti-phishing education app,” in *International Workshop on Security and Trust Management*. Springer, 2014, p. 188–192. [Online]. Available: https://doi.org/10.1007/978-3-319-11851-2_14

- [69] J. M. Carroll, “Creating a design science of human-computer interaction,” *Interacting with computers*, vol. 5, no. 1, pp. 3–12, 1993. [Online]. Available: [https://doi.org/10.1016/0953-5438\(93\)90022-L](https://doi.org/10.1016/0953-5438(93)90022-L)
- [70] J. M. Carroll, “Human-computer interaction: psychology as a science of design,” *Annual review of psychology*, vol. 48, no. 1, pp. 61–83, 1997. [Online]. Available: <https://doi.org/10.1146/annurev.psych.48.1.61>
- [71] J. M. Carroll, “Conceptualizing a possible discipline of human–computer interaction,” *Interacting with Computers*, vol. 22, no. 1, pp. 3–12, 2010. [Online]. Available: <https://doi.org/10.1016/j.intcom.2009.11.008>
- [72] J. Casanueva and E. Blake, “The effects of avatars on co-presence in a collaborative virtual environment,” 2001, <http://hdl.handle.net/10500/24749>, accessed 18 February 2023.
- [73] N. C. S. Centre. (2020) Official statistics cyber security breaches survey 2020– chapter 5: Incidence and impact of breaches or attacks. [Online]. Available: <https://www.gov.uk/government/statistics/cyber-security-breaches-survey-2020/cyber-security-breaches-survey-2020>
- [74] E. Cheon, Y. Shin, J. Huh, H. Kim, and I. Oakley, “Gesture authentication for smartphones: Evaluation of gesture password selection policies,” in *2020 IEEE Symposium on Security and Privacy (SP)*. Los Alamitos, CA, USA: IEEE Computer Society, may 2020, pp. 327–345. [Online]. Available: <https://doi.org/10.1109/SP40000.2020.00034>
- [75] S. Chiasson, R. Biddle, and P. C. van Oorschot, “A second look at the usability of click-based graphical passwords,” in *Proceedings of the 3rd Symposium on Usable Privacy and Security*, ser. SOUPS '07. New York, NY, USA: Association for Computing Machinery, 2007, p. 1–12. [Online]. Available: <https://doi.org/10.1145/1280680.1280682>
- [76] I. Choi, E. Ofek, H. Benko, M. Sinclair, and C. Holz, *CLAW: A Multifunctional Handheld Haptic Controller for Grasping, Touching, and Triggering in Virtual Reality*. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–13. [Online]. Available: <https://doi.org/10.1145/3173574.3174228>
- [77] R. Ciesielski and M. Zierer. (2022) How biometric devices are putting afghans in danger. <https://interaktiv.br.de/biometrie-afghanistan/en/index.html>, accessed 18 February 2023.

- [78] A. Cockburn, C. Gutwin, and A. Dix, “Hark no more: On the preregistration of chi experiments,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–12. [Online]. Available: <https://doi.org/10.1145/3173574.3173715>
- [79] C. Coelho, J. Tichon, T. J. Hine, G. Wallis, and G. Riva, “Media presence and inner presence: the sense of presence in virtual reality technologies,” *From communication to presence: Cognition, emotions and culture towards the ultimate communicative experience*, vol. 11, pp. 25–45, 2006, <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.150.1784&rep=rep1&type=pdf>, accessed 18 February 2023.
- [80] O. S. Collaboration, “Estimating the reproducibility of psychological science,” *Science*, vol. 349, no. 6251, 2015. [Online]. Available: <https://doi.org/10.1126/science.aac4716>
- [81] M. Colley, P. Jansen, E. Rukzio, and J. Gugenheimer, “Swivr-car-seat: Exploring vehicle motion effects on interaction quality in virtual reality automated driving using a motorized swivel seat,” *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 5, no. 4, dec 2022. [Online]. Available: <https://doi.org/10.1145/3494968>
- [82] J. Colnago, S. Devlin, M. Oates, C. Swoopes, L. Bauer, L. Cranor, and N. Christin, ““it’s not actually that horrible”: Exploring adoption of two-factor authentication at a university,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–11. [Online]. Available: <https://doi.org/10.1145/3173574.3174030>
- [83] T. D. Cook, D. T. Campbell, and A. Day, *Quasi-experimentation: Design & analysis issues for field settings*. Houghton Mifflin Boston, 1979, vol. 351.
- [84] L. Coventry, A. De Angeli, and G. Johnson, “Usability and biometric verification at the atm interface,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’03. New York, NY, USA: Association for Computing Machinery, 2003, p. 153–160. [Online]. Available: <https://doi.org/10.1145/642611.642639>
- [85] L. F. Cranor and N. Buchler, “Better together: Usability and security go hand in hand,” *IEEE Security & Privacy*, vol. 12, no. 6, pp. 89–93, 2014. [Online]. Available: <https://doi.org/10.1109/MSP.2014.109>
- [86] J. J. Cummings and J. N. Bailenson, “How immersive is enough? a meta-analysis of the effect of immersive technology on user presence,” *Media Psychology*, vol. 19, no. 2, pp. 272–309, 2016. [Online]. Available: <https://doi.org/10.1080/15213269.2015.1015740>

- [87] N. Dahlbäck, A. Jönsson, and L. Ahrenberg, “Wizard of oz studies: Why and how,” in *Proceedings of the 1st International Conference on Intelligent User Interfaces*, ser. IUI ’93. New York, NY, USA: Association for Computing Machinery, 1993, p. 193–200. [Online]. Available: <https://doi.org/10.1145/169891.169968>
- [88] F. Dandurand, T. R. Shultz, and K. H. Onishi, “Comparing online and lab methods in a problem-solving experiment,” *Behavior research methods*, vol. 40, no. 2, pp. 428–434, 2008. [Online]. Available: <https://doi.org/10.3758/BRM.40.2.428>
- [89] H. Dang, L. Mecke, and D. Buschek, “Ganslider: How users control generative models for images using multiple sliders with and without feedforward information,” in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’22. New York, NY, USA: Association for Computing Machinery, 2022. [Online]. Available: <https://doi.org/10.1145/3491102.3502141>
- [90] S. Das, G. Laput, C. Harrison, and J. I. Hong, “Thumprint: Socially-Inclusive Local Group Authentication Through Shared Secret Knocks,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’17. New York, NY, USA: ACM, 2017, p. 3764–3774. [Online]. Available: <https://doi.org/10.1145/3025453.3025991>
- [91] S. Das, D. Lu, T. Lee, J. Lo, and J. I. Hong, “The memory palace: Exploring visual-spatial paths for strong, memorable, infrequent authentication,” in *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, ser. UIST ’19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1109–1121. [Online]. Available: <https://doi.org/10.1145/3332165.3347917>
- [92] B. David and F. Myron, “Eye gaze tracking using an active stereo head,” in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, vol. 2. USA: IEEE Computer Society, 2003, pp. II–451. [Online]. Available: <https://doi.org/10.1109/CVPR.2003.1211502>
- [93] J. T. Davin, A. J. Aviv, F. Wolf, and R. Kuber, “Baseline measurements of shoulder surfing analysis and comparability for smartphone unlock authentication,” in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA ’17. New York, NY, USA: ACM, 2017, pp. 2496–2503. [Online]. Available: <https://doi.org/10.1145/3027063.3053221>
- [94] A. De Luca, “Designing usable and secure authentication mechanisms for public spaces,” Ph.D. dissertation, lmu, 2011, https://edoc.ub.uni-muenchen.de/13155/1/De_Luca_Alexander.pdf, accessed 18 February 2023.

- [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*, ser. SOUPS '09. New York, NY, USA: Association for Computing Machinery, 2009. [Online]. Available: <https://doi.org/10.1145/1572532.1572542>
- [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*, ser. CHI '12. New York, NY, USA: Association for Computing Machinery, 2012, p. 987–996. [Online]. Available: <https://doi.org/10.1145/2207676.2208544>
- [97] A. De Luca, A. Hang, E. von Zezschwitz, and H. Hussmann, “I feel like i’m taking selfies all day! towards understanding biometric authentication on smartphones,” in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 1411–1414. [Online]. Available: <https://doi.org/10.1145/2702123.2702141>
- [98] A. De Luca, M. Harbach, E. von Zezschwitz, M.-E. Maurer, B. E. Slawik, H. Hussmann, and M. Smith, “Now you see me, now you don’t: Protecting smartphone authentication from shoulder surfers,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 2937–2946. [Online]. Available: <https://doi.org/10.1145/2556288.2557097>
- [99] A. De Luca, K. Hertzschuch, and H. Hussmann, “Colorpin: Securing pin entry through indirect input,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI. New York, NY, USA: ACM, 2010, p. 1103–1106. [Online]. Available: <https://doi.org/10.1145/1753326.1753490>
- [100] A. De Luca, M. Langheinrich, and H. Hussmann, “Towards understanding atm security: A field study of real world atm use,” in *Proc. of the Sixth Symp. on Usable Privacy and Security*, ser. SOUPS '10. New York, NY, USA: ACM, 2010. [Online]. Available: <https://doi.org/10.1145/1837110.1837131>
- [101] 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*, ser. CHI '09. New York, NY, USA: Association for Computing Machinery, 2009, p. 913–916. [Online]. Available: <https://doi.org/10.1145/1518701.1518840>
- [102] A. De Luca, E. von Zezschwitz, N. D. H. Nguyen, M.-E. Maurer, E. Rubegni, M. P. Scipioni, and M. Langheinrich, “Back-of-device authentication on smartphones,” in

- Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 2389–2398. [Online]. Available: <https://doi.org/10.1145/2470654.2481330>
- [103] A. De Luca, E. von Zezschwitz, L. Pichler, and H. Hussmann, “Using fake cursors to secure on-screen password entry,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 2399–2402. [Online]. Available: <https://doi.org/10.1145/2470654.2481331>
- [104] 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*, ser. OZCHI '07. New York, NY, USA: Association for Computing Machinery, 2007, p. 199–202. [Online]. Available: <https://doi.org/10.1145/1324892.1324932>
- [105] N. Dell, V. Vaidyanathan, I. Medhi, E. Cutrell, and W. Thies, ““yours is better!”: Participant response bias in hci,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '12. New York, NY, USA: ACM, 2012, p. 1321–1330. [Online]. Available: <https://doi.org/10.1145/2207676.2208589>
- [106] O. Developers. Oculus integration sdk. <https://developer.oculus.com/downloads/package/unity-integration/>, accessed 18 February 2023.
- [107] T. Deyle and V. Roth, “Accessible authentication via tactile pin entry,” *Computer Graphics Topics*, vol. 2, pp. 24–26, 2006, http://www.travisdeyle.com/publications/pdf/2006_cgtopics_tactile_pin_entry.pdf, accessed 18 February 2023.
- [108] G. Dhandapani, J. Ferguson, and E. Freeman, “Hapticlock: Eyes-free authentication for mobile devices,” in *Proceedings of the 2021 International Conference on Multimodal Interaction*, ser. ICMI '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 195–202. [Online]. Available: <https://doi.org/10.1145/3462244.3481001>
- [109] M. Di Luca, H. Seifi, S. Egan, and M. Gonzalez-Franco, “Locomotion vault: The extra mile in analyzing vr locomotion techniques,” in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI '21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: <https://doi.org/10.1145/3411764.3445319>
- [110] V. Distler, M. Fassl, H. Habib, K. Krombholz, G. Lenzini, C. Lallemand, L. F. Cranor, and V. Koenig, “A systematic literature review of empirical methods and

- risk representation in usable privacy and security research,” *ACM Transactions on Computer-Human Interaction*, vol. 28, no. 6, dec 2021. [Online]. Available: <https://doi.org/10.1145/3469845>
- [111] A. Dix, J. Finlay, G. D. Abowd, and R. Beale, *Human-computer interaction*. Pearson Education, 2003, <https://hcibook.com/>, accessed 18 February 2023.
- [112] S. Djamasbi, D. F. Galletta, F. F.-H. Nah, X. Page, L. P. Robert Jr., and P. J. Wisniewski, “Bridging a bridge: Bringing two hci communities together,” in *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–8. [Online]. Available: <https://doi.org/10.1145/3170427.3170612>
- [113] Z. Doffman. (2019) New data breach has exposed millions of fingerprint and facial recognition records: Report. <https://www.forbes.com/sites/zakdoffman/2019/08/14/new-data-breach-has-exposed-millions-of-fingerprint-and-facial-recognition-records-report/>, accessed 18 February 2023.
- [114] L. Dole and W. Ju, “Face and ecological validity in simulations: Lessons from search-and-rescue hri,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–8. [Online]. Available: <https://doi.org/10.1145/3290605.3300681>
- [115] B. Dosono, J. Hayes, and Y. Wang, “Toward accessible authentication: Learning from people with visual impairments,” *IEEE Internet Computing*, vol. 22, no. 2, pp. 62–70, 2018. [Online]. Available: <https://doi.org/10.1109/MIC.2018.112101619>
- [116] P. Dourish, “What we talk about when we talk about context,” *Personal and ubiquitous computing*, vol. 8, no. 1, pp. 19–30, 2004. [Online]. Available: <https://doi.org/10.1007/s00779-003-0253-8>
- [117] J. S. Downs, M. B. Holbrook, and L. F. Cranor, “Decision strategies and susceptibility to phishing,” in *Proceedings of the Second Symposium on Usable Privacy and Security*, ser. SOUPS '06. New York, NY, USA: Association for Computing Machinery, 2006, p. 79–90. [Online]. Available: <https://doi.org/10.1145/1143120.1143131>
- [118] P. Dunphy, A. Fitch, and P. Olivier, “Gaze-contingent passwords at the atm,” in *In 4th Conference on Communication by Gaze Interaction (COGAIN)*, 2008.
- [119] P. Dunphy, A. P. Heiner, and N. Asokan, “A closer look at recognition-based graphical passwords on mobile devices,” in *Proceedings of the Sixth*

- Symposium on Usable Privacy and Security*, ser. SOUPS '10. New York, NY, USA: Association for Computing Machinery, 2010. [Online]. Available: <https://doi.org/10.1145/1837110.1837114>
- [120] N. Durlach, A. S. Mavor, and G. B. Newby, "Virtual reality: Scientific and technological challenges," *Library and Information Science Research*, vol. 18, no. 3, pp. 278–280, 1996. [Online]. Available: <https://doi.org/10.1177/135485659600200111>
- [121] S. Egelman, J. King, R. C. Miller, N. Ragouzis, and E. Shehan, "Security user studies: Methodologies and best practices," in *CHI '07 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '07. New York, NY, USA: Association for Computing Machinery, 2007, p. 2833–2836. [Online]. Available: <https://doi.org/10.1145/1240866.1241089>
- [122] S. Egelman and E. Peer, "Scaling the security wall: Developing a security behavior intentions scale (sebis)," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2015, p. 2873–2882. [Online]. Available: <https://doi.org/10.1145/2702123.2702249>
- [123] P. Eghbali, K. Väänänen, and T. Jokela, "Social acceptability of virtual reality in public spaces: Experiential factors and design recommendations," in *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*, ser. MUM '19. New York, NY, USA: Association for Computing Machinery, 2019. [Online]. Available: <https://doi.org/10.1145/3365610.3365647>
- [124] egosimx. (2021) 3d smartphone model. <https://free3d.com/3d-model/iphonex-113534.html>, accessed 18 February 2023.
- [125] 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 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI '17. New York, NY, USA: ACM, 2017, p. 4254–4265. [Online]. Available: <https://doi.org/10.1145/3025453.3025636>
- [126] L. A. Elkin, M. Kay, J. J. Higgins, and J. O. Wobbrock, "An aligned rank transform procedure for multifactor contrast tests," New York, NY, USA, p. 754–768, 2021. [Online]. Available: <https://doi.org/10.1145/3472749.3474784>
- [127] C. Elsdén, D. Chatting, A. C. Durrant, A. Garbett, B. Nissen, J. Vines, and D. S. Kirk, "On speculative enactments," in *Proceedings of the 2017 CHI conference on human factors in computing systems*, 2017, pp. 5386–5399. [Online]. Available: <https://doi.org/10.1145/3025453.3025503>

- [128] D. C. Engelbart and W. K. English, "A research center for augmenting human intellect," in *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, 1968, pp. 395–410. [Online]. Available: <https://doi.org/10.1145/1476589.1476645>
- [129] W. K. English, D. C. Engelbart, and M. L. Berman, "Display-selection techniques for text manipulation," *IEEE Transactions on Human Factors in Electronics*, no. 1, pp. 5–15, 1967. [Online]. Available: <https://doi.org/10.1109/THFE.1967.232994>
- [130] S. Eroglu, F. Stefan, A. Chevalier, D. Roettger, D. Zielasko, T. W. Kuhlen, and B. Weyers, "Design and evaluation of a free-hand vr-based authoring environment for automated vehicle testing," in *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, 2021, pp. 1–10. [Online]. Available: <https://doi.org/10.1109/VR50410.2021.00020>
- [131] U. Erra, D. Malandrino, and L. Pepe, "Virtual reality interfaces for interacting with three-dimensional graphs," *International Journal of Human-Computer Interaction*, vol. 35, no. 1, pp. 75–88, 2019. [Online]. Available: <https://doi.org/10.1080/10447318.2018.1429061>
- [132] K. M. Everitt, T. Bragin, J. Fogarty, and T. Kohno, "A comprehensive study of frequency, interference, and training of multiple graphical passwords," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '09. New York, NY, USA: Association for Computing Machinery, 2009, p. 889–898. [Online]. Available: <https://doi.org/10.1145/1518701.1518837>
- [133] S. Fahl, M. Harbach, Y. Acar, and M. Smith, "On the ecological validity of a password study," in *Proceedings of the Ninth Symposium on Usable Privacy and Security*, ser. SOUPS '13. New York, NY, USA: Association for Computing Machinery, 2013. [Online]. Available: <https://doi.org/10.1145/2501604.2501617>
- [134] D. Fallman, "Design-oriented human-computer interaction," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '03. New York, NY, USA: Association for Computing Machinery, 2003, p. 225–232. [Online]. Available: <https://doi.org/10.1145/642611.642652>
- [135] M. Feick, N. Kleer, A. Tang, and A. Krüger, "The virtual reality questionnaire toolkit," ser. UIST Adjunct. New York, NY, USA: ACM, 2020. [Online]. Available: <https://doi.org/10.1145/3379350.3416188>
- [136] A. P. Felt, R. W. Reeder, H. Almuhimedi, and S. Consolvo, "Experimenting at Scale with Google Chrome's SSL Warning," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '14. New York, NY, USA: ACM, 2014, p. 2667–2670. [Online]. Available: <https://doi.org/10.1145/2556288.2557292>

