Saarland University
Department of Computer Science

Averting Security Theater: Methods to Investigate and Integrate Secure Experience in a User-Centered Security Design Process

Dissertation
zur Erlangung des Grades
des Doktors der Ingenieurwissenschaften
der Fakultät für Mathematik und Informatik
der Universität des Saarlandes

von
Dipl.-Ing. Matthias Fassl, BSc

Saarbrücken, 2024

Tag des Kolloquiums:	15. Juli 2024
Dekan:	Prof. Dr. Roland Speicher

Prüfungsausschuss:

Vorsitzender:	Prof. Dr. Jürgen Steimle
Berichterstattende:	Dr. Katharina Krombholz
	Prof. Dr. Antonio Krüger
	Prof. Dr. M. Angela Sasse
Akademischer Mitarbeiter:	Dr. Ashwin Ram

Zusammenfassung

Die Interaktion von Endbenutzer*innen mit Computersicherheitsmechanismen kann dazu führen, dass sie sich geschützt fühlen, was zu einer sogenannten *Secure Experience* (sicheren Nutzungserfahrung) führt. Eine Secure Experience kann jedoch trügerisch sein.

Erstens führt sie dazu, dass Benutzer*innen ihre Vorstellung von Sicherheit darauf aufbauen und diese als Grundlage für die Auswahl ihrer Sicherheitspraktiken verwenden. Bei Fehlen einer Secure Experience können Endbenutzer*innen daher oft keine fundierte Entscheidungen mehr über ihre Sicherheitspraktiken treffen. Unter Umständen würden sie dann unsichere Kommunikationstools wählen, obwohl sie Zugang zu sichereren Alternativen haben. Zweitens sind ungerechtfertigte Secure Experiences – die Endbenutzer*innen das Gefühl geben, geschützt zu sein, obwohl sie es nicht sind – gefährlich. Diese von Bruce Schneier als *Sicherheitstheater* bezeichnete Situation kann dazu führen, dass Benutzer*innen auf tatsächlich wirksame Sicherheitspraktiken verzichten oder ein riskanteres Verhalten an den Tag legen, weil sie glauben, ohnehin ausreichend geschützt wägen.

Um eine Täuschung der Benutzer*innen zu vermeiden, ist das bewusste Design dieser Secure Experiences entscheidend. Etablierte Designprozesse sollten die Secure Experience mit dem eigentlichen technologischen Schutz in Einklang bringen. Derzeit gibt es im Forschungsfeld Usable Privacy and Security keine etablierten Methoden, um die Konzepte der Secure Experience und des Sicherheitstheaters in den Designprozess zu integrieren.

Diese Arbeit befasst sich mit beiden Arten von trügerischen Secure Experience, indem sie bestehende Methoden aus der Mensch-Computer-Interaktion (HCI) auf die Rahmenbedingungen der Security-Forschung anpasst und neue Designmethoden vorschlägt. Der erste Teil dieser Arbeit beschreibt eine gründliche Untersuchung von Sicherheits- und Datenschutzproblemen, die trügerische Secure Experiences zur Folge hat: Cookie-Zustimmungshinweise im Internet, Anti-Stalker-Software auf Smartphones und die kombinierte Nutzung von Private-Browsing-Tools. Im zweiten Teil dieser Arbeit werden bekannte Methoden aus dem Bereich HCI auf die Sicherheitsforschung angewandt und adaptiert. Im dritten Teil wird gezeigt, wie man die Konzepte der Secure Experience und des Sicherheitstheaters in einen nutzerzentrierten Sicherheitsdesignprozess integriert.

Abstract

End users' interaction with computer security mechanisms can make them *feel* protected – resulting in a so-called *secure experience*. However, these secure experiences can be deceiving.