- [137] I. Fléchais and S. Faily, “Security and usability: Searching for the philosopher’s stone,” 2010. [Online]. Available: <https://rgu-repository.worktribe.com/output/1545167>
- [138] D. Florencio and C. Herley, “A large-scale study of web password habits,” in *Proceedings of the 16th International Conference on World Wide Web*, ser. WWW ’07. New York, NY, USA: Association for Computing Machinery, 2007, p. 657–666. [Online]. Available: <https://doi.org/10.1145/1242572.1242661>
- [139] Florian Mathis, K. Vaniea, and M. Khamis, “Replicueauth: Validating the use of a lab-based virtual reality setup for evaluating authentication systems.” in *Proceedings of the 39th Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI ’21. New York, NY, USA: ACM, 2021. [Online]. Available: <https://doi.org/10.1145/3411764.3445478>
- [140] A. Forget, S. Chiasson, P. C. van Oorschot, and R. Biddle, “Improving text passwords through persuasion,” in *Proceedings of the 4th Symposium on Usable Privacy and Security*, ser. SOUPS ’08. New York, NY, USA: Association for Computing Machinery, 2008, p. 1–12. [Online]. Available: <https://doi.org/10.1145/1408664.1408666>
- [141] T. Forman and A. Aviv, “Double patterns: A usable solution to increase the security of android unlock patterns,” in *Annual Computer Security Applications Conference*, ser. ACSAC ’20. New York, NY, USA: Association for Computing Machinery, 2020, p. 219–233. [Online]. Available: <https://doi.org/10.1145/3427228.3427252>
- [142] T. Franke, C. Attig, and D. Wessel, “A personal resource for technology interaction: Development and validation of the affinity for technology interaction (ati) scale,” *International Journal of Human–Computer Interaction*, vol. 35, no. 6, pp. 456–467, 2019. [Online]. Available: <https://doi.org/10.1080/10447318.2018.1456150>
- [143] Free3D. (2019) Free3d: 3d atm model. <https://free3d.com/3d-model/atm-57251.html>, accessed 18 February 2023.
- [144] D. Freeman, N. Evans, R. Lister, A. Antley, G. Dunn, and M. Slater, “Height, social comparison, and paranoia: An immersive virtual reality experimental study,” *Psychiatry research*, vol. 218, no. 3, pp. 348–352, 2014. [Online]. Available: <https://doi.org/10.1016/j.psychres.2013.12.014>
- [145] E. Freeman, S. Brewster, and V. Lantz, “Tactile feedback for above-device gesture interfaces: Adding touch to touchless interactions,” in *Proceedings of the 16th International Conference on Multimodal Interaction*, ser. ICMI ’14. New York, NY, USA: Association for Computing Machinery, 2014, p. 419–426. [Online]. Available: <https://doi.org/10.1145/2663204.2663280>

- [146] E. Freeman, S. Brewster, and V. Lantz, “Do that, there: An interaction technique for addressing in-air gesture systems,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 2319–2331. [Online]. Available: <https://doi.org/10.1145/2858036.2858308>
- [147] L. Freina and M. Ott, “A literature review on immersive virtual reality in education: state of the art and perspectives,” in *The international scientific Conf. elearning and software for education*, vol. 1, 2015, https://www.researchgate.net/publication/280566372_A_Literature_Review_on_Immersive_Virtual_Reality_in_Education_State_Of_The_Art_and_Perspectives, accessed 18 February 2023.
- [148] S. J. Friston, B. J. Congdon, D. Swapp, L. Izzouzi, K. Brandstätter, D. Archer, O. Olkkonen, F. J. Thiel, and A. Steed, “Ubiq: A system to build flexible social virtual reality experiences,” in *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*, ser. VRST '21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: <https://doi.org/10.1145/3489849.3489871>
- [149] S. G. Frydenberg and K. Nordby, “Virtual fieldwork on a ship’s bridge: virtual reality-reconstructed operation scenarios as contextual substitutes for fieldwork in design education,” *Virtual Reality*, pp. 1–12, 2022. [Online]. Available: <https://doi.org/10.1007/s10055-022-00655-1>
- [150] M. Funk, K. Marky, I. Mizutani, M. Kritzler, S. Mayer, and F. Michahelles, “Lookunlock: Using spatial-targets for user-authentication on hmds,” in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '19. New York, NY, USA: ACM, 2019, pp. LBW0114:1–LBW0114:6. [Online]. Available: <http://doi.acm.org/10.1145/3290607.3312959>
- [151] A. Gamero-Garrido, S. Savage, K. Levchenko, and A. C. Snoeren, “Quantifying the pressure of legal risks on third-party vulnerability research,” in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 1501–1513. [Online]. Available: <https://doi.org/10.1145/3133956.3134047>
- [152] M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogni, A. Steed, and M. A. Sasse, “The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '03. New York, NY, USA: ACM, 2003, p. 529–536. [Online]. Available: <https://doi.org/10.1145/642611.642703>

- [153] S. Garcia, R. Kauer, D. Laesker, J. Nguyen, and M. Andujar, "A virtual reality experience for learning languages," in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–4. [Online]. Available: <https://doi.org/10.1145/3290607.3313253>
- [154] S. Garfinkel, "Design principles and patterns for computer systems that are simultaneously secure and usable," Ph.D. dissertation, Massachusetts Institute of Technology, 2005, <http://hdl.handle.net/1721.1/33204>, accessed 18 February 2023.
- [155] S. Garfinkel and H. R. Lipford, "Usable security: History, themes, and challenges," *Synthesis Lectures on Information Security, Privacy, and Trust*, vol. 5, no. 2, pp. 1–124, 2014. [Online]. Available: <https://www.morganclaypool.com/doi/abs/10.2200/S00594ED1V01Y201408SPT011>
- [156] C. George, D. Buschek, A. Ngao, and M. Khamis, "Gazeroomlock: Using gaze and head-pose to improve the usability and observationresistance of 3d passwords in virtual reality," in *Augmented Reality, Virtual Reality, and Computer Graphics*. Springer International Publishing, 2020. [Online]. Available: https://doi.org/10.1007/978-3-030-58465-8_5
- [157] 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)*, March 2019, pp. 277–285. [Online]. Available: <https://doi.org/10.1109/VR.2019.8797862>
- [158] C. George, M. Khamis, E. von Zezschwitz, M. Burger, H. Schmidt, F. Alt, and H. Hussmann, "Seamless and secure vr: Adapting and evaluating established authentication systems for virtual reality," in *Network and Distributed System Security Symposium (NDSS 2017)*, ser. USEC '17. NDSS, February 2017. [Online]. Available: <http://dx.doi.org/10.14722/usec.2017.23028>
- [159] J. Gideon, L. Cranor, S. Egelman, and A. Acquisti, "Power strips, prophylactics, and privacy, oh my!" in *Proceedings of the Second Symposium on Usable Privacy and Security*, ser. SOUPS '06. New York, NY, USA: Association for Computing Machinery, 2006, p. 133–144. [Online]. Available: <https://doi.org/10.1145/1143120.1143137>
- [160] L. M. Given, *The Sage encyclopedia of qualitative research methods*. Sage publications, 2008. [Online]. Available: <http://dx.doi.org/10.4135/9781412963909.n381>

- [161] G. V. Glass, P. D. Peckham, and J. R. Sanders, "Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance," *Review of Educational Research*, vol. 42, no. 3, pp. 237–288, 1972. [Online]. Available: <https://doi.org/10.3102/00346543042003237>
- [162] E. Glassman, P. Guo, D. Jackson, D. Karger, J. Kim, R. Miller, C. Sims, and H. Zhang. (2016) User interface design & implementation. <http://web.mit.edu/6.813/www/sp16/classes/11-experiment-design/>, accessed 18 February 2023.
- [163] D. Goedicke, J. Li, V. Evers, and W. Ju, "Vr-oom: Virtual reality on-road driving simulation," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2018, p. 1–11. [Online]. Available: <https://doi.org/10.1145/3173574.3173739>
- [164] L. Gong, M. A. Lomas, R. M. Needham, and J. H. Saltzer, "Protecting poorly chosen secrets from guessing attacks," *IEEE Journal on Selected Areas in Communications*, vol. 11, no. 5, pp. 648–656, June 1993. [Online]. Available: <https://doi.org/10.1109/49.223865>
- [165] M. Gonzalez-Franco, B. Cohn, E. Ofek, D. Burin, and A. Maselli, "The self-avatar follower effect in virtual reality," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2020, pp. 18–25. [Online]. Available: <https://doi.org/10.1109/VR46266.2020.00019>
- [166] M. Gonzalez-Franco, Z. Egan, M. Peachey, A. Antley, T. Randhavane, P. Panda, Y. Zhang, C. Y. Wang, D. F. Reilly, T. C. Peck, A. S. Won, A. Steed, and E. Ofek, "Movebox: Democratizing mocap for the microsoft rocketbox avatar library," in *2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, 2020, pp. 91–98. [Online]. Available: <https://doi.org/10.1109/AIVR50618.2020.00026>
- [167] M. Gonzalez-Franco, E. Ofek, C. Holz, A. Steed, J. Lanier, B. Buxton, K. Hinckley, and M. Sinclair, "Taxonomy of hand-object haptics for virtual reality," 2022. [Online]. Available: <https://doi.org/10.36227/techrxiv.20182229.v1>
- [168] M. Gonzalez Franco, E. Ofek, Y. Pan, A. Antley, A. Steed, B. Spanlang, A. Maselli, D. Banakou, N. Pelechano, S. Orts-Escolano, V. Orvalho, L. Trutoiu, M. Wojcik, M. V. Sanchez-Vives, J. Bailenson, M. Slater, and J. Lanier, "The rocketbox library and the utility of freely available rigged avatars," *Frontiers in Virtual Reality*, November 2020. [Online]. Available: <https://doi.org/10.3389/frvir.2020.561558>

- [169] M. Gonzalez-Franco, M. Wojcik, E. Ofek, A. Steed, and D. Garagan, “Microsoft rocketbox,” 2020, <https://github.com/microsoft/Microsoft-Rocketbox>, accessed 18 February 2023.
- [170] M. González-Franco, D. Pérez-Marcos, B. Spanlang, and M. Slater, “The contribution of real-time mirror reflections of motor actions on virtual body ownership in an immersive virtual environment,” in *2010 IEEE Virtual Reality Conference (VR)*, 2010, pp. 111–114. [Online]. Available: <https://doi.org/10.1109/VR.2010.5444805>
- [171] L. Götz, R. Rivu, F. Alt, A. Schmidt, and V. Mäkelä, “Real-world methods of autobiographical recall in virtual reality,” 2022. [Online]. Available: <https://doi.org/10.1145/3546155.3546704>
- [172] P. Green and L. Wei-Haas, “The rapid development of user interfaces: Experience with the wizard of oz method,” in *Proceedings of the Human Factors Society Annual Meeting*, vol. 29, no. 5. SAGE Publications Sage CA: Los Angeles, CA, 1985, pp. 470–474. [Online]. Available: <https://doi.org/10.1177/154193128502900515>
- [173] S. Greenberg and B. Buxton, “Usability evaluation considered harmful (some of the time),” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’08. New York, NY, USA: Association for Computing Machinery, 2008, p. 111–120. [Online]. Available: <https://doi.org/10.1145/1357054.1357074>
- [174] S. Greenberg and C. Fitchett, “Phidgets: Easy development of physical interfaces through physical widgets,” in *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST ’01. New York, NY, USA: ACM, 2001, p. 209–218. [Online]. Available: <https://doi.org/10.1145/502348.502388>
- [175] S. W. Greenwald, Z. Wang, M. Funk, and P. Maes, “Investigating social presence and communication with embodied avatars in room-scale virtual reality,” in *International Conference on Immersive Learning*. Springer, 2017, pp. 75–90. [Online]. Available: https://doi.org/10.1007/978-3-319-60633-0_7
- [176] M. Guerar, M. Benmohammed, and V. Alimi, “Color wheel pin: Usable and resilient atm authentication,” *Journal of High Speed Networks*, vol. 22, no. 3, pp. 231–240, 2016. [Online]. Available: <https://doi.org/10.3233/JHS-160545>
- [177] M. Guerar, L. Verderame, A. Merlo, F. Palmieri, M. Migliardi, and L. Vallerini, “Circlepin: A novel authentication mechanism for smartwatches to prevent unauthorized access to iot devices,” *ACM Trans. Cyber-Phys. Syst.*, vol. 4, no. 3, mar 2020. [Online]. Available: <https://doi.org/10.1145/3365995>

- [178] G. Guest, A. Bunce, and L. Johnson, “How many interviews are enough? an experiment with data saturation and variability,” *Field methods*, vol. 18, no. 1, pp. 59–82, 2006. [Online]. Available: <https://doi.org/10.1177/1525822X05279903>
- [179] G. Guest, E. Namey, and K. McKenna, “How many focus groups are enough? building an evidence base for nonprobability sample sizes,” *Field methods*, vol. 29, no. 1, pp. 3–22, 2017. [Online]. Available: <https://doi.org/10.1177/1525822X16639015>
- [180] J. Gugenheimer, A. De Luca, H. Hess, S. Karg, D. Wolf, and E. Rukzio, “ColorSnakes: Using colored decoys to secure authentication in sensitive contexts,” in *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, ser. MobileHCI '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 274–283. [Online]. Available: <https://doi.org/10.1145/2785830.2785834>
- [181] J. Gugenheimer, C. Mai, M. McGill, J. Williamson, F. Steinicke, and K. Perlin, “Challenges using head-mounted displays in shared and social spaces,” in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–8. [Online]. Available: <https://doi.org/10.1145/3290607.3299028>
- [182] J. Gugenheimer, E. Stemasov, J. Frommel, and E. Rukzio, “Sharevr: Enabling co-located experiences for virtual reality between hmd and non-hmd users,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2017, p. 4021–4033. [Online]. Available: <https://doi.org/10.1145/3025453.3025683>
- [183] M. Gutfleisch, M. Peiffer, S. Erk, and M. A. Sasse, *Microsoft Office Macro Warnings: A Design Comedy of Errors with Tragic Security Consequences*. New York, NY, USA: Association for Computing Machinery, 2021, p. 9–22. [Online]. Available: <https://doi.org/10.1145/3481357.3481512>
- [184] S. Guttinger, “The limits of replicability,” *European Journal for Philosophy of Science*, vol. 10, no. 2, p. 10, 2020. [Online]. Available: <https://doi.org/10.1007/s13194-019-0269-1>
- [185] V. Ha, K. Inkpen, F. Al Shaar, and L. Hdeib, “An examination of user perception and misconception of internet cookies,” in *CHI '06 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '06. New York, NY, USA: Association for Computing Machinery, 2006, p. 833–838. [Online]. Available: <https://doi.org/10.1145/1125451.1125615>

- [186] E. T. Hall, *The hidden dimension*. Garden City, NY: Doubleday, 1966, https://books.google.co.uk/books/about/The_Hidden_Dimension.html?id=zGYPwLj2dCoC&redir_esc=y, accessed 18 February 2023.
- [187] O. Hamdy and I. Traoré, “Homogeneous physio-behavioral visual and mouse-based biometric,” *ACM Trans. Comput.-Hum. Interact.*, vol. 18, no. 3, aug 2011. [Online]. Available: <https://doi.org/10.1145/1993060.1993062>
- [188] J. Han, A. V. Moere, and A. L. Simeone, “Foldable spaces: An overt redirection approach for natural walking in virtual reality,” in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2022, pp. 167–175. [Online]. Available: <https://doi.org/10.1109/VR51125.2022.00035>
- [189] 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*, ser. CHI ’16. New York, NY, USA: ACM, 2016, p. 4806–4817. [Online]. Available: <https://doi.org/10.1145/2858036.2858267>
- [190] M. Harbach, A. De Luca, N. Malkin, and S. Egelman, “Keep on lockin’ in the free world: A multi-national comparison of smartphone locking,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’16. New York, NY, USA: ACM, 2016, p. 4823–4827. [Online]. Available: <https://doi.org/10.1145/2858036.2858273>
- [191] M. Harbach, E. von Zezschwitz, A. Fichtner, A. D. Luca, and M. Smith, “It’s a hard lock life: A field study of smartphone (Un)Locking behavior and risk perception,” in *10th Symposium On Usable Privacy and Security (SOUPS 2014)*. Menlo Park, CA: USENIX Association, Jul. 2014, pp. 213–230, <https://www.usenix.org/conference/soups2014/proceedings/presentation/harbach>, accessed 18 February 2023.
- [192] J. Hardy, E. Rukzio, and N. Davies, “Real world responses to interactive gesture based public displays,” in *Proceedings of the 10th International Conference on Mobile and Ubiquitous Multimedia*, ser. MUM ’11. New York, NY, USA: Association for Computing Machinery, 2011, p. 33–39. [Online]. Available: <https://doi.org/10.1145/2107596.2107600>
- [193] S. G. Hart, “Nasa-task load index (nasa-tlx); 20 years later,” in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, no. 9. Sage publications Sage CA: Los Angeles, CA, 2006, pp. 904–908. [Online]. Available: <https://doi.org/10.1177/154193120605000909>