First, users build their mental models of security and choose their security practices based on their prior secure experiences with the available tools. Hence, the lack of secure experience reduces end users' ability to make informed decisions about their security practices. Without this knowledge foundation, users may, for example, choose insecure communication tools despite having access to more secure alternatives. Second, unjustified secure experiences – that make end users feel protected while they are not – are dangerous. This situation, termed *security theater* by Bruce Schneier, may result in users forgoing effective security practices or engaging in riskier behavior because they believe to be sufficiently protected.

To avoid deceiving users, the deliberate design of these secure experiences is crucial to align them with the technology's ability to protect users from security risks. Currently, the field of Usable Privacy and Security has no methods to integrate the concepts of secure experience and security into the design process.

This thesis tackles both types of deceptive secure experiences by adapting existing methods from Human-Computer Interaction to Security and proposing new design methods: Contextual Inquiry at scale, Autoethnography, and a user-centered security design approach. The first part of this work describes a thorough investigation of security and privacy issues that have deceptive secure experiences: cookie consent notices on the web, anti-stalkerware on smartphones, and the combined use of private browsing tools. The second part of this work applies and adapts known methods from HCI to security. In the third part, this work demonstrates how to integrate the concepts of secure experience and security theater into a user-centered security design process.

Background of this Dissertation

This thesis is based on the following six papers. I was the main author of five of these papers [P3, P2, P5, P4, P6] and shared the main authorship with my colleague Lea Gröber for one paper [P1]. One paper [P6] was the result of a significant extension of my master's thesis. My collaboration partners were from CISPA Helmholtz Center for Information Security, Umeå University, and TU Wien. Two of these papers were published at the *CHI Conference on Human Factors in Computing Systems*, ranked A* in CORE2021, one was published at the *ACM Conference on Computer Supported Cooperative Work (CSCW)*, ranked A in CORE2021, and one was published at the premier venue for Usable Privacy and Security research, the *Symposium for Usable Privacy and Security (SOUPS)*, ranked B in CORE2021.

The CSCW paper on *"Investigating Security Folklore: A Case Study on the Tor over VPN Phenomenon"* [P3] received a best paper honorable mention award 🏆, awarded to the top 3% of all submissions, and a methods recognition 🏆, which are awarded to *"strong examples of work that include well developed, explained, or implemented methods, and/or methodological innovation."*

[P1] **Fassl, M.**, Gröber, L. T., and Krombholz, K. Stop the Consent Theater. In: *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, May 2021, 1–7.

[P2] **Fassl, M.**, Anell, S., Houy, S., Lindorfer, M., and Krombholz, K. Comparing User Perceptions of Anti-Stalkerware Apps with the Technical Reality. In: *Proceedings of the Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*. USENIX Association, Boston, MA, USA, 2022, 135–154.

[P3] **Fassl, M.**, Ponticello, A., Dabrowski, A., and Krombholz, K. Investigating Security Folklore: A Case Study on the Tor over VPN Phenomenon. *Proc. ACM Hum.-Comput. Interact.* CSCW2 (Nov. 2023).

[P4] **Fassl, M.**, Neumayr, M., Schedler, O., and Krombholz, K. Transferring Update Behavior from Smartphones to Smart Consumer Devices. In: *Computer Security. ESORICS 2021 International Workshops*. Ed. by Katsikas, S., Lambrinoudakis, C., Cuppens, N., Mylopoulos, J., Kalloniatis, C., Meng, W., Furnell, S., Pallas, F., Pohle, J., Sasse, M. A., Abie, H., Ranise, S., Verderame, L., Cambiaso, E., Maestre Vidal, J., and Sotelo Monge, M. A. Vol. 13106. Springer International Publishing, Cham, 2022, 357–383.

[P5] **Fassl, M.** and Krombholz, K. Why I Can't Authenticate — Understanding the Low Adoption of Authentication Ceremonies with Autoethnography. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI '23. ACM, Hamburg, Germany, 2023, 15.