- [194] S. Hart and L. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," in *Human mental workload*, <http://humanfactors.arc.nasa.gov/groups/TLX/downloads/NASA-TLXChapter.pdf>, accessed 18 February 2023.
- [195] E. Hayashi, S. Das, S. Amini, J. Hong, and I. Oakley, "Casa: Context-aware scalable authentication," in *Proceedings of the Ninth Symposium on Usable Privacy and Security*, ser. SOUPS '13. New York, NY, USA: Association for Computing Machinery, 2013. [Online]. Available: <https://doi.org/10.1145/2501604.2501607>
- [196] E. Hayashi, R. Dhamija, N. Christin, and A. Perrig, "Use your illusion: Secure authentication usable anywhere," in *Proceedings of the 4th Symposium on Usable Privacy and Security*, ser. SOUPS '08. New York, NY, USA: Association for Computing Machinery, 2008, p. 35–45. [Online]. Available: <https://doi.org/10.1145/1408664.1408670>
- [197] H. Hecht, R. Welsch, J. Viehoff, and M. R. Longo, "The shape of personal space," *Acta Psychologica*, vol. 193, pp. 113–122, 2019. [Online]. Available: <https://doi.org/10.1016/j.actpsy.2018.12.009>
- [198] P. Heidicker, E. Langbehn, and F. Steinicke, "Influence of avatar appearance on presence in social vr," in *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, 01 2017, pp. 233–234. [Online]. Available: <https://doi.org/10.1109/3DUI.2017.7893357>
- [199] M. L. Heilig, "Sensorama simulator," *US PAT.* 3,050,870, 1962, <https://patents.google.com/patent/US3050870A/en>, accessed 18 February 2023.
- [200] N. Henze, M. Pielot, B. Poppinga, T. Schinke, and S. Boll, "My app is an experiment: Experience from user studies in mobile app stores," *International Journal of Mobile Human Computer Interaction (IJMHCI)*, vol. 3, no. 4, pp. 71–91, 2011. [Online]. Available: <http://dx.doi.org/10.4018/jmhci.2011100105>
- [201] N. Henze, E. Rukzio, and S. Boll, "100,000,000 taps: Analysis and improvement of touch performance in the large," in *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, ser. MobileHCI '11. New York, NY, USA: Association for Computing Machinery, 2011, p. 133–142. [Online]. Available: <https://doi.org/10.1145/2037373.2037395>
- [202] N. Henze, E. Rukzio, and S. Boll, "Observational and experimental investigation of typing behaviour using virtual keyboards for mobile devices," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '12. New York, NY, USA: Association for Computing Machinery, 2012, p. 2659–2668. [Online]. Available: <https://doi.org/10.1145/2207676.2208658>

- [203] C. Herley and P. V. Oorschot, “A research agenda acknowledging the persistence of passwords,” *IEEE Security Privacy*, vol. 10, no. 1, pp. 28–36, 2012. [Online]. Available: <https://doi.org/10.1109/MSP.2011.150>
- [204] J. L. Hintze and R. D. Nelson, “Violin plots: A box plot-density trace synergism,” *The American Statistician*, vol. 52, no. 2, pp. 181–184, 1998. [Online]. Available: <https://www.tandfonline.com/doi/abs/10.1080/00031305.1998.10480559>
- [205] L. Hodges, P. Anderson, G. Burdea, H. Hoffmann, and B. Rothbaum, “Treating psychological and physical disorders with vr,” *IEEE Computer Graphics and Applications*, vol. 21, no. 6, pp. 25–33, 2001. [Online]. Available: <https://doi.org/10.1109/38.963458>
- [206] S. Holm, “A simple sequentially rejective multiple test procedure,” *Scandinavian journal of statistics*, pp. 65–70, 1979, <https://www.jstor.org/stable/4615733>, accessed 18 February 2023.
- [207] C. Holz, S. Buthpitiya, and M. Knaust, “Bodyprint: Biometric user identification on mobile devices using the capacitive touchscreen to scan body parts,” in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI ’15. New York, NY, USA: Association for Computing Machinery, 2015, p. 3011–3014. [Online]. Available: <https://doi.org/10.1145/2702123.2702518>
- [208] M. Hoppe, J. Karolus, F. Dietz, P. W. Woźniak, A. Schmidt, and T.-K. Machulla, “Vrsneaky: Increasing presence in vr through gait-aware auditory feedback,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–9. [Online]. Available: <https://doi.org/10.1145/3290605.3300776>
- [209] S. Houde and C. Hill, “What do prototypes prototype?” in *Handbook of human-computer interaction*. Elsevier, 1997, pp. 367–381. [Online]. Available: <https://doi.org/10.1016/B978-044481862-1.50082-0>
- [210] N. Huaman, A. Krause, D. Wermke, J. H. Klemmer, C. Stransky, Y. Acar, and S. Fahl, “If you Can’t get them to the lab: Evaluating a virtual study environment with security information workers,” in *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*. Boston, MA: USENIX Association, Aug. 2022, pp. 313–330. [Online]. Available: <https://www.usenix.org/conference/soups2022/presentation/huaman>
- [211] B. Huang and H. Ling, “Deprocams: Simultaneous relighting, compensation and shape reconstruction for projector-camera systems,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 5, pp. 2725–2735, 2021. [Online]. Available: <https://doi.org/10.1109/TVCG.2021.3067771>

- [212] T. hundred fifty-five (255) pixel studios. (2021) City package. <https://assetstore.unity.com/packages/3d/environments/urban/city-package-107224>, accessed 18 February 2023.
- [213] D. Hunt and Y.-H. Eakes. (2019) Introducing the animation rigging preview package for unity 2019.1. <https://blogs.unity3d.com/2019/05/14/introducing-the-animation-rigging-preview-package-for-unity-2019-1/>, accessed 18 February 2023.
- [214] S. Hwang, M. Ahn, and K.-y. Wohn, “Maggetz: Customizable passive tangible controllers on and around conventional mobile devices,” in *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST ’13. New York, NY, USA: Association for Computing Machinery, 2013, p. 411–416. [Online]. Available: <https://doi.org/10.1145/2501988.2501991>
- [215] G. Iachello and L. Terrenghi, “Mobile hci 2004: Experience and reflection,” *IEEE Pervasive Computing*, vol. 4, no. 1, pp. 88–91, 2005. [Online]. Available: <https://doi.org/10.1109/MPRV.2005.19>
- [216] 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*, ser. CHI ’10. New York, NY, USA: ACM, 2010, p. 383–392. [Online]. Available: <https://doi.org/10.1145/1753326.1753384>
- [217] I. Ion, R. Reeder, and S. Consolvo, “{“... No} one can hack my {Mind”}: Comparing expert and {Non-Expert} security practices,” in *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*, 2015, pp. 327–346, <https://www.usenix.org/system/files/conference/soups2015/soups15-paper-ion.pdf>, accessed 18 February 2023.
- [218] A. Irlitti, T. Hoang, and F. Vetere, *Surrogate-Aloud: A Human Surrogate Method for Remote Usability Evaluation and Ideation in Virtual Reality*. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: <https://doi.org/10.1145/3411763.3451764>
- [219] M. R. Islam, D. Lee, L. S. Jahan, and I. Oakley, “Glasspass: Tapping gestures to unlock smart glasses,” in *Proceedings of the 9th Augmented Human International Conference*, ser. AH ’18. New York, NY, USA: Association for Computing Machinery, 2018. [Online]. Available: <https://doi.org/10.1145/3174910.3174936>
- [220] Y. Itoh, T. Kaminokado, and K. Akşit, “Beaming displays,” vol. 27, no. 5, 2021, pp. 2659–2668. [Online]. Available: <https://doi.org/10.1109/TVCG.2021.3067764>

- [221] L. Izzouzi and A. Steed, "Integrating rocketbox avatars with the ubiq social vr platform," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2022, pp. 69–70. [Online]. Available: <https://doi.org/10.1109/VRW55335.2022.00025>
- [222] M. R. Jamnik and D. J. Lane, "The use of reddit as an inexpensive source for high-quality data," *Practical Assessment, Research, and Evaluation*, vol. 22, no. 1, p. 5, 2017. [Online]. Available: <https://doi.org/10.7275/j18t-c009>
- [223] D. Jo, K. Kim, G. F. Welch, W. Jeon, Y. Kim, K.-H. Kim, and G. J. Kim, "The impact of avatar-owner visual similarity on body ownership in immersive virtual reality," in *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, 2017, pp. 1–2. [Online]. Available: <https://doi.org/10.1145/3139131.3141214>
- [224] D. Jo, K.-H. Kim, and G. J. Kim, "Effects of avatar and background types on users' co-presence and trust for mixed reality-based teleconference systems," in *Proceedings the 30th Conference on Computer Animation and Social Agents*, 2017, pp. 27–36. [Online]. Available: <https://doi.org/10.1145/2992138.2992146>
- [225] B. E. John and D. E. Kieras, "The goms family of user interface analysis techniques: Comparison and contrast," *ACM Trans. Comput.-Hum. Interact.*, vol. 3, no. 4, p. 320–351, dec 1996. [Online]. Available: <https://doi.org/10.1145/235833.236054>
- [226] P. Johnson, "Usability and mobility; interactions on the move," in *Proceedings of the First Workshop on Human-Computer Interaction with Mobile Devices*, 1998, <http://www.dcs.gla.ac.uk/johnson/papers/mobile/HCIMD1.html>, accessed 18 February 2023.
- [227] R. Johnstone, N. McDonnell, and J. R. Williamson, "When virtuality surpasses reality: Possible futures of ubiquitous xr," in *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '22. New York, NY, USA: Association for Computing Machinery, 2022. [Online]. Available: <https://doi.org/10.1145/3491101.3516396>
- [228] G. Jonathan, "A moving target: The evolution of human-computer interaction," *Human-computer interaction handbook: Fundamentals, evolving technologies, and emerging applications.(3rd edition)*. Taylor and Francis, 2012. [Online]. Available: <https://doi.org/10.1201/b11963-ch-101>
- [229] J. M. Jones, R. Duezguen, P. Mayer, M. Volkamer, and S. Das, "A literature review on virtual reality authentication," in *Human Aspects of Information Security and Assurance*, S. Furnell and N. Clarke, Eds. Cham: Springer International Publishing, 2021, pp. 189–198. [Online]. Available: https://doi.org/10.1007/978-3-030-81111-2_16

- [230] T. Jung, M. Dieck, H. Lee, and N. Chung, “Effects of virtual reality and augmented reality on visitor experiences in museum,” in *Information and communication technologies in tourism 2016*. Springer, 2016, pp. 621–635. [Online]. Available: https://doi.org/10.1007/978-3-319-28231-2_45
- [231] R. Kainda, I. Fléchais, and A. Roscoe, “Security and usability: Analysis and evaluation,” in *2010 International Conference on Availability, Reliability and Security*, 2010, pp. 275–282. [Online]. Available: <https://doi.org/10.1109/ARES.2010.77>
- [232] M. Kamarushi, S. Watson, G. Tigwell, and R. Peiris, “Onebuttonpin: A single button authentication method for blind and low vision users to improve accessibility and prevent eavesdropping,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 6, 08 2022. [Online]. Available: [10.1145/3546747](https://doi.org/10.1145/3546747)
- [233] M. Kaptein and J. Robertson, “Rethinking statistical analysis methods for chi,” in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, ser. CHI ’12. New York, NY, USA: ACM, 2012, p. 1105–1114. [Online]. Available: <https://doi.org/10.1145/2207676.2208557>
- [234] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, “Analyzing and improving the image quality of stylegan,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 8110–8119. [Online]. Available: <https://doi.org/10.48550/arXiv.1912.04958>
- [235] 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,” in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’20. New York, NY, USA: ACM, 2020, p. 1–21. [Online]. Available: <https://doi.org/10.1145/3313831.3376840>
- [236] C. Katsini, M. Belk, C. Fidas, N. Avouris, and G. Samaras, “Security and usability in knowledge-based user authentication: A review,” in *Proceedings of the 20th Pan-Hellenic Conference on Informatics*, ser. PCI ’16. New York, NY, USA: Association for Computing Machinery, 2016. [Online]. Available: <https://doi.org/10.1145/3003733.3003764>
- [237] 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*, 2012, pp. 523–537. [Online]. Available: <https://doi.org/10.1109/SP.2012.38>

- [238] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness," *The international journal of aviation psychology*, vol. 3, no. 3, pp. 203–220, 1993. [Online]. Available: https://doi.org/10.1207/s15327108ijap0303_3
- [239] 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 34th Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '16. New York, NY, USA: ACM, 2016. [Online]. Available: <https://doi.org/10.1145/2851581.2892314>
- [240] 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*, ser. MUM '17. New York, NY, USA: ACM, 2017. [Online]. Available: <https://doi.org/10.1145/3152832.3152851>
- [241] 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*, vol. 2, no. 3, 2018. [Online]. Available: <https://doi.org/10.3390/mti2030045>
- [242] M. Khamis, R. Hasholzner, A. Bulling, and F. Alt, "Gtmopass: Two-factor authentication on public displays using gaze-touch passwords and personal mobile devices," in *Proceedings of the 6th ACM International Symposium on Pervasive Displays*, ser. PerDis '17. New York, NY, USA: ACM, 2017. [Online]. Available: <https://doi.org/10.1145/3078810.3078815>
- [243] 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*, ser. ICMI 2017. New York, NY, USA: ACM, 2017. [Online]. Available: <https://doi.org/10.1145/3136755.3136809>
- [244] M. Khamis, A. Hoesl, A. Klimczak, M. Reiss, F. Alt, and A. Bulling, "Eyescout: Active eye tracking for position and movement independent gaze interaction with large public displays," in *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '17. New York, NY, USA: ACM, 2017, p. 155–166. [Online]. Available: <https://doi.org/10.1145/3126594.3126630>
- [245] 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,” *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 4, Dec. 2018. [Online]. Available: <https://doi.org/10.1145/3287052>
- [246] H. Khan, U. Hengartner, and D. Vogel, “Evaluating attack and defense strategies for smartphone pin shoulder surfing,” in *Proc. of the 2018 CHI Conf. on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2018, p. 1–10. [Online]. Available: <https://doi.org/10.1145/3173574.3173738>
- [247] R. Khan, R. Hasan, and J. Xu, “Sepia: Secure-pin-authentication-as-a-service for atm using mobile and wearable devices,” in *2015 3rd IEEE International Conference on Mobile Cloud Computing, Services, and Engineering*, 2015, pp. 41–50. [Online]. Available: <https://doi.org/10.1109/MobileCloud.2015.16>
- [248] D. Kim, P. Dunphy, P. Briggs, J. Hook, J. W. Nicholson, J. Nicholson, and P. Olivier, “Multi-touch authentication on tabletops,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’10. New York, NY, USA: Association for Computing Machinery, 2010, p. 1093–1102. [Online]. Available: <https://doi.org/10.1145/1753326.1753489>
- [249] Y. Kim, H. J. Kim, and Y. J. Kim, “Encountered-type haptic display for large vr environment using per-plane reachability maps,” *Computer Animation and Virtual Worlds*, vol. 29, no. 3-4, p. e1814, 2018. [Online]. Available: <https://doi.org/doi:10.1002/cav.1814>
- [250] I. Kirlappos and A. M. Sasse, “Security education against phishing: A modest proposal for a major rethink,” *IEEE Security Privacy*, vol. 10, no. 2, pp. 24–32, March 2012. [Online]. Available: <https://doi.org/10.1109/MSP.2011.179>
- [251] I. Kirlappos and M. A. Sasse, “What usable security really means: Trusting and engaging users,” in *International Conference on Human Aspects of Information Security, Privacy, and Trust*. Springer, 2014, pp. 69–78. [Online]. Available: https://doi.org/10.1007/978-3-319-07620-1_7
- [252] J. Kjeldskov and C. Graham, “A review of mobile hci research methods,” in *International Conference on Mobile Human-Computer Interaction*. Springer, 2003, pp. 317–335. [Online]. Available: https://doi.org/10.1007/978-3-540-45233-1_23
- [253] J. Kjeldskov and M. B. Skov, “Was it worth the hassle? ten years of mobile hci research discussions on lab and field evaluations,” in *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services*, ser. MobileHCI ’14. New York, NY, USA: Association for Computing Machinery, 2014, p. 43–52. [Online]. Available: <https://doi.org/10.1145/2628363.2628398>

- [254] J. Kjeldskov, M. B. Skov, B. S. Als, and R. T. Høegh, “Is it worth the hassle? exploring the added value of evaluating the usability of context-aware mobile systems in the field,” in *Mobile Human-Computer Interaction - MobileHCI 2004*, S. Brewster and M. Dunlop, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 61–73. [Online]. Available: https://doi.org/10.1007/978-3-540-28637-0_6
- [255] P. Knierim, V. Schwind, A. M. Feit, F. Nieuwenhuizen, and N. Henze, “Physical keyboards in virtual reality: Analysis of typing performance and effects of avatar hands,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–9. [Online]. Available: <https://doi.org/10.1145/3173574.3173919>
- [256] M. Koelle, S. Boll, T. Olsson, J. Williamson, H. Profita, S. Kane, and R. Mitchell, “(un)acceptable!?! re-thinking the social acceptability of emerging technologies,” in *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA ’18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–8. [Online]. Available: <https://doi.org/10.1145/3170427.3170620>
- [257] L. Koeman. (2018) How many participants do researchers recruit? a look at 678 ux/hci studies. <https://lisakoeman.nl/blog/how-many-participants-do-researchers-recruit-a-look-at-678-ux-hci-studies/>, accessed 18 February 2023.
- [258] L. Koeman. (2021) Hci/ux research: What methods do we use? <https://lisakoeman.nl/blog/hci-ux-research-what-methods-do-we-use/>, accessed 18 February 2023.
- [259] 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*, ser. CHI ’11. New York, NY, USA: Association for Computing Machinery, 2011, p. 2595–2604. [Online]. Available: <https://doi.org/10.1145/1978942.1979321>
- [260] ———, “Of passwords and people: Measuring the effect of password-composition policies,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’11. New York, NY, USA: Association for Computing Machinery, 2011, p. 2595–2604. [Online]. Available: <https://doi.org/10.1145/1978942.1979321>
- [261] R. Kovacs, E. Ofek, M. Gonzalez Franco, A. F. Siu, S. Marwecki, C. Holz, and M. Sinclair, “Haptic pivot: On-demand handhelds in vr,” in *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. New York, NY,

- USA: Association for Computing Machinery, 2020, p. 1046–1059. [Online]. Available: <https://doi.org/10.1145/3379337.3415854>
- [262] S. Kozubaev, C. Elsdén, N. Howell, M. L. J. Søndergaard, N. Merrill, B. Schulte, and R. Y. Wong, “Expanding modes of reflection in design futuring,” in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–15. [Online]. Available: <https://doi.org/10.1145/3313831.3376526>
- [263] L. Kraus, R. Schmidt, M. Walch, F. Schaub, and S. Möller, “On the use of emojis in mobile authentication,” in *ICT Systems Security and Privacy Protection*, S. De Capitani di Vimercati and F. Martinelli, Eds. Cham: Springer International Publishing, 2017, pp. 265–280. [Online]. Available: https://doi.org/10.1007/978-3-319-58469-0_18
- [264] V. Krauß, A. Boden, L. Oppermann, and R. Reiners, “Current practices, challenges, and design implications for collaborative ar/vr application development,” in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: <https://doi.org/10.1145/3411764.3445335>
- [265] K. Krawiecka, J. Sturgess, A. Petrova, and I. Martinovic, *Plug-and-Play: Framework for Remote Experimentation in Cyber Security*. New York, NY, USA: Association for Computing Machinery, 2021, p. 48–58. [Online]. Available: <https://doi.org/10.1145/3481357.3481518>
- [266] P. Krishnasamy, S. Belongie, and D. Kriegman, “Wet fingerprint recognition: Challenges and opportunities,” in *2011 International Joint Conference on Biometrics (IJCB)*, 2011, pp. 1–7. [Online]. Available: <https://doi.org/10.1109/IJCB.2011.6117594>
- [267] K. Krol, J. M. Spring, S. Parkin, and M. A. Sasse, “Towards robust experimental design for user studies in security and privacy,” in *The LASER Workshop: Learning from Authoritative Security Experiment Results (LASER 2016)*. San Jose, CA: USENIX Association, May 2016. [Online]. Available: <https://www.usenix.org/conference/laser2016/program/presentation/krol>
- [268] K. Krombholz, H. Hobel, M. Huber, and E. Weippl, “Advanced social engineering attacks,” *Journal of Information Security and Applications*, vol. 22, pp. 113–122, 2015, special Issue on Security of Information and Networks. [Online]. Available: <https://doi.org/10.1016/j.jisa.2014.09.005>
- [269] K. Krombholz, T. Hupperich, and T. Holz, “Use the force: Evaluating force-sensitive authentication for mobile devices,” in *Proceedings of the Twelfth USENIX Conference*

- on Usable Privacy and Security*, ser. SOUPS '16. USA: USENIX Association, 2016, p. 207–219, <https://www.usenix.org/conference/soups2016/technical-sessions/presentation/krombholz>, accessed 18 February 2023.
- [270] K. Krombholz, W. Mayer, M. Schmiedecker, and E. Weippl, ““i have no idea what i’m doing” - on the usability of deploying {HTTPS},” in *26th USENIX Security Symposium (USENIX Security 17)*, 2017, pp. 1339–1356, <https://www.usenix.org/system/files/conference/usenixsecurity17/sec17-krombholz.pdf>, accessed 18 February 2023.
- [271] R. A. Krueger and M. A. Casey, *Focus groups: A practical guide for applied research*. Sage publications, 2014, <https://us.sagepub.com/en-us/nam/focus-groups/book243860>, accessed 18 February 2023.
- [272] 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*, ser. SOUPS '07. New York, NY, USA: ACM, 2007, p. 13–19. [Online]. Available: <https://doi.org/10.1145/1280680.1280683>
- [273] O. M. Lapenta, A. P. Xavier, S. C. Côrrea, and P. S. Boggio, “Human biological and nonbiological point-light movements: Creation and validation of the dataset,” *Behavior research methods*, vol. 49, no. 6, pp. 2083–2092, 2017. [Online]. Available: <https://doi.org/10.3758/s13428-016-0843-9>
- [274] M. E. Latoschik, D. Roth, D. Gall, J. Achenbach, T. Waltemate, and M. Botsch, “The effect of avatar realism in immersive social virtual realities,” in *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, ser. VRST '17. New York, NY, USA: Association for Computing Machinery, 2017. [Online]. Available: <https://doi.org/10.1145/3139131.3139156>
- [275] B. Laugwitz, T. Held, and M. Schrepp, “Construction and evaluation of a user experience questionnaire,” in *Symposium of the Austrian HCI and usability engineering group*. Springer, 2008, pp. 63–76. [Online]. Available: https://doi.org/10.1007/978-3-540-89350-9_6
- [276] J. Lazar, J. H. Feng, and H. Hochheiser, *Research methods in human-computer interaction*. Morgan Kaufmann, 2017, <https://www.elsevier.com/books/research-methods-in-human-computer-interaction/lazar/978-0-12-805390-4>, accessed 18 February 2023.
- [277] J. Lee, R. Natarajan, S. S. Rodriguez, P. Panda, and E. Ofek, “Remotelab: A vr remote study toolkit,” in *The 35th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '22. New York, NY, USA: Association for

- Computing Machinery, 2022, <https://www.microsoft.com/en-us/research/publication/remotelab-virtual-reality-remote-study-toolkit/>, accessed 18 February 2023.
- [278] M.-K. Lee, “Security notions and advanced method for human shoulder-surfing resistant pin-entry,” *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 4, 2014. [Online]. Available: <https://doi.org/10.1109/TIFS.2014.2307671>
- [279] P. G. Leon, B. Ur, Y. Wang, M. Sleeper, R. Balebako, R. Shay, L. Bauer, M. Christodorescu, and L. F. Cranor, “What matters to users? factors that affect users’ willingness to share information with online advertisers,” in *Proceedings of the Ninth Symposium on Usable Privacy and Security*, ser. SOUPS ’13. New York, NY, USA: ACM, 2013. [Online]. Available: <https://doi.org/10.1145/2501604.2501611>
- [280] V. I. Levenshtein, “Binary codes capable of correcting deletions, insertions, and reversals,” in *Soviet physics doklady*, vol. 10, no. 8, 1966, pp. 707–710, <https://nymity.ch/sybilhunting/pdf/Levenshtein1966a.pdf>, accessed 18 February 2023.
- [281] C. Liang, C. Yu, X. Wei, X. Xu, Y. Hu, Y. Wang, and Y. Shi, “Auth+track: Enabling authentication free interaction on smartphone by continuous user tracking,” in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: <https://doi.org/10.1145/3411764.3445624>
- [282] J. Liebers, S. Brockel, U. Gruenefeld, and S. Schneegass, “Identifying users by their hand tracking data in augmented and virtual reality,” *International Journal of Human–Computer Interaction*, vol. 0, no. 0, pp. 1–16, 2022. [Online]. Available: <https://doi.org/10.1080/10447318.2022.2120845>
- [283] J. Liebers and S. Schneegass, “Introducing functional biometrics: Using body-reflections as a novel class of biometric authentication systems,” in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA ’20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1–7. [Online]. Available: <https://doi.org/10.1145/3334480.3383059>
- [284] S. Linxen, C. Sturm, F. Brühlmann, V. Cassau, K. Opwis, and K. Reinecke, “How weird is chi?” in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: <https://doi.org/10.1145/3411764.3445488>
- [285] L. Little, “Attitudes towards technology use in public zones: The influence of external factors on atm use,” in *EA on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2003, p. 990–991. [Online]. Available: <https://doi.org/10.1145/765891.766110>

- [286] Y. Liu, J. Goncalves, D. Ferreira, B. Xiao, S. Hosio, and V. Kostakos, “Chi 1994-2013: Mapping two decades of intellectual progress through co-word analysis,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 3553–3562. [Online]. Available: <https://doi.org/10.1145/2556288.2556969>
- [287] M. Lombard, T. B. Ditton, and L. Weinstein, “Measuring presence: the temple presence inventory,” in *Proceedings of the 12th annual international workshop on presence*, 2009, pp. 1–15, http://matthewlombard.com/ISPR/Proceedings/2009/Lombard_et_al.pdf, accessed 18 February 2023.
- [288] A. D. Luca, S. Das, M. Ortlieb, I. Ion, and B. Laurie, “Expert and Non-Expert attitudes towards (secure) instant messaging,” in *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*. Denver, CO: USENIX Association, Jun. 2016, pp. 147–157. [Online]. Available: <https://dl.acm.org/doi/10.5555/3235895.3235908>
- [289] J.-L. Lugrin, M. Wiedemann, D. Bieberstein, and M. E. Latoschik, “Influence of avatar realism on stressful situation in vr,” in *2015 IEEE Virtual Reality (VR)*, 2015, pp. 227–228. [Online]. Available: <https://doi.org/10.1109/VR.2015.7223378>
- [290] C. Mai and M. Khamis, “Public hmds: Modeling and understanding user behavior around public head-mounted displays,” in *Proceedings of the 7th ACM International Symposium on Pervasive Displays*, ser. PerDis '18. New York, NY, USA: Association for Computing Machinery, 2018. [Online]. Available: <https://doi.org/10.1145/3205873.3205879>
- [291] V. Mäkelä, R. Radiah, S. Alsherif, M. Khamis, C. Xiao, L. Borchert, A. Schmidt, and F. Alt, “Virtual field studies: Conducting studies on public displays in virtual reality,” in *Proc. of the 2020 CHI Conf. on Human Factors in Comp. Systems*, ser. CHI '20. New York, NY, USA: ACM, 2020, p. 1–15. [Online]. Available: <https://doi.org/10.1145/3313831.3376796>
- [292] M. Malheiros, S. Brostoff, C. Jennett, and M. A. Sasse, “Would you sell your mother’s data? personal data disclosure in a simulated credit card application,” in *The economics of information security & privacy*, 2013. [Online]. Available: https://doi.org/10.1007/978-3-642-39498-0_11
- [293] Y. Malhotra and D. Galletta, “Extending the technology acceptance model to account for social influence: theoretical bases and empirical validation,” in *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers*, vol. Track1, 1999, pp. 14 pp.–. [Online]. Available: <https://doi.org/10.1109/HICSS.1999.772658>

- [294] N. Malkin, M. Harbach, A. De Luca, and S. Egelman, “The anatomy of smartphone unlocking: Why and how android users around the world lock their phones,” *GetMobile: Mobile Comp. and Comm.*, vol. 20, no. 3, p. 42–46, Jan. 2017. [Online]. Available: <https://doi.org/10.1145/3036699.3036712>
- [295] K. Malterud, “Qualitative research: standards, challenges, and guidelines,” *The lancet*, vol. 358, no. 9280, pp. 483–488, 2001. [Online]. Available: [https://doi.org/10.1016/s0140-6736\(01\)05627-6](https://doi.org/10.1016/s0140-6736(01)05627-6)
- [296] S. Mare, M. Baker, and J. Gummesson, “A study of authentication in daily life,” in *Proceedings of the Twelfth USENIX Conference on Usable Privacy and Security*, ser. SOUPS ’16. USA: USENIX Association, 2016, p. 189–206, <https://www.usenix.org/conference/soups2016/technical-sessions/presentation/mare>, accessed 18 February 2023.
- [297] P. Markert, D. V. Bailey, M. Golla, M. Dürmuth, and A. J. Aviv, “This PIN Can Be Easily Guessed: Analyzing the Security of Smartphone Unlock PINs,” in *IEEE Symposium on Security and Privacy*, ser. SP ’20. San Francisco, California, USA: IEEE, May 2020, pp. 1525–1542. [Online]. Available: <https://doi.org/10.1109/SP40000.2020.00100>
- [298] P. Markert, D. V. Bailey, M. Golla, M. Dürmuth, and A. J. Aviv, “On the security of smartphone unlock pins,” *ACM Transactions on Privacy and Security*, vol. 24, no. 4, sep 2021. [Online]. Available: <https://doi.org/10.1145/3473040>
- [299] K. Marky, M. Schmitz, V. Zimmermann, M. Herbers, K. Kunze, and M. Mühlhäuser, “3d-auth: Two-factor authentication with personalized 3d-printed items,” in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1–12. [Online]. Available: <https://doi.org/10.1145/3313831.3376189>
- [300] D. Marques, T. Guerreiro, L. Carriço, I. Beschastnikh, and K. Beznosov, “Vulnerability & blame: Making sense of unauthorized access to smartphones,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–13. [Online]. Available: <https://doi.org/10.1145/3290605.3300819>
- [301] F. Mathis, “Moving usable security and privacy research out of the lab: Adding virtual reality to the research arsenal,” 2022, http://fmathis.com/publications/LightningTalk_FlorianMathis_SOUPS2022.pdf, accessed 18 February 2023.

- [302] F. Mathis, J. O’Hagan, K. Vaniea, and M. Khamis, “Stay home! conducting remote usability evaluations of novel real-world authentication systems using virtual reality,” in *Proceedings of the 2022 International Conference on Advanced Visual Interfaces*, ser. AVI 2022. New York, NY, USA: Association for Computing Machinery, 2022. [Online]. Available: <https://doi.org/10.1145/3531073.3531087>
- [303] F. Mathis, J. O’Hagan, M. Khamis, and K. Vaniea, “Virtual reality observations: Using virtual reality to augment lab-based shoulder surfing research,” in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2022, pp. 291–300. [Online]. Available: <https://doi.org/10.1109/VR51125.2022.00048>
- [304] F. Mathis, K. Vaniea, and M. Khamis, “Observing virtual avatars: The impact of avatars’ fidelity on identifying interactions,” in *Proc. of the 24th International Conf. on Academic Mindtrek*, ser. AcademicMindtrek ’21. New York, NY, USA: ACM, 2021.
- [305] F. Mathis, K. Vaniea, and M. Khamis, “Can i borrow your atm? using virtual reality for (simulated) in situ authentication research,” in *IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR)*. IEEE, 2022, pp. 301–310. [Online]. Available: <https://doi.org/10.1109/VR51125.2022.00049>
- [306] F. Mathis, K. Vaniea, and M. Khamis, “Prototyping usable privacy and security systems: Insights from experts,” *International Journal of Human–Computer Interaction*, 2022. [Online]. Available: <https://doi.org/10.1080/10447318.2021.1949134>
- [307] F. Mathis, J. H. Williamson, V. Kami, and M. Khamis, “RubikAuth: Fast and Secure Authentication in Virtual Reality,” in *Proceedings of the 38th Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA ’20. New York, NY, USA: ACM, 2020. [Online]. Available: <http://dx.doi.org/10.1145/3334480.3382827>
- [308] 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.*, vol. 28, no. 1, Jan. 2021. [Online]. Available: <https://doi.org/10.1145/3428121>
- [309] F. Mathis, X. Zhang, J. O’Hagan, D. Medeiros, P. Saeghe, M. McGill, S. Brewster, and M. Khamis, “Remote xr studies: The golden future of hci research?” in *Proceedings of the CHI 2021 Workshop on XR Remote Research*, 2021, https://www.research.manchester.ac.uk/portal/files/195248903/chi2021_workshop_remoteXR.pdf, accessed 18 February 2023.