[P6] **Fassl, M.**, Gröber, L. T., and Krombholz, K. Exploring User-Centered Security Design for Usable Authentication Ceremonies. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI '21. ACM, Yokohama, Japan, May 2021, 1–15.

Further Contributions of the Author

I also contributed as a co-author to several other papers with collaboration partners from CISPA Helmholtz Center for Information Security, University of Luxembourg, Carnegie Mellon University, Eindhoven University of Technology, University of Bonn, and Fraunhofer FKIE. Two of the papers [S3, S5] were main-authored by my colleagues for their master's thesis.

[S1] Distler, V., **Fassl, M.**, Habib, H., Krombholz, K., Lenzini, G., Lallemand, C., Cranor, L. F., and Koenig, V. A Systematic Literature Review of Empirical Methods and Risk Representation in Usable Privacy and Security Research. *ACM Transactions on Computer-Human Interaction* 28, 6 (Dec. 2021), 1–50.

[S2] Distler, V., **Fassl, M.**, Habib, H., Krombholz, K., Lenzini, G., Lallemand, C., Koenig, V., and Cranor, L. F. Empirical research methods in usable privacy and security. In: *Human Factors in Privacy Research*. Ed. by Gerber, N., Stöver, A., and Marky, K. Springer International Publishing, Cham, 2023, 29–53.

[S3] Gröber, L., **Fassl, M.**, Gupta, A., and Krombholz, K. Investigating Car Drivers' Information Demand after Safety and Security Critical Incidents. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI. ACM, Yokohama, Japan, May 2021, 1–17.

[S4] Ortloff, A.-M., **Fassl, M.**, Ponticello, A., Martius, F., Mertens, A., Krombholz, K., and Smith, M. Different Researchers, Different Results? Analyzing the Influence of Researcher Experience and Data Type During Qualitative Analysis of an Interview and Survey Study on Security Advice. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI '23. ACM, Hamburg, Germany, 2023, 21.

[S5] Ponticello, A., **Fassl, M.**, and Krombholz, K. Exploring Authentication for Security-Sensitive Tasks on Smart Home Voice Assistants. In: *Proceedings of the Seventeenth Symposium on Usable Privacy and Security*. SOUPS. USENIX Association, Virtual Conference, 2021, 18.

Acknowledgments

Seven years ago, after my first meeting with Katharina, she cautiously asked me *"Have you ever thought about doing a PhD?"* At that time, I just laughed and admitted that, yes, I have thought about doing one as long as it's not in Austria. It was the start of a relationship that is quite unlike any other. Throughout the years, she has been a valuable friend, mentor, and boss, always available to give kind and emphatic words of encouragement and support when I needed them the most. Katharina, I have seen you go above and beyond your duties as a Professor to help students in need, which only elevated my deep respect for you. Try to keep this character trait alive in the years to come.

My time at CISPA would not have been the same amount of fun without my fantastic PhD colleagues, Lea, Carolyn, Alexander, Divyanshu, Dañiel, Simon, Sebastian, and our group's postdoc, Adrian. The fruitful collaborations, office shenanigans, sports activities, and parties forged our friendship. Over time, you became my second family away from home, and I will miss you dearly.

During my first summer school in Bochum, I also made some academic friends: Miranda, Mohammad, Nikita, Verena, and Viktor. Apart from following their papers, they always gave me something to look forward to, knowing that I would see familiar faces at small and large conferences and have the time to catch up with them.

I also had the good fortune to work with excellent collaboration partners from CISPA, TU Wien, Umeå Universitet, University of Luxembourg, Carnegie Mellon University, University of Bonn, and Fraunhofer FKIE: Simon Anell, Lorrie Cranor, Adrian Dabrowski, Verena Distler, Hana Habib, Sabine Houy, Lea Gröber, Abhilash Gupta, Vincent Koenig, Katharina Krombholz, Carine Lallemand, Gabriele Lenzini, Martina Lindorfer, Florin Martius, Anne Mertens, Michaela Neumayr, Anna-Marie Ortloff, Alexander Ponticello, Oliver Schedler, and Matthew Smith. It was a pleasure to work with you and follow your career success. I look forward to seeing what you will do in the future.

State-of-the-art research depends on a lot of other people's work as well. At CISPA, we can always rely on the fantastic work of the corporate communications department, all the help we get from the front office and the indispensable work of the facility managers and cleaning staff. I also want to give all the members of the workscouncil a shoutout: keep on fighting.

Even abroad, I could count on my friends' and family's support in Vienna. Thank you, Stefan, Georg, Viktor, Mathias, and Michael, for always finding time to meet up and rekindle old friendships. Some of my friends from Vienna also started their own PhD journey. Thank you, Gabriel, Sabrina, and Hans-Jörg, for having an open ear for the shared PhD struggles at different institutions. Thank you, moms and dads, for supporting me and keeping an interest in my PhD journey.

Over the years, I received many reviews for my paper submissions. I thank all the anonymous reviewers who invested time in reading my papers and providing helpful and constructive feedback. Similarly, I thank all the reviewers of this thesis and the committee members for being there at the final stage of this journey: Katharina Krombholz, Antonio Krüger, M. Angela Sasse, Jürgen Steimle, and Ashwin Ram.

Now there is not much more left for me to say, except maybe: *"It's been a fun ride, but it's time to say goodbye"* or as we would rather say in Vienna *"Es war mir ein Fest, auf wiederschaun und baba."*

Contents

List of Figures

List of Tables

1 Introduction

In interaction design, the term *user experience* describes how users feel when interacting with software. Analogously, a *secure experience* describes how protected users feel when using security technology. These secure experiences affect users' beliefs about how technology works, their intention to use security technology, and, consequently, security practices. In doing so, they may deceive users in multiple ways, leading to faulty conclusions and expectations about security.

Automating security and making it invisible to users has been and continues to be a popular approach to improve the usability of security [68]. While this approach works, it removes users' secure experience. Without a secure experience, end users cannot distinguish between secure and insecure practices from the user interface alone. Hence, a lack of secure experiences erodes end users' knowledge foundation for making informed decisions about their security practices [70]. This effect is observable in user studies: For example, when tasked with choosing a secure communication medium, users can not tell if sending an email, SMS, or WhatsApp message will give them the highest level of security [3, 54]. They think SMS gives them more security than WhatsApp because their bank uses that method to communicate with them. In contrast, the security guarantees of the much more secure WhatsApp are invisible to them. A similar confusion exists around anonymous browsing tools, specifically the Tor browser and commercial VPN providers. Users who experienced both have difficulties explaining the difference it makes to their online anonymity [264], even though the Tor browser offers much better anonymity guarantees than VPN providers.

While a lack of secure experience can result in flawed security decisions and practices, the reverse – an excellent but unjustified secure experience – is more dangerous. One of the dangers is that users, who feel adequately protected, ignore more helpful security advice and practices due to a limited compliance budget [23]. Another danger is that users might take on more risks to compensate for the perceived security improvement [243]. Bruce Schneier coined the term *security theater* [252] for such performative security practices that do little to mitigate risks. Unjustified secure experience is, for example, observable in the user interaction of anti-virus and anti-stalkerware software. However, it is also a common complaint amongst cybersecurity industry experts [71].

This thesis tackles both types of user deception stemming from misaligned secure experience: the lack of secure experience that reduces end users' decision-making abilities and the unjustified secure experience that lulls end users into a false sense

of security. Incorporating the concept of secure experience into the security design process is necessary to begin tackling the potential of user deception. However, the Usable Privacy and Security field – a comparatively recent field of research that is still developing – currently has no methods of investigating and designing user experience for security. Hence, developing and tailoring these methods for security and privacy is the first step to the in-depth exploration of secure experiences. Since user experience has been a research topic in the closely related field of Human-Computer Interaction (HCI), this thesis explores examples of how existing methods can be transferred to Usable Privacy and Security.