- [310] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino, “Impact of artificial ”gummy” fingers on fingerprint systems,” in *Optical Security and Counterfeit Deterrence Techniques IV*, R. L. van Renesse, Ed., vol. 4677, International Society for Optics and Photonics. SPIE, 2002, pp. 275 – 289. [Online]. Available: <https://doi.org/10.1117/12.462719>
- [311] A. M. Matwyshyn, A. Cui, A. D. Keromytis, and S. J. Stolfo, “Ethics in security vulnerability research,” *IEEE Security Privacy*, vol. 8, no. 2, pp. 67–72, 2010. [Online]. Available: <https://doi.org/10.1109/MSP.2010.67>
- [312] M.-E. Maurer, A. De Luca, and S. Kempe, “Using data type based security alert dialogs to raise online security awareness,” in *Proceedings of the Seventh Symposium on Usable Privacy and Security*, ser. SOUPS ’11. New York, NY, USA: Association for Computing Machinery, 2011. [Online]. Available: <https://doi.org/10.1145/2078827.2078830>
- [313] T. Mazuryk and M. Gervautz, “Virtual reality-history, applications, technology and future,” 1996, <https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.42.7849>, accessed 18 February 2023.
- [314] M. McGill, D. Boland, R. Murray-Smith, and S. Brewster, “A dose of reality: Overcoming usability challenges in vr head-mounted displays,” in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI ’15. New York, NY, USA: Association for Computing Machinery, 2015, p. 2143–2152. [Online]. Available: <https://doi.org/10.1145/2702123.2702382>
- [315] M. R. McGrail, C. M. Rickard, and R. Jones, “Publish or perish: A systematic review of interventions to increase academic publication rates,” *Higher Education Research & Development*, vol. 25, no. 1, pp. 19–35, 2006. [Online]. Available: <https://doi.org/10.1080/07294360500453053>
- [316] J. E. McGrath, J. Martin, and R. A. Kulka, “Some quasi-rules for making judgment calls in research,” *American Behavioral Scientist*, vol. 25, no. 2, pp. 211–224, 1981. [Online]. Available: <https://doi.org/10.1177/000276428102500206>
- [317] J. McGrenere, R. M. Baecker, and K. S. Booth, “An evaluation of a multiple interface design solution for bloated software,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2002, pp. 164–170. [Online]. Available: <https://doi.org/10.1145/503376.503406>
- [318] K. Y. McKenna and J. A. Bargh, “Plan 9 from cyberspace: The implications of the internet for personality and social psychology,” *Personality and social*

- psychology review*, vol. 4, no. 1, pp. 57–75, 2000. [Online]. Available: https://doi.org/10.1207/S15327957PSPR0401_6
- [319] R. P. McMahan, D. A. Bowman, D. J. Zielinski, and R. B. Brady, “Evaluating display fidelity and interaction fidelity in a virtual reality game,” vol. 18, no. 4, 2012, pp. 626–633. [Online]. Available: <https://doi.org/10.1109/TVCG.2012.43>
- [320] R. P. McMahan, C. Lai, and S. K. Pal, “Interaction fidelity: the uncanny valley of virtual reality interactions,” in *International conference on virtual, Augmented and Mixed Reality*. Springer, 2016, pp. 59–70. [Online]. Available: https://doi.org/10.1007/978-3-319-39907-2_6
- [321] M. McNutt, “Reproducibility,” 2014. [Online]. Available: <https://doi.org/10.1126/science.1250475>
- [322] J. McVeigh-Schultz, M. Kreminski, K. Prasad, P. Hoberman, and S. S. Fisher, “Immersive design fiction: Using vr to prototype speculative interfaces and interaction rituals within a virtual storyworld,” in *Proceedings of the 2018 Designing Interactive Systems Conference*, ser. DIS ’18. New York, NY, USA: Association for Computing Machinery, 2018, p. 817–829. [Online]. Available: <https://doi.org/10.1145/3196709.3196793>
- [323] L. Mecke, K. Pfeuffer, S. Prange, and F. Alt, “Open sesame! user perception of physical, biometric, and behavioural authentication concepts to open doors,” in *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, ser. MUM 2018. New York, NY, USA: Association for Computing Machinery, 2018, p. 153–159. [Online]. Available: <https://doi.org/10.1145/3282894.3282923>
- [324] L. I. Meho, “E-mail interviewing in qualitative research: A methodological discussion,” *Journal of the American Society for Information Science and Technology*, vol. 57, no. 10, pp. 1284–1295, 2006. [Online]. Available: <https://doi.org/10.1002/asi.20416>
- [325] A.-S. Melenhorst, W. A. Rogers, and E. C. Caylor, “The use of communication technologies by older adults: exploring the benefits from the user’s perspective,” in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 45, no. 3. SAGE Publications Sage CA: Los Angeles, CA, 2001, pp. 221–225. [Online]. Available: <https://doi.org/10.1177/154193120104500305>
- [326] M. Melo, G. Gonçalves, J. Vasconcelos-Raposo, and M. Bessa, “How much presence is enough? qualitative scales for interpreting the igroup presence questionnaire score,” *IEEE Access*, vol. 11, pp. 24 675–24 685, 2023.

- [327] N. Memon, “How biometric authentication poses new challenges to our security and privacy [in the spotlight],” *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 196–194, 2017.
- [328] Meta. Meta quest 2: Immersive all-in-one vr headset. <https://www.meta.com/gb/quest/products/quest-2/>, accessed 18 February 2023.
- [329] Microsoft. (2019, 08) Hand menu - mixed reality — microsoft docs. <https://docs.microsoft.com/en-us/windows/mixed-reality/design/hand-menu>, accessed 18 February 2023.
- [330] P. Milgram and F. Kishino, “A taxonomy of mixed reality visual displays,” *IEICE TRANSACTIONS on Information and Systems*, vol. 77, no. 12, pp. 1321–1329, 1994.
- [331] R. Miller, N. K. Banerjee, and S. Banerjee, “Within-system and cross-system behavior-based biometric authentication in virtual reality,” in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 2020, pp. 311–316. [Online]. Available: <https://doi.org/10.1109/VRW50115.2020.00070>
- [332] M. Minderer, C. D. Harvey, F. Donato, and E. I. Moser, “Virtual reality explored,” *Nature*, vol. 533, no. 7603, pp. 324–325, 2016. [Online]. Available: <https://doi.org/10.1038/nature17899>
- [333] Miro. (2021) Miro — online whiteboard for visual collaboration. Accessed 18 February 2023. [Online]. Available: <https://miro.com/>
- [334] B. J. Mohler, S. H. Creem-Regehr, W. B. Thompson, and H. H. Bühlhoff, “The effect of viewing a self-avatar on distance judgments in an hmd-based virtual environment,” *Presence*, vol. 19, no. 3, pp. 230–242, 2010. [Online]. Available: <https://doi.org/10.1162/pres.19.3.230>
- [335] F. Monroe and A. D. Rubin, “Keystroke dynamics as a biometric for authentication,” *Future Generation Computer Systems*, vol. 16, no. 4, pp. 351–359, 2000. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0167739X9900059X>
- [336] R. D. Morey *et al.*, “Confidence intervals from normalized data: A correction to cousineau (2005),” vol. 4, no. 2, pp. 61–64, 2008, <https://www.tqmp.org/RegularArticles/vol04-2/p061/p061.pdf>, accessed 18 February 2023.
- [337] M. Mori, K. F. MacDorman, and N. Kageki, “The uncanny valley [from the field],” *IEEE Robotics & Automation Magazine*, vol. 19, no. 2, pp. 98–100, 2012.
- [338] L. Motion. (2019) Unity assets for leap motion orion beta. <https://developer.leapmotion.com/unity>, accessed 18 February 2023.

- [339] A. Mottelson and K. Hornbæk, “Virtual reality studies outside the laboratory,” in *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, ser. VRST ’17. New York, NY, USA: Association for Computing Machinery, 2017. [Online]. Available: <https://doi.org/10.1145/3139131.3139141>
- [340] A. Mottelson, G. B. Petersen, K. Lilija, and G. Makransky, “Conducting unsupervised virtual reality user studies online,” *Frontiers in Virtual Reality*, 2021. [Online]. Available: <https://doi.org/10.3389/frvir.2021.681482>
- [341] Mozilla. (2022) Mozilla labs - hubs by mozilla. <https://hubs.mozilla.com/>, accessed 18 February 2023.
- [342] F. F. Mueller, P. Lopes, P. Strohmeier, W. Ju, C. Seim, M. Weigel, S. Nanayakkara, M. Obrist, Z. Li, J. Delfa, J. Nishida, E. M. Gerber, D. Svanaes, J. Grudin, S. Greuter, K. Kunze, T. Erickson, S. Greenspan, M. Inami, J. Marshall, H. Reiterer, K. Wolf, J. Meyer, T. Schiphorst, D. Wang, and P. Maes, “Next steps for human-computer integration,” in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2020, p. 1–15. [Online]. Available: <https://doi.org/10.1145/3313831.3376242>
- [343] J. Müller, R. Walter, G. Bailly, M. Nischt, and F. Alt, “Looking glass: A field study on noticing interactivity of a shop window,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’12. New York, NY, USA: Association for Computing Machinery, 2012, p. 297–306. [Online]. Available: <https://doi.org/10.1145/2207676.2207718>
- [344] M. Murcia-López and A. Steed, “A comparison of virtual and physical training transfer of bimanual assembly tasks,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 4, pp. 1574–1583, 2018. [Online]. Available: <https://doi.org/10.1109/TVCG.2018.2793638>
- [345] T. Mustafa, R. Matovu, A. Serwadda, and N. Muirhead, “Unsure how to authenticate on your vr headset? come on, use your head!” in *Proceedings of the Fourth ACM International Workshop on Security and Privacy Analytics*, ser. IWSPA ’18. New York, NY, USA: Association for Computing Machinery, 2018, p. 23–30. [Online]. Available: <https://doi.org/10.1145/3180445.3180450>
- [346] A. K. Mutasim, A. U. Batmaz, and W. Stuerzlinger, “Pinch, click, or dwell: Comparing different selection techniques for eye-gaze-based pointing in virtual reality,” in *ACM Symposium on Eye Tracking Research and Applications*, ser. ETRA ’21 Short Papers. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: <https://doi.org/10.1145/3448018.3457998>

- [347] B. A. Myers, “A brief history of human-computer interaction technology,” *Interactions*, vol. 5, no. 2, p. 44–54, mar 1998. [Online]. Available: <https://doi.org/10.1145/274430.274436>
- [348] A. Naiakshina, A. Danilova, C. Tiefenau, and M. Smith, “Deception task design in developer password studies: Exploring a student sample,” in *Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018)*, 2018, pp. 297–313. [Online]. Available: <https://www.usenix.org/conference/soups2018/presentation/naiakshina>
- [349] NationalCash. (2021) Atm statistics. <http://www.nationalcash.com/statistics/>, accessed 18 February 2023.
- [350] A. Ng, D. Medeiros, M. McGill, J. Williamson, and S. Brewster, “The passenger experience of mixed reality virtual display layouts in airplane environments,” in *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2021, pp. 265–274. [Online]. Available: <https://doi.org/10.1109/ISMAR52148.2021.00042>
- [351] D. C. Nguyen, E. Derr, M. Backes, and S. Bugiel, “Short text, large effect: Measuring the impact of user reviews on android app security & privacy,” *2019 IEEE Symposium on Security and Privacy (SP)*, pp. 555–569, 2019. [Online]. Available: <https://doi.org/10.1109/SP.2019.00012>
- [352] D. C. Nguyen, D. Wermke, Y. Acar, M. Backes, C. Weir, and S. Fahl, “A stitch in time: Supporting android developers in writingsecure code,” in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS ’17. New York, NY, USA: Association for Computing Machinery, 2017, p. 1065–1077. [Online]. Available: <https://doi.org/10.1145/3133956.3133977>
- [353] T. Nguyen and N. Memon, “Tap-based user authentication for smartwatches,” *Computers & Security*, vol. 78, pp. 174–186, 2018. [Online]. Available: <https://doi.org/10.1016/j.cose.2018.07.001>
- [354] A. L. Nichols and J. K. Maner, “The good-subject effect: Investigating participant demand characteristics,” *The Journal of general psychology*, vol. 135, no. 2, pp. 151–166, 2008. [Online]. Available: <https://doi.org/10.3200/GENP.135.2.151-166>
- [355] J. Nicholson, L. Coventry, and P. Briggs, “Age-related performance issues for pin and face-based authentication systems,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’13. New York, NY, USA: Association for Computing Machinery, 2013, p. 323–332. [Online]. Available: <https://doi.org/10.1145/2470654.2470701>

- [356] J. Nicholson, J. Terry, H. Beckett, and P. Kumar, *Understanding Young People's Experiences of Cybersecurity*. New York, NY, USA: Association for Computing Machinery, 2021, p. 200–210. [Online]. Available: <https://doi.org/10.1145/3481357.3481520>
- [357] C. M. Nielsen, M. Overgaard, M. B. Pedersen, J. Stage, and S. Stenild, "It's worth the hassle! the added value of evaluating the usability of mobile systems in the field," in *Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles*, ser. NordiCHI '06. New York, NY, USA: Association for Computing Machinery, 2006, p. 272–280. [Online]. Available: <https://doi.org/10.1145/1182475.1182504>
- [358] J. Nielsen. (2000) Security & human factors. [Online]. Available: <https://www.nngroup.com/articles/security-and-human-factors/>
- [359] J. Nielsen and R. Molich, "Heuristic evaluation of user interfaces," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '90. New York, NY, USA: Association for Computing Machinery, 1990, p. 249–256. [Online]. Available: <https://doi.org/10.1145/97243.97281>
- [360] B. Nouredin, P. D. Lawrence, and C. Man, "A non-contact device for tracking gaze in a human computer interface," *Computer Vision and Image Understanding*, vol. 98, no. 1, pp. 52–82, 2005. [Online]. Available: <https://doi.org/10.1016/j.cviu.2004.07.005>
- [361] D. Nyang, A. Mohaisen, and J. Kang, "Keylogging-resistant visual authentication protocols," *IEEE Transactions on Mobile Computing*, vol. 13, no. 11, pp. 2566–2579, 2014. [Online]. Available: <https://doi.org/10.1109/TMC.2014.2307331>
- [362] Oculus. (2021) Oculus integration sdk: Hand tracking in unity. <https://developer.oculus.com/documentation/unity/unity-handtracking/>, accessed 18 February 2023.
- [363] Oculus. (2022) Oculus guardian. <https://support.oculus.com/guardian>, accessed 18 February 2023.
- [364] N. Ogawa, T. Narumi, and M. Hirose, "Virtual hand realism affects object size perception in body-based scaling," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2019, pp. 519–528. [Online]. Available: <https://doi.org/10.1109/VR.2019.8798040>
- [365] A. A. Ogunyemi, D. Lamas, M. K. Lárusdóttir, and F. Loizides, "A systematic mapping study of hci practice research," *International Journal of Human-Computer Interaction*, vol. 35, no. 16, pp. 1461–1486, 2019. [Online]. Available: <https://doi.org/10.1080/10447318.2018.1541544>

- [366] J. O'Hagan, J. R. Williamson, and M. Khamis, "Bystander interruption of vr users," in *Proceedings of the 9TH ACM International Symposium on Pervasive Displays*, ser. PerDis '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 19–27. [Online]. Available: <https://doi.org/10.1145/3393712.3395339>
- [367] I. Olade, H.-N. Liang, C. Fleming, and C. Champion, "Exploring the vulnerabilities and advantages of swipe or pattern authentication in virtual reality (vr)," in *Proceedings of the 2020 4th International Conference on Virtual and Augmented Reality Simulations*, ser. ICVARS 2020. New York, NY, USA: ACM, 2020, p. 45–52. [Online]. Available: <https://doi.org/10.1145/3385378.3385385>
- [368] E. A. Oladimeji, S. Supakkul, and L. Chung, "Security threat modeling and analysis: A goal-oriented approach." Citeseer, <https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.103.2997>, accessed 18 February 2023.
- [369] S. Olinsky, P. M. Desai, S. Turkay, E. M. Heitkemper, E. G. Mitchell, L. Mamykina, and M. L. Hwang, "Meals for monsters: A mobile application for the feasibility of gaming and social mechanisms," in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: <https://doi.org/10.1145/3411763.3451789>
- [370] OptiTrack. (2022) Optitrack system. <https://optitrack.com/>, accessed 18 February 2023.
- [371] A. S. Originals. (2020) Snaps prototype — office. <https://assetstore.unity.com/packages/3d/environments/snaps-prototype-office-137490>, accessed 18 February 2023.
- [372] M. T. Orne, "On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications." *American psychologist*, vol. 17, no. 11, p. 776, 1962. [Online]. Available: <https://doi.org/10.4324/9781315129945-26>
- [373] J. M. Orr, P. R. Sackett, and C. L. Dubois, "Outlier detection and treatment in i/o psychology: A survey of researcher beliefs and an empirical illustration," *Personnel Psychology*, vol. 44, no. 3, pp. 473–486, 1991. [Online]. Available: <https://doi.org/10.1111/j.1744-6570.1991.tb02401.x>
- [374] M. Osadchy, B. Pinkas, A. Jarrous, and B. Moskovich, "Scifi-a system for secure face identification," in *2010 IEEE Symposium on Security and Privacy*. IEEE, 2010, pp. 239–254. [Online]. Available: <https://doi.org/10.1109/SP.2010.39>

- [375] A. Oulasvirta, “Field experiments in hci: promises and challenges,” in *Future interaction design II*. Springer, 2009, pp. 87–116. [Online]. Available: https://doi.org/10.1007/978-1-84800-385-9_5
- [376] V. Paneva, M. Bachynskyi, and J. Müller, “Levitation simulator: Prototyping ultrasonic levitation interfaces in virtual reality,” in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1–12. [Online]. Available: <https://doi.org/10.1145/3313831.3376409>
- [377] C. L. Paul, E. Morse, A. Zhang, Y.-Y. Choong, and M. Theofanos, “A field study of user behavior and perceptions in smartcard authentication,” in *IFIP Conference on Human-Computer Interaction*. Springer, 2011, pp. 1–17. [Online]. Available: https://doi.org/10.1007/978-3-642-23768-3_1
- [378] T. C. Peck and M. Gonzalez-Franco, “Avatar embodiment. a standardized questionnaire,” *Frontiers in Virtual Reality*, vol. 1, p. 44, 2021. [Online]. Available: <https://doi.org/10.3389/frvir.2020.575943>
- [379] G. B. Petersen, A. Mottelson, and G. Makransky, “Pedagogical agents in educational vr: An in the wild study,” in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: <https://doi.org/10.1145/3411764.3445760>
- [380] K. Pfeuffer, J. Alexander, M. K. Chong, and H. Gellersen, “Gaze-touch: Combining gaze with multi-touch for interaction on the same surface,” in *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST ’14. New York, NY, USA: Association for Computing Machinery, 2014, p. 509–518. [Online]. Available: <https://doi.org/10.1145/2642918.2647397>
- [381] 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*, ser. CHI ’19. New York, NY, USA: ACM, 2019, pp. 110:1–110:12. [Online]. Available: <http://doi.acm.org/10.1145/3290605.3300340>
- [382] 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*, ser. UIST ’13. New York, NY, USA: Association for Computing Machinery, 2013, p. 261–270. [Online]. Available: <https://doi.org/10.1145/2501988.2501998>