Research Questions. This thesis proposes a methodological framework for designing secure user experiences that match technical security guarantees and avoid security theater.

Designing secure experiences depends on successfully tackling two problems: First, identifying instances of security theater and secure experience mismatches, in general, is necessary, preferably systematically. Without identifying secure experience mismatches, it is hardly possible to decide if the secure experience design process was successful. Second, the Usable Privacy and Security field needs additional methods to investigate and design user experience. This thesis explores the possibility of adapting appropriate methods from HCI, where user experience is an already established topic of research, Then, this thesis uses real-world examples to demonstrate integrating the secure experience and security theater concepts into a user-centered design process for security mechanisms. This thesis investigates these *secure experience* design issues with three research questions:

RQ1: How can we systematically detect instances of security theater?

RQ2: How can we adapt established HCI methods to security mechanisms?

RQ3: How can we effectively match secure experiences with actual security?

Part I to III of this thesis each tackle the corresponding research question. Each of the chapters, which are based on published papers, illustrates a methodological approach using concrete security research problems.

1.1. Contributions of this thesis

The contribution of this thesis is primarily methods-driven, i.e., it proposes methods that are necessary and useful to handle the evolving types of research questions in the Usable Security and Privacy field. Each chapter illustrates a developed method using an example study. Table 1.1 offers an overview of the papers' contribution to answering the three research questions.

Recognizing Instances of Security Theater

Security theater in ICT is a mismatch of user experience with technical abilities. Chapter 3 to 5 describe three different methods of systematically recognizing security

Table 1.1.: Contribution of this thesis' papers

RQs	Paper	Contribution
Recognizing Security Theater	[P1]	Literature review to identify and understand security theater
	[P2]	User perception and technical analysis of a specific security tool
	[P3]	Case study of security tool combination
Adapting HCI Methods to Security	[P4]	Contextual inquiry at scale for understanding and improving update behavior for smart consumer devices
	[P5]	Autoethnography to understand secure experience of cooperative security mechanisms
Designing Secure Experiences	[P6]	Integrating secure experience into a user-centered security design approach including evaluation

theater: for well-studied areas of research, for checking standalone non-sensitive apps, and for instances where users combine different security tools expecting increased security.

A systematic literature review helps recognize and understand security theater for well-studied research areas. In Chapter 3, we used this approach to understand the consent theater on the web. Comparing the detailed prior work on user understanding of online behavioral advertising with literature on the technical side of online advertising and the existing literature on dark patterns in cookie consent banners showed that most users are not fully aware of what they are consenting to.

In Chapter 4, we present findings on the security theater of specific security-focused Android apps. We exemplify this approach using anti-stalkerware apps, which are supposed to detect and mitigate the dangers of stalkerware on users' phones. We use a qualitative analysis approach and UI walkthroughs to understand the secure experience of using these apps and contrast them with the static and dynamic program analysis results. The results demonstrate mismatching secure experiences and technical abilities of the apps. However, users over- and underestimation of the apps' security depends on their positive and negative experiences with them, respectively.

In Chapter 5, we investigate how combining different security tools may result in a security theater. We analyzed related work on experts' opinions on the effectiveness of the combination, quantified the phenomenon in a privacy-preserving way, conducted an online survey to understand participants' security beliefs about the combination, and systematically analyzed the background information sources to understand the origins of these beliefs. The results indicate that many Tor over VPN users expect general security benefits from the combination without a specific threat model. This expectation appears to spread through normative beliefs, where users copy security practices from others for primarily social reasons.

The three approaches demonstrate how a combination of careful literature review,

behavior measurement, online user surveys, app-store reviews, UI walkthroughs, and qualitative analysis of background information sources help systematically recognize and understand the origins of security theater. While the three approaches already cover a variety of use cases, researchers may also recombine