- [383] C. A. Pierce and H. Aguinis, "Using virtual reality technology in organizational behavior research," *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, vol. 18, no. 5, pp. 407–410, 1997. [Online]. Available: [http://dx.doi.org/10.1002/\(SICI\)1099-1379\(199709\)18:5%3C407::AID-JOB869%3E3.0.CO;2-P](http://dx.doi.org/10.1002/(SICI)1099-1379(199709)18:5%3C407::AID-JOB869%3E3.0.CO;2-P)
- [384] T. Piumsomboon, G. Lee, R. W. Lindeman, and M. Billinghurst, "Exploring natural eye-gaze-based interaction for immersive virtual reality," in *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, 2017, pp. 36–39. [Online]. Available: <https://doi.org/10.1109/3DUI.2017.7893315>
- [385] T. Piumsomboon, A. Clark, M. Billinghurst, and A. Cockburn, "User-defined gestures for augmented reality," in *IFIP Conference on Human-Computer Interaction*. Springer, 2013, pp. 282–299. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-40480-1_18
- [386] T. Piumsomboon, G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman, B. H. Thomas, and M. Billinghurst, "Mini-me: An adaptive avatar for mixed reality remote collaboration," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–13. [Online]. Available: <https://doi.org/10.1145/3173574.3173620>
- [387] P. G. Polson, C. Lewis, J. Rieman, and C. Wharton, "Cognitive walkthroughs: a method for theory-based evaluation of user interfaces," *International Journal of man-machine studies*, vol. 36, no. 5, pp. 741–773, 1992. [Online]. Available: [https://doi.org/10.1016/0020-7373\(92\)90039-N](https://doi.org/10.1016/0020-7373(92)90039-N)
- [388] A. P. Pons and P. Polak, "Understanding user perspectives on biometric technology," *Commun. ACM*, vol. 51, no. 9, pp. 115–118, Sep. 2008. [Online]. Available: <https://doi.org/10.1145/1378727.1389971>
- [389] R. Poppe, R. Rienks, and B. v. Dijk, "Evaluating the future of hci: challenges for the evaluation of emerging applications," in *Artificial Intelligence for Human Computing*. Springer, 2007, pp. 234–250. [Online]. Available: https://doi.org/10.1007/978-3-540-72348-6_12
- [390] S. Prange, C. George, and F. Alt, "Design considerations for usable authentication in smart homes," in *Mensch Und Computer 2021*, ser. MuC '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 311–324. [Online]. Available: <https://doi.org/10.1145/3473856.3473878>

- [391] S. Prange, L. Mecke, A. Nguyen, M. Khamis, and F. Alt, *Don't Use Fingerprint, It's Raining! How People Use and Perceive Context-Aware Selection of Mobile Authentication*. New York, NY, USA: Association for Computing Machinery, 2020. [Online]. Available: <https://doi.org/10.1145/3399715.3399823>
- [392] S. Prange, A. Shams, R. Piening, Y. Abdelrahman, and F. Alt, "Priview– exploring visualisations to support users' privacy awareness," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI '21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: <https://doi.org/10.1145/3411764.3445067>
- [393] L. Press, "Before the altar: The history of personal computing," *Communications of the ACM*, vol. 36, no. 9, pp. 27–33, 1993. [Online]. Available: <https://doi.org/10.1145/162685.162697>
- [394] Prolific. (2022) Quickly find research participants you can trust. <https://www.prolific.co/>, accessed 18 February 2023,.
- [395] S. Putze, D. Alexandrovsky, F. Putze, S. Höffner, J. D. Smeddinck, and R. Malaka, "Breaking the experience: Effects of questionnaires in vr user studies," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI '20. New York, NY, USA: ACM, 2020, p. 1–15. [Online]. Available: <https://doi.org/10.1145/3313831.3376144>
- [396] L. Qiu, A. De Luca, I. Muslukhov, and K. Beznosov, "Towards understanding the link between age and smartphone authentication," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19. New York, NY, USA: ACM, 2019. [Online]. Available: <https://doi.org/10.1145/3290605.3300393>
- [397] Qualtrics. (2005) Qualtrics experience managements. <https://www.qualtrics.com/>, accessed 18 February 2023.
- [398] Qualtrics. (2005) Qualtrics experience managements randomizer. <https://www.qualtrics.com/support/survey-platform/survey-module/survey-flow/standard-elements/randomizer/>, accessed 18 February 2023.
- [399] E. Rader, R. Wash, and B. Brooks, "Stories as informal lessons about security," in *Proceedings of the Eighth Symposium on Usable Privacy and Security*, ser. SOUPS '12. New York, NY, USA: Association for Computing Machinery, 2012. [Online]. Available: <https://doi.org/10.1145/2335356.2335364>
- [400] 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.*, vol. 28, no. 6, nov 2021. [Online]. Available: <https://doi.org/10.1145/3472617>
- [401] E. D. Ragan, D. A. Bowman, R. Kopper, C. Stinson, S. Scerbo, and R. P. McMahan, “Effects of field of view and visual complexity on virtual reality training effectiveness for a visual scanning task,” *IEEE transactions on visualization and computer graphics*, vol. 21, no. 7, pp. 794–807, 2015. [Online]. Available: <https://doi.org/10.1109/TVCG.2015.2403312>
- [402] K. Ragozin, Y. S. 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*. ACM, 2019, p. 18. [Online]. Available: <https://doi.org/10.1145/3338286.3340129>
- [403] Razer. (2020) Razer blade 15: Nvidia geforce rtx 2080. <https://www.razer.com/gb-en/gaming-laptops/razer-blade\protect\leavevmode@ifvmode\kern+.1667em\relaxaccessed18February2023>.
- [404] F. Rebelo, P. Noriega, E. Duarte, and M. Soares, “Using virtual reality to assess user experience,” *Human Factors*, vol. 54, no. 6, pp. 964–982, 2012. [Online]. Available: <https://doi.org/10.1177/0018720812465006>
- [405] E. M. Redmiles, ““should i worry?” a cross-cultural examination of account security incident response,” in *2019 2019 IEEE Symposium on Security and Privacy (SP)*. Los Alamitos, CA, USA: IEEE Computer Society, may 2019, pp. 1107–1121. [Online]. Available: <https://doi.ieeecomputersociety.org/10.1109/SP.2019.00059>
- [406] E. M. Redmiles, Y. Acar, S. Fahl, and M. L. Mazurek, “A summary of survey methodology best practices for security and privacy researchers,” Tech. Rep., 2017. [Online]. Available: <https://doi.org/10.13016/M22K2W>
- [407] E. M. Redmiles, S. Kross, and M. L. Mazurek, “Where is the digital divide? a survey of security, privacy, and socioeconomics,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’17. New York, NY, USA: Association for Computing Machinery, 2017, p. 931–936. [Online]. Available: <https://doi.org/10.1145/3025453.3025673>
- [408] E. M. Redmiles, S. Kross, and M. L. Mazurek, “How well do my results generalize? comparing security and privacy survey results from mturk, web, and telephone samples,” in *2019 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2019, pp. 1326–1343. [Online]. Available: <https://doi.org/10.1109/SP.2019.00014>

- [409] E. M. Redmiles, Z. Zhu, S. Kross, D. Kuchhal, T. Dumitras, and M. L. Mazurek, "Asking for a friend: Evaluating response biases in security user studies," in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1238–1255. [Online]. Available: <https://doi.org/10.1145/3243734.3243740>
- [410] H. Regenbrecht and T. Schubert, "Real and illusory interactions enhance presence in virtual environments," *Presence: Teleoperators & Virtual Environments*, vol. 11, no. 4, pp. 425–434, 2002. [Online]. Available: <https://doi.org/10.1162/105474602760204318>
- [411] M. S. Remland, T. S. Jones, and H. Brinkman, "Interpersonal distance, body orientation, and touch: Effects of culture, gender, and age," *The Journal of Social Psychology*, vol. 135, no. 3, pp. 281–297, 1995, PMID: 7650932. [Online]. Available: <https://doi.org/10.1080/00224545.1995.9713958>
- [412] H. Rex Hartson, "Human–computer interaction: Interdisciplinary roots and trends," *Journal of Systems and Software*, vol. 43, no. 2, pp. 103–118, 1998. [Online]. Available: [https://doi.org/10.1016/S0164-1212\(98\)10026-2](https://doi.org/10.1016/S0164-1212(98)10026-2)
- [413] K. A. Reynolds, P. M. Watt, S. A. Boone, and C. P. Gerba, "Occurrence of bacteria and biochemical markers on public surfaces," *International journal of environmental health research*, vol. 15, no. 3, pp. 225–234, 2005. [Online]. Available: <https://doi.org/10.1080/09603120500115298>
- [414] G. M. Rhineberger, D. J. Hartmann, and T. L. Van Valey, "Triangulated research designs—a justification?" *Journal of Applied Sociology*, no. 1, pp. 56–66, 2005.
- [415] S. Riches, S. Elghany, P. Garety, M. Rus-Calafell, and L. Valmaggia, "Factors affecting sense of presence in a virtual reality social environment: a qualitative study," *Cyberpsychology, Behavior, and Social Networking*, vol. 22, no. 4, pp. 288–292, 2019.
- [416] J. Rico and S. Brewster, "Usable gestures for mobile interfaces: Evaluating social acceptability," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '10. New York, NY, USA: Association for Computing Machinery, 2010, p. 887–896. [Online]. Available: <https://doi.org/10.1145/1753326.1753458>
- [417] A. Riener, "Assessment of simulator fidelity and validity in simulator and on-the-road studies," *International Journal On Advances in Systems and Measurements*, vol. 3, pp. 110–124 (15), 02 2011.

- [418] G. Robertson, S. Card, and J. Mackinlay, “Three views of virtual reality: nonimmersive virtual reality,” *Computer*, vol. 26, 1993. [Online]. Available: <https://doi.org/10.1109/2.192002>
- [419] C. Robson, *Real world research: A resource for social scientists and practitioner-researchers*. Wiley-Blackwell, 2002, https://books.google.co.uk/books/about/Real_World_Research.html?id=DkplMcAysFQC, accessed 18 February 2023.
- [420] RockVR. (2020) Video capture by rockvr. <https://assetstore.unity.com/publishers/24830>, accessed 18 February 2023. [Online]. Available: <https://assetstore.unity.com/publishers/24830>
- [421] T. Rodden, K. Cheverst, K. Davies, and A. Dix, “Exploiting context in hci design for mobile systems,” in *Workshop on human computer interaction with mobile devices*, vol. 12. Citeseer, 1998, <https://www.alandix.com/academic/papers/exploiting-context-1998/>, accessed 18 February 2023.
- [422] F. Roesner, T. Kohno, and D. Molnar, “Security and privacy for augmented reality systems,” *Commun. ACM*, vol. 57, no. 4, p. 88–96, Apr. 2014. [Online]. Available: <https://doi.org/10.1145/2580723.2580730>
- [423] S. Rosenbaum, G. Cockton, K. Coyne, M. Muller, and T. Rauch, “Focus groups in hci: Wealth of information or waste of resources?” in *CHI '02 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '02. New York, NY, USA: Association for Computing Machinery, 2002, p. 702–703. [Online]. Available: <https://doi.org/10.1145/506443.506554>
- [424] D. Roth, J.-L. Lugrin, D. Galakhov, A. Hofmann, G. Bente, M. E. Latoschik, and A. Fuhrmann, “Avatar realism and social interaction quality in virtual reality,” in *2016 IEEE Virtual Reality (VR)*, 2016, pp. 277–278. [Online]. Available: <https://doi.org/10.1109/VR.2016.7504761>
- [425] 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*, ser. CCS '04. New York, NY, USA: ACM, 2004, p. 236–245. [Online]. Available: <https://doi.org/10.1145/1030083.1030116>
- [426] E. Rukzio, “Physical mobile interactions: Mobile devices as pervasive mediators for interactions with the real world,” Ph.D. dissertation, University of Munich, 2006.
- [427] S. Ruoti and K. Seamons, “Standard metrics and scenarios for usable authentication,” in *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*. Denver,

- CO: USENIX Association, Jun. 2016. [Online]. Available: https://www.usenix.org/conference/soups2016/workshop-program/way2016/presentation/ruoti_metrics
- [428] A. Saad, J. Liebers, U. Gruenefeld, F. Alt, and S. Schneegass, “Understanding bystanders’ tendency to shoulder surf smartphones using 360-degree videos in virtual reality,” 2021. [Online]. Available: <https://doi.org/10.1145/3447526.3472058>
- [429] N. Sae-Bae, K. Ahmed, K. Isbister, and N. Memon, “Biometric-rich gestures: A novel approach to authentication on multi-touch devices,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’12. New York, NY, USA: Association for Computing Machinery, 2012, p. 977–986. [Online]. Available: <https://doi.org/10.1145/2207676.2208543>
- [430] P. Saeghe, B. Weir, M. McGill, S. Clinch, and R. Stevens, “Augmenting a nature documentary with a lifelike hologram in virtual reality,” in *ACM International Conference on Interactive Media Experiences*, ser. IMX ’22. New York, NY, USA: Association for Computing Machinery, 2022, p. 275–280. [Online]. Available: <https://doi.org/10.1145/3505284.3532974>
- [431] D. Saffo, S. Bartolomeo, C. Yildirim, and C. Dunne, “Remote and collaborative virtual reality experiments via social vr platforms,” in *Proceedings of the 39th Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI ’21. New York, NY, USA: ACM, 2021. [Online]. Available: <https://doi.org/10.1145/3411764.3445426>
- [432] S. Safikhani, M. Holly, A. Kainz, and J. Pirker, “The influence of in-vr questionnaire design on the user experience,” in *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*, ser. VRST ’21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: <https://doi.org/10.1145/3489849.3489884>
- [433] J. Saldaña, *The coding manual for qualitative researchers*. Sage, 2015, <https://us.sagepub.com/en-us/nam/book/coding-manual-qualitative-researchers-1>, accessed 18 February 2023.
- [434] H. Sasamoto, N. Christin, and E. Hayashi, “Undercover: Authentication usable in front of prying eyes,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’08. New York, NY, USA: Association for Computing Machinery, 2008, p. 183–192. [Online]. Available: <https://doi.org/10.1145/1357054.1357085>
- [435] A. M. Sasse, S. Brostoff, and D. Weirich, “Transforming the ‘weakest link’ — a human/computer interaction approach to usable and effective security,” *BT*

- Technology Journal*, vol. 19, no. 3, pp. 122–131, Jul 2001. [Online]. Available: <https://doi.org/10.1023/A:1011902718709>
- [436] M. A. Sasse, “Red-eye blink, bendy shuffle, and the yuck factor: A user experience of biometric airport systems,” *IEEE S & P*, pp. 78–81, 2007. [Online]. Available: <https://doi.org/10.1109/MSP.2007.69>
- [437] M. A. Sasse and I. Flechais, “Usable security: Why do we need it? how do we get it?” O’Reilly, 2005, <https://discovery.ucl.ac.uk/id/eprint/20345/>, accessed 18 February 2023.
- [438] M. A. Sasse, M. Steves, K. Krol, and D. Chisnell, “The great authentication fatigue – and how to overcome it,” in *Cross-Cultural Design*, P. L. P. Rau, Ed. Cham: Springer International Publishing, 2014. [Online]. Available: https://doi.org/10.1007/978-3-319-07308-8_23
- [439] G.-L. Savino, N. Emanuel, S. Kowalzik, F. Kroll, M. C. Lange, M. Laudan, R. Leder, Z. Liang, D. Markhabayeva, M. Schmeißer, N. Schütz, C. Stellmacher, Z. Xu, K. Bub, T. Kluss, J. Maldonado, E. Kruijff, and J. Schöning, “Comparing pedestrian navigation methods in virtual reality and real life,” in *2019 International Conference on Multimodal Interaction*, ser. ICMI ’19. New York, NY, USA: Association for Computing Machinery, 2019, p. 16–25. [Online]. Available: <https://doi.org/10.1145/3340555.3353741>
- [440] Y. Sawaya, M. Sharif, N. Christin, A. Kubota, A. Nakarai, and A. Yamada, “Self-confidence trumps knowledge: A cross-cultural study of security behavior,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’17. New York, NY, USA: Association for Computing Machinery, 2017, p. 2202–2214. [Online]. Available: <https://doi.org/10.1145/3025453.3025926>
- [441] F. Schaub, B. Könings, P. Lang, B. Wiedersheim, C. Winkler, and M. Weber, “Prical: Context-adaptive privacy in ambient calendar displays,” in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp ’14. New York, NY, USA: Association for Computing Machinery, 2014, p. 499–510. [Online]. Available: <https://doi.org/10.1145/2632048.2632087>
- [442] F. Schaub, M. Walch, B. Könings, and M. Weber, “Exploring the design space of graphical passwords on smartphones,” in *Proceedings of the Ninth Symposium on Usable Privacy and Security*, ser. SOUPS ’13. New York, NY, USA: Association for Computing Machinery, 2013. [Online]. Available: <https://doi.org/10.1145/2501604.2501615>

- [443] S. E. Schechter, R. Dhamija, A. Ozment, and I. Fischer, “The emperor’s new security indicators,” in *2007 IEEE Symposium on Security and Privacy (SP’07)*. IEEE, 2007, pp. 51–65. [Online]. Available: <https://doi.org/10.1109/SP.2007.35>
- [444] E. Schmider, M. Ziegler, E. Danay, L. Beyer, and M. Bühner, “Is it really robust? reinvestigating the robustness of anova against violations of the normal distribution assumption.” *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, vol. 6, no. 4, p. 147, 2010. [Online]. Available: <https://psycnet.apa.org/doi/10.1027/1614-2241/a000016>
- [445] A. Schmidt, F. Alt, and V. Mäkelä, *Evaluation in Human-Computer Interaction – Beyond Lab Studies*. ACM, 2021.
- [446] A. Schmidt, M. Beigl, and H.-W. Gellersen, “There is more to context than location,” *Computers & Graphics*, vol. 23, no. 6, pp. 893–901, 1999. [Online]. Available: [https://doi.org/10.1016/S0097-8493\(99\)00120-X](https://doi.org/10.1016/S0097-8493(99)00120-X)
- [447] S. Schmidt, “Shall we really do it again? the powerful concept of replication is neglected in the social sciences,” *Review of General Psychology*, vol. 13, no. 2, pp. 90–100, 2009. [Online]. Available: <https://doi.org/10.1037/a0015108>
- [448] S. Schneegass, Y. Oualil, and A. Bulling, “Skullconduct: Biometric user identification on eyewear computers using bone conduction through the skull,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’16. New York, NY, USA: Association for Computing Machinery, 2016, p. 1379–1384. [Online]. Available: <https://doi.org/10.1145/2858036.2858152>
- [449] M. Schrepp. (2019) User experience questionnaire handbook. <https://www.ueq-online.org/Material/Handbook.pdf>, accessed 18 February 2023.
- [450] H. Schrom-Feiertag, G. Regal, J. Puthenkalam, and S. Suetterle, “Immersive experience prototyping: Using mixed reality to integrate real devices in virtual simulated contexts to prototype experiences with mobile apps,” in *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, 2021, pp. 75–81. [Online]. Available: <https://doi.org/10.1109/ISMAR-Adjunct54149.2021.00025>
- [451] T. Schubert, F. Friedmann, and H. Regenbrecht, “The experience of presence: Factor analytic insights,” *Presence: Teleoperators & Virtual Environments*, vol. 10, no. 3, pp. 266–281, 2001. [Online]. Available: <https://doi.org/10.1162/105474601300343603>
- [452] T. W. Schubert, “The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness.” *Z.*

- für Medienpsychologie*, vol. 15, no. 2, pp. 69–71, 2003. [Online]. Available: <http://dx.doi.org/10.1026//1617-6383.15.2.69>
- [453] V. Schwind, P. Knierim, N. Haas, and N. Henze, “Using presence questionnaires in virtual reality,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–12. [Online]. Available: <https://doi.org/10.1145/3290605.3300590>
- [454] T. Seitz and H. Hussmann, “Pasdjo: Quantifying password strength perceptions with an online game,” in *Proceedings of the 29th Australian Conference on Computer-Human Interaction*, ser. OZCHI ’17. New York, NY, USA: Association for Computing Machinery, 2017, p. 117–125. [Online]. Available: <https://doi.org/10.1145/3152771.3152784>
- [455] T. Seitz, F. Mathis, and H. Hussmann, “The bird is the word: A usability evaluation of emojis inside text passwords,” in *Proceedings of the 29th Australian Conference on Computer-Human Interaction*, ser. OZCHI ’17. New York, NY, USA: ACM, 2017, p. 10–20. [Online]. Available: <https://doi.org/10.1145/3152771.3152773>
- [456] M. Seymour, K. Riemer, and J. Kay, “Interactive realistic digital avatars-revisiting the uncanny valley,” 2017.
- [457] R. Sharp, J. Scott, and A. R. Beresford, “Secure mobile computing via public terminals,” in *International Conference on Pervasive Computing*. Springer, 2006, pp. 238–253. [Online]. Available: https://doi.org/10.1007/11748625_15
- [458] I. Shatz, “Fast, free, and targeted: Reddit as a source for recruiting participants online,” *Social Science Computer Review*, vol. 35, no. 4, pp. 537–549, 2017. [Online]. Available: <https://doi.org/10.1177/0894439316650163>
- [459] Y. Shen, H. Wen, C. Luo, W. Xu, T. Zhang, W. Hu, and D. Rus, “Gaitlock: Protect virtual and augmented reality headsets using gait,” *IEEE Transactions on Dependable and Secure Computing*, vol. 16, no. 3, pp. 484–497, May 2019. [Online]. Available: <https://doi.org/10.1109/TDSC.2018.2800048>
- [460] J. Shigeyama, T. Hashimoto, S. Yoshida, T. Narumi, T. Tanikawa, and M. Hirose, “Transcalibur: A weight shifting virtual reality controller for 2d shape rendering based on computational perception model,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–11. [Online]. Available: <https://doi.org/10.1145/3290605.3300241>

- [461] B. Shin, S. Kim, V. Druzhin, P. Malinina, S. Dubynin, A. Bolotova, S. Kopenkin, A. Putilin, W. Seo, C.-K. Lee *et al.*, “Compact augmented-reality glasses using holographic optical element combiner,” in *Practical Holography XXXIII: Displays, Materials, and Applications*, vol. 10944. SPIE, 2019, pp. 93–99. [Online]. Available: <https://doi.org/10.1117/12.2507339>
- [462] B. Shneiderman, C. Plaisant, M. Cohen, S. Jacobs, N. Elmqvist, and N. Diakopoulos, *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, 6th ed. Pearson, 2017, https://books.google.co.uk/books/about/Designing_the_User_Interface.html?id=nhDYtQEACAAJ&redir_esc=y, accessed 18 February 2023.
- [463] L. Sidenmark and H. Gellersen, “Eye, head and torso coordination during gaze shifts in virtual reality,” *ACM Trans. Comput.-Hum. Interact.*, vol. 27, no. 1, Dec. 2019. [Online]. Available: <https://doi.org/10.1145/3361218>
- [464] L. Sidenmark and H. Gellersen, “Eye&head: Synergetic eye and head movement for gaze pointing and selection,” in *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, ser. UIST ’19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1161–1174. [Online]. Available: <https://doi.org/10.1145/3332165.3347921>
- [465] L. Sidenmark, D. Potts, B. Bapisch, and H. Gellersen, “Radi-eye: Hands-free radial interfaces for 3d interaction using gaze-activated head-crossing,” in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: <https://doi.org/10.1145/3411764.3445697>
- [466] S. Siltanen, H. Heinonen, A. Burova, P. B. Palma, P. Truong, V. Opas, and M. Turunen, “There is always a way: Organizing vr user tests with remote and hybrid setups during a pandemic—learnings from five case studies,” *Multimodal Technologies and Interaction*, vol. 5, no. 10, 2021. [Online]. Available: <https://www.mdpi.com/2414-4088/5/10/62>
- [467] A. L. Simeone, R. Cools, S. Depuydt, J. a. M. Gomes, P. Goris, J. Grocott, A. Esteves, and K. Gerling, “Immersive speculative enactments: Bringing future scenarios and technology to life using virtual reality,” in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’22. New York, NY, USA: Association for Computing Machinery, 2022. [Online]. Available: <https://doi.org/10.1145/3491102.3517492>
- [468] A. L. Simeone, N. C. Nilsson, A. Zenner, M. Speicher, and F. Daiber, “The space bender: Supporting natural walking via overt manipulation of the virtual environment,”

- in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2020, pp. 598–606.
- [469] M. Sinclair, M. Pahud, and H. Benko, “Touchmover: Actuated 3d touchscreen with haptic feedback,” in *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces*, ser. ITS '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 287–296. [Online]. Available: <https://doi.org/10.1145/2512349.2512805>
- [470] Y. N. Singh and S. K. Singh, “Evaluation of electrocardiogram for biometric authentication,” 2011. [Online]. Available: <http://dx.doi.org/10.4236/jis.2012.31005>
- [471] R. Skarbes, M. Smith, and M. Whitton, “Mixed reality doesn’t need standardized evaluation methods,” 2021. [Online]. Available: <http://dx.doi.org/10.13140/RG.2.2.14305.02407>
- [472] R. Skarbez, F. P. Brooks, Jr., and M. C. Whitton, “A survey of presence and related concepts,” *ACM Comput. Surv.*, vol. 50, no. 6, Nov. 2017. [Online]. Available: <https://doi.org/10.1145/3134301>
- [473] R. Skarbez, J. Gabbard, D. A. Bowman, T. Ogle, and T. Tucker, “Virtual replicas of real places: Experimental investigations,” *IEEE Transactions on Visualization and Computer Graphics*, 2021. [Online]. Available: <https://doi.org/10.1109/TVCG.2021.3096494>
- [474] M. Slater, “A note on presence terminology,” *Presence connect*, vol. 3, no. 3, pp. 1–5, 2003, http://www0.cs.ucl.ac.uk/research/vr/Projects/Presencia/ConsortiumPublications/ucl_cs_papers/presence-terminology.htm, accessed 18 February 2023.
- [475] M. Slater, “Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1535, pp. 3549–3557, 2009. [Online]. Available: <https://doi.org/10.1098/rstb.2009.0138>
- [476] M. Slater, D. Banakou, A. Beacco, J. Gallego, F. Macia-Varela, and R. Oliva, “A separate reality: An update on place illusion and plausibility in virtual reality,” *Frontiers in Virtual Reality*, p. 81, 2022. [Online]. Available: <https://doi.org/10.3389/frvir.2022.914392>
- [477] M. Slater, P. Khanna, J. Mortensen, and I. Yu, “Visual realism enhances realistic response in an immersive virtual environment,” *IEEE computer graphics*

- and applications*, vol. 29, no. 3, pp. 76–84, 2009. [Online]. Available: <https://doi.org/10.1162/105474601753272844>
- [478] M. Slater and A. Steed, “A virtual presence counter,” *Presence*, vol. 9, no. 5, pp. 413–434, 2000. [Online]. Available: <http://dx.doi.org/10.1162/105474600566925>
- [479] M. Slater and M. Usoh, “Body centred interaction in immersive virtual environments,” *Artificial life and virtual reality*, vol. 1, no. 1994, 1994, <https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.109.7613>, accessed 18 February 2023.
- [480] O. B. Software. (2012) Obs (open broadcaster software) — free and open source software. <https://obsproject.com/de>, accessed 18 February 2023.
- [481] P. Song, W. B. Goh, W. Hutama, C.-W. Fu, and X. Liu, “A handle bar metaphor for virtual object manipulation with mid-air interaction,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’12. New York, NY, USA: Association for Computing Machinery, 2012, p. 1297–1306. [Online]. Available: <https://doi.org/10.1145/2207676.2208585>
- [482] A. Sotirakopoulos, K. Hawkey, and K. Beznosov, “On the challenges in usable security lab studies: Lessons learned from replicating a study on ssl warnings,” in *Proceedings of the Seventh Symposium on Usable Privacy and Security*, ser. SOUPS ’11. New York, NY, USA: Association for Computing Machinery, 2011. [Online]. Available: <https://doi.org/10.1145/2078827.2078831>
- [483] R. Spears and M. Lea, “Panacea or panopticon? the hidden power in computer-mediated communication,” *Communication research*, vol. 21, no. 4, pp. 427–459, 1994. [Online]. Available: <https://doi.org/10.1177/009365094021004001>
- [484] M. Speicher, B. D. Hall, and M. Nebeling, “What is mixed reality?” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–15. [Online]. Available: <https://doi.org/10.1145/3290605.3300767>
- [485] S. Spiekermann, J. Grossklags, and B. Berendt, “E-privacy in 2nd generation e-commerce: Privacy preferences versus actual behavior,” in *Proceedings of the 3rd ACM Conference on Electronic Commerce*, ser. EC ’01. New York, NY, USA: Association for Computing Machinery, 2001, p. 38–47. [Online]. Available: <https://doi.org/10.1145/501158.501163>
- [486] A. Steed, S. Frlston, M. M. Lopez, J. Drummond, Y. Pan, and D. Swapp, “An ‘in the wild’ experiment on presence and embodiment using consumer virtual reality equipment,” *IEEE Transactions on Visualization and Computer*

- Graphics*, vol. 22, no. 4, pp. 1406–1414, 2016. [Online]. Available: <https://doi.org/10.1109/TVCG.2016.2518135>
- [487] A. Steed, L. Izzouzi, K. Brandstätter, S. Friston, B. Congdon, O. Olkkonen, D. Giunchi, N. Numan, and D. Swapp, “Ubiq-exp: A toolkit to build and run remote and distributed mixed reality experiments,” *Frontiers in Virtual Reality*, vol. 3, 2022. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/frvir.2022.912078>
- [488] A. Steed, F. Ortega, A. Williams, E. Kruijff, W. Stuerzlinger, A. Batmaz, A. Won, E. Rosenberg, A. Simeone, and A. Hayes, “Evaluating immersive experiences during covid-19 and beyond,” *Interactions*, vol. 27, no. 4, p. 62–67, jul 2020. [Online]. Available: <https://doi.org/10.1145/3406098>
- [489] A. Steed, Y. Pan, F. Zisch, and W. Steptoe, “The impact of a self-avatar on cognitive load in immersive virtual reality,” in *2016 IEEE Virtual Reality (VR)*, 2016, pp. 67–76. [Online]. Available: <https://doi.org/10.1109/VR.2016.7504689>
- [490] A. Steed and R. Schroeder, “Collaboration in immersive and non-immersive virtual environments,” in *Immersed in Media*. Springer, 2015, pp. 263–282. [Online]. Available: https://doi.org/10.1007/978-3-319-10190-3_11
- [491] J. Steuer, “Defining virtual reality: Dimensions determining telepresence,” *Journal of communication*, vol. 42, no. 4, pp. 73–93, 1992. [Online]. Available: <https://doi.org/10.1111/j.1460-2466.1992.tb00812.x>
- [492] M. Stokkenes, R. Ramachandra, and C. Busch, “Biometric authentication protocols on smartphones: An overview,” in *Proceedings of the 9th International Conference on Security of Information and Networks*, ser. SIN ’16. New York, NY, USA: ACM, 2016, pp. 136–140. [Online]. Available: <https://doi.org/10.1145/2947626.2951962>
- [493] U. A. Store. (2022) Unity asset store: 3d. <https://assetstore.unity.com/3d>, accessed 18 February 2023.
- [494] S. Sultana, M. Deb, A. Bhattacharjee, S. Hasan, S. Alam, T. Chakraborty, P. Roy, S. F. Ahmed, A. Moitra, M. A. Amin, A. N. Islam, and S. I. Ahmed, “‘unmochon’: A tool to combat online sexual harassment over facebook messenger,” in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: <https://doi.org/10.1145/3411764.3445154>
- [495] S.-T. Sun, E. Pospisil, I. Muslukhov, N. Dindar, K. Hawkey, and K. Beznosov, “What makes users refuse web single sign-on? an empirical investigation of openid,” in *Proceedings of the Seventh Symposium on Usable Privacy and Security*, ser. SOUPS

- '11. New York, NY, USA: Association for Computing Machinery, 2011. [Online]. Available: <https://doi.org/10.1145/2078827.2078833>
- [496] J. Sunshine, S. Egelman, H. Almuhiemedi, N. Atri, and L. F. Cranor, "Crying wolf: An empirical study of ssl warning effectiveness," in *Proceedings of the 18th Conference on USENIX Security Symposium*, ser. SSYM'09. USA: USENIX Association, 2009, p. 399–416, https://www.usenix.org/events/sec09/tech/full_papers/sunshine.pdf, accessed 18 February 2023.
- [497] H. B. Surale, A. Gupta, M. Hancock, and D. Vogel, "Tabletinvr: Exploring the design space for using a multi-touch tablet in virtual reality," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–13. [Online]. Available: <https://doi.org/10.1145/3290605.3300243>
- [498] H. B. Surale, F. Matulic, and D. Vogel, "Experimental analysis of barehand mid-air mode-switching techniques in virtual reality," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–14. [Online]. Available: <https://doi.org/10.1145/3290605.3300426>
- [499] S. W. Sussman and L. Sproull, "Straight talk: Delivering bad news through electronic communication," *Information Systems Research*, vol. 10, no. 2, pp. 150–166, 1999. [Online]. Available: <https://doi.org/10.1287/isre.10.2.150>
- [500] I. E. Sutherland, "Sketchpad a man-machine graphical communication system," *Simulation*, vol. 2, no. 5, pp. R–3, 1964. [Online]. Available: <https://doi.org/10.1145/800265.810742>
- [501] I. E. Sutherland, "A head-mounted three dimensional display," in *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, 1968, pp. 757–764. [Online]. Available: <https://doi.org/10.1145/1476589.1476686>
- [502] Z. Syed, S. Banerjee, Q. Cheng, and B. Cukic, "Effects of user habituation in keystroke dynamics on password security policy," in *2011 IEEE 13th International Symposium on High-Assurance Systems Engineering*. IEEE, 2011, pp. 352–359. [Online]. Available: <https://doi.org/10.1109/HASE.2011.16>
- [503] M. Tahaei, A. Frik, and K. Vaniea, "Privacy champions in software teams: Understanding their motivations, strategies, and challenges," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI '21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: <https://doi.org/10.1145/3411764.3445768>

- [504] M. Tahaei and K. Vaniea, "Recruiting participants with programming skills: A comparison of four crowdsourcing platforms and a cs student mailing list," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 1–16. [Online]. Available: <https://doi.org/10.1145/3491102.3501957>
- [505] D. S. Tan, P. Keyani, and M. Czerwinski, "Spy-resistant keyboard: more secure password entry on public touch screen displays," in *Proceedings of the 17th Australia conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future*. Citeseer, 2005, pp. 1–10, <https://www.microsoft.com/en-us/research/publication/spy-resistant-keyboard-towards-more-secure-password-entry-on-publicly-observable-touch-screens/>, accessed 18 February 2023.
- [506] F. Tari, A. A. Ozok, and S. H. Holden, "A comparison of perceived and real shoulder-surfing risks between alphanumeric and graphical passwords," in *Proceedings of the Second Symposium on Usable Privacy and Security*, ser. SOUPS '06. New York, NY, USA: Association for Computing Machinery, 2006, p. 56–66. [Online]. Available: <https://doi.org/10.1145/1143120.1143128>
- [507] M. J. Tarr and W. H. Warren, "Virtual reality in behavioral neuroscience and beyond," *Nature neuroscience*, vol. 5, no. 11, pp. 1089–1092, 2002. [Online]. Available: <https://doi.org/10.1038/nn948>
- [508] J. Taylor. (2019) Major breach found in biometrics system used by banks, uk police and defence firms. <https://www.theguardian.com/technology/2019/aug/14/major-breach-found-in-biometrics-system-used-by-banks-uk-police-and-defence-firms>, accessed 18 February 2023.
- [509] U. Technologies. (2022) Unity developer documentation: Start bringing your vision to life today with the unity real-time 3d development platform. <https://docs.unity3d.com/ScriptReference/Collider.html>, accessed 18 February 2023.
- [510] T. Technology. (2020) Tobii htc vive devkit. <https://vr.tobii.com/sdk/develop/unity/getting-started/tobii-htc-dev-kit/>, accessed 18 February 2023.
- [511] S.-Y. Teng, T.-S. Kuo, C. Wang, C.-h. Chiang, D.-Y. Huang, L. Chan, and B.-Y. Chen, "Pupop: Pop-up prop on palm for virtual reality," in *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 5–17. [Online]. Available: <https://doi.org/10.1145/3242587.3242628>

- [512] TI. (2021) Ultimatereplay. <https://assetstore.unity.com/packages/tools/camera/ultimate-replay-2-0-178602>, accessed 18 February 2023.
- [513] G. W. Tigwell, B. M. Gorman, and R. Menzies, “Emoji accessibility for visually impaired people,” in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1–14. [Online]. Available: <https://doi.org/10.1145/3313831.3376267>
- [514] Tobii. (2022) Tobii pro vr integration. <https://www.tobii.com/product-listing/vr-integration/>, accessed 18 February 2023.
- [515] E. Toch, Y. Wang, and L. Cranor, “Personalization and privacy: A survey of privacy risks and remedies in personalization-based systems,” *User Modeling and User-Adapted Interaction*, vol. 22, 04 2012. [Online]. Available: <https://doi.org/10.1007/s11257-011-9110-z>
- [516] D. Trinitatova and D. Tsetserukou, “Touchvr: A wearable haptic interface for vr aimed at delivering multi-modal stimuli at the user’s palm,” in *SIGGRAPH Asia 2019 XR*, ser. SA ’19. New York, NY, USA: Association for Computing Machinery, 2019, p. 42–43. [Online]. Available: <https://doi.org/10.1145/3355355.3361896>
- [517] C. W. Turner, J. R. Lewis, and J. Nielsen, “Determining usability test sample size,” *International encyclopedia of ergonomics and human factors*, vol. 3, no. 2, pp. 3084–3088, 2006, <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.110.2521&rep=rep1&type=pdf>, accessed 18 February 2023.
- [518] I. T. Union. (2020) Internet surge slows, leaving 2.7 billion people offline in 2022. <https://www.itu.int/en/mediacentre/Pages/PR-2022-09-16-Internet-surge-slows.aspx>, accessed 18 February 2023.
- [519] 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*, ser. CHI ’17. New York, NY, USA: Association for Computing Machinery, 2017, p. 3775–3786. [Online]. Available: <https://doi.org/10.1145/3025453.3026050>
- [520] M. Usoh, E. Catena, S. Arman, and M. Slater, “Using presence questionnaires in reality,” *Presence*, vol. 9, no. 5, pp. 497–503, 2000. [Online]. Available: <http://dx.doi.org/10.1162/105474600566989>

- [521] C. Vandeweerd, T. Luong, M. Atchapero, A. Mottelson, C. Holz, G. Makransky, and R. Böhm, “Virtual reality reduces covid-19 vaccine hesitancy in the wild: a randomized trial,” *Scientific Reports*, vol. 12, no. 1, pp. 1–7, 2022. [Online]. Available: <https://doi.org/10.1038/s41598-022-08120-4>
- [522] S. Ventura, E. Brivio, G. Riva, and R. M. Baños, “Immersive versus non-immersive experience: Exploring the feasibility of memory assessment through 360 technology,” *Frontiers in psychology*, vol. 10, p. 2509, 2019. [Online]. Available: <https://doi.org/10.3389/fpsyg.2019.02509>
- [523] I. Verhulst, A. Woods, L. Whittaker, J. Bennett, and P. Dalton, “Do vr and ar versions of an immersive cultural experience engender different user experiences?” *Computers in Human Behavior*, vol. 125, p. 106951, 2021. [Online]. Available: <https://doi.org/10.1016/j.chb.2021.106951>
- [524] Verizon. (2020) 2020 data breach investigations report. [Online]. Available: <https://www.verizon.com/business/resources/reports/2020-data-breach-investigations-report.pdf>
- [525] 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*, ser. UbiComp '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 439–448. [Online]. Available: <https://doi.org/10.1145/2493432.2493477>
- [526] V. Vinayagamoorthy, A. Brogni, M. Gillies, M. Slater, and A. Steed, “An investigation of presence response across variations in visual realism,” in *The 7th Annual International Presence Workshop*. Citeseer, 2004, pp. 148–155, <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.492.5517&rep=rep1&type=pdf>, accessed 18 February 2023.
- [527] V. Vinayagamoorthy, A. Steed, and M. Slater, “Building characters: Lessons drawn from virtual environments,” in *Proceedings of Toward Social Mechanisms of Android Science: A CogSci 2005 Workshop*. COGSCI 2005, 2005, pp. 119–126, <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.573.3047&rep=rep1&type=pdf>, accessed 18 February 2023.
- [528] A. Voit, S. Mayer, V. Schwind, and N. Henze, “Online, vr, ar, lab, and in-situ: Comparison of research methods to evaluate smart artifacts,” in *Proc. of the 2019 CHI Conf. on Human Factors in Comp. Systems*, ser. CHI '19. New York, NY, USA: ACM, 2019, p. 1–12. [Online]. Available: <https://doi.org/10.1145/3290605.3300737>

- [529] M. Volkamer, A. Gutmann, K. Renaud, P. Gerber, and P. Mayer, “Replication study: A cross-country field observation study of real world PIN usage at ATMs and in various electronic payment scenarios,” in *Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018)*. Baltimore, MD: USENIX Association, Aug. 2018, pp. 1–11, <https://www.usenix.org/conference/soups2018/presentation/volkamer>, accessed 18 February 2023.
- [530] 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*, ser. CHI '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 1403–1406. [Online]. Available: <https://doi.org/10.1145/2702123.2702212>
- [531] 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*, ser. MobileHCI '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 261–270. [Online]. Available: <https://doi.org/10.1145/2493190.2493231>
- [532] E. von Zezschwitz, A. Koslow, A. De Luca, and H. Hussmann, “Making graphic-based authentication secure against smudge attacks,” in *Proceedings of the 2013 International Conference on Intelligent User Interfaces*, ser. IUI '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 277–286. [Online]. Available: <https://doi.org/10.1145/2449396.2449432>
- [533] C. Wacharamanotham, L. Eisenring, S. Haroz, and F. Echtler, “Transparency of CHI Research Artifacts: Results of a Self-Reported Survey,” in *Proceedings of the 38th Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '20. New York, NY, USA: ACM, 2020. [Online]. Available: <http://dx.doi.org/10.1145/3313831.3376448>
- [534] N. Wagener, J. Niess, Y. Rogers, and J. Schöning, “Mood worlds: A virtual environment for autonomous emotional expression,” in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, ser. CHI '22. New York, NY, USA: Association for Computing Machinery, 2022. [Online]. Available: <https://doi.org/10.1145/3491102.3501861>
- [535] C.-H. Wang, S. Yong, H.-Y. Chen, Y.-S. Ye, and L. Chan, “Hmd light: Sharing in-vr experience via head-mounted projector for asymmetric interaction,” in *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. New

- York, NY, USA: Association for Computing Machinery, 2020, p. 472–486. [Online]. Available: <https://doi.org/10.1145/3379337.3415847>
- [536] Y. Wang, H. Xia, Y. Yao, and Y. Huang, “Flying eyes and hidden controllers: A qualitative study of people’s privacy perceptions of civilian drones in the us,” *Proc. on Privacy Enhancing Tech.*, vol. 2016, no. 3, 2016. [Online]. Available: <http://dx.doi.org/10.1515/popets-2016-0022>
- [537] K. Watson, R. Bretin, M. Khamis, and F. Mathis, “The feet in human-centred security: Investigating foot-based user authentication for public displays,” 2022. [Online]. Available: <https://doi.org/10.1145/3491101.3519838>
- [538] M. Weiser, “The computer for the 21 st century,” *Scientific american*, vol. 265, no. 3, pp. 94–105, 1991, <http://www.jstor.org/stable/24938718>, accessed 18 February 2023.
- [539] M. Weiß, K. Angerbauer, A. Voit, M. Schwarzl, M. Sedlmair, and S. Mayer, “Revisited: Comparison of empirical methods to evaluate visualizations supporting crafting and assembly purposes,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 2, pp. 1204–1213, 2020. [Online]. Available: <https://doi.org/10.1109/TVCG.2020.3030400>
- [540] 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*, ser. NordiCHI ’08. New York, NY, USA: ACM, 2008, p. 383–392. [Online]. Available: <https://doi.org/10.1145/1463160.1463202>
- [541] J. Whiteside, J. Bennett, and K. Holtzblatt, “Usability engineering: Our experience and evolution,” in *Handbook of Human-Computer Interaction*, M. HELANDER, Ed. Amsterdam: North-Holland, 1988, pp. 791–817. [Online]. Available: <https://doi.org/10.1016/B978-0-444-70536-5.50041-5>
- [542] A. Whitten and J. D. Tygar, “Why Johnny Can’t Encrypt: A Usability Evaluation of PGP 5.0,” in *Proceedings of the 8th Conference on USENIX Security Symposium - Volume 8*, ser. SSYM’99. USA: USENIX Association, 1999, p. 14, https://www.usenix.org/legacy/events/sec99/full_papers/whitten/whitten_html/index.html, accessed 18 February 2023.
- [543] S. Wiedenbeck, J. Waters, J.-C. Birget, A. Brodskiy, and N. Memon, “Authentication using graphical passwords: Effects of tolerance and image choice,” in *Proceedings of the 2005 Symposium on Usable Privacy and Security*, ser. SOUPS ’05. New York, NY, USA: Association for Computing Machinery, 2005, p. 1–12. [Online]. Available: <https://doi.org/10.1145/1073001.1073002>

- [544] O. Wiese and V. Roth, "Pitfalls of shoulder surfing studies." NDSS Workshop on Usable Security, 2015. [Online]. Available: <https://doi.org/10.14722/usec.2015.23007>
- [545] O. Wiese and V. Roth, "See you next time: A model for modern shoulder surfers," in *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, ser. MobileHCI '16. New York, NY, USA: ACM, 2016, p. 453–464. [Online]. Available: <https://doi.org/10.1145/2935334.2935388>
- [546] J. Williamson, J. Li, V. Vinayagamoorthy, D. A. Shamma, and P. Cesar, "Proxemics and social interactions in an instrumented virtual reality workshop," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI '21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: <https://doi.org/10.1145/3411764.3445729>
- [547] J. R. Williamson, "User experience, performance, and social acceptability: usable multimodal mobile interaction," Ph.D. dissertation, University of Glasgow, 2012, <https://theses.gla.ac.uk/3260/>, accessed 18 February 2023.
- [548] C. E. Wilson, "Triangulation: the explicit use of multiple methods, measures, and approaches for determining core issues in product development," *Interactions*, vol. 13, no. 6, pp. 46–ff, 2006. [Online]. Available: <https://doi.org/10.1145/1167948.1167980>
- [549] G. Wilson and M. McGill, "Violent video games in virtual reality: Re-evaluating the impact and rating of interactive experiences," in *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, ser. CHI PLAY '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 535–548. [Online]. Available: <https://doi.org/10.1145/3242671.3242684>
- [550] M. Wilson, W. Mackay, E. Chi, M. Bernstein, and J. Nichols, "Replichi sig: From a panel to a new submission venue for replication," in *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '12. New York, NY, USA: ACM, 2012, p. 1185–1188. [Online]. Available: <https://doi.org/10.1145/2212776.2212419>
- [551] M. L. Wilson, E. H. Chi, S. Reeves, and D. Coyle, "Replichi: The workshop ii," in *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '14. New York, NY, USA: ACM, 2014, p. 33–36. [Online]. Available: <https://doi.org/10.1145/2559206.2559233>
- [552] M. L. L. Wilson, P. Resnick, D. Coyle, and E. H. Chi, "Replichi: The workshop," in *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 3159–3162. [Online]. Available: <https://doi.org/10.1145/2468356.2479636>

- [553] C. Winkler, J. Gugenheimer, A. De Luca, G. Haas, P. Speidel, D. Dobbstein, and E. Rukzio, “Glass unlock: Enhancing security of smartphone unlocking through leveraging a private near-eye display,” in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 1407–1410. [Online]. Available: <https://doi.org/10.1145/2702123.2702316>
- [554] B. G. Witmer and M. J. Singer, “Measuring presence in virtual environments: A presence questionnaire,” *Presence*, vol. 7, no. 3, pp. 225–240, 1998. [Online]. Available: <https://doi.org/10.1162/105474698565686>
- [555] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins, “The aligned rank transform for nonparametric factorial analyses using only anova procedures,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2011, p. 143–146. [Online]. Available: <https://doi.org/10.1145/1978942.1978963>
- [556] J. O. Wobbrock and J. A. Kientz, “Research contributions in human-computer interaction,” *Interactions*, vol. 23, no. 3, p. 38–44, Apr. 2016. [Online]. Available: <https://doi.org/10.1145/2907069>
- [557] C. G. Wolf, J. M. Carroll, T. K. Landauer, B. E. John, and J. Whiteside, “The role of laboratory experiments in hci: help, hindrance, or ho-hum?” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1989, pp. 265–268. [Online]. Available: <https://doi.org/10.1145/67450.67500>
- [558] E. Wolf, N. Döllinger, D. Mal, C. Wienrich, M. Botsch, and M. E. Latoschik, “Body weight perception of females using photorealistic avatars in virtual and augmented reality,” in *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2020, pp. 462–473. [Online]. Available: <https://doi.org/10.1109/ISMAR50242.2020.00071>
- [559] J. D. Woodward, “Biometrics: privacy’s foe or privacy’s friend?” *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1480–1492, 1997. [Online]. Available: <https://doi.org/10.1109/5.628723>
- [560] N. Wouters, J. Downs, M. Harrop, T. Cox, E. Oliveira, S. Webber, F. Vetere, and A. Vande Moere, “Uncovering the honeypot effect: How audiences engage with public interactive systems,” in *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, ser. DIS '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 5–16. [Online]. Available: <https://doi.org/10.1145/2901790.2901796>

- [561] B. Xin, G. Chen, Y. Wang, G. Bai, X. Gao, J. Chu, J. Xiao, and T. Liu, “The efficacy of immersive virtual reality surgical simulator training for pedicle screw placement: a randomized double-blind controlled trial,” *World neurosurgery*, 2019. [Online]. Available: <https://doi.org/10.1007/s00264-020-04488-y>
- [562] H. Xu, Y. Zhou, and M. R. Lyu, “Towards continuous and passive authentication via touch biometrics: An experimental study on smartphones,” in *Symposium on usable privacy and security, SOUPS*, vol. 14, 2014, pp. 187–198.
- [563] S. Yi, Z. Qin, N. Carter, and Q. Li, “Wearlock: Unlocking your phone via acoustics using smartwatch,” in *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*, 2017, pp. 469–479. [Online]. Available: <https://doi.org/10.1109/ICDCS.2017.183>
- [564] C. Youngblut, “Experience of presence in virtual environments,” Institute for Defense Analyses Alexandria VA, Tech. Rep., 2003.
- [565] N. H. Zakaria, D. Griffiths, S. Brostoff, and J. Yan, “Shoulder surfing defence for recall-based graphical passwords,” in *Proceedings of the Seventh Symposium on Usable Privacy and Security*, ser. SOUPS '11. New York, NY, USA: Association for Computing Machinery, 2011. [Online]. Available: <https://doi.org/10.1145/2078827.2078835>
- [566] M. R. Zhang and S. Zhai, “Phraseflow: Designs and empirical studies of phrase-level input,” in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI '21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: <https://doi.org/10.1145/3411764.3445166>
- [567] Z. Zhang, M. Sun, B. Gao, and L. Wang, “2-thumbs typing: A novel bimanual text entry method in virtual reality environments,” in *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2021, pp. 530–531. [Online]. Available: <https://doi.org/10.1109/VRW52623.2021.00147>
- [568] M. E. Zurko and R. T. Simon, “User-centered security,” in *Proceedings of the 1996 workshop on New security paradigms*, ser. NSPW '96. New York, NY, USA: ACM, 1996, pp. 27–33. [Online]. Available: <https://doi.org/10.1145/304851.304859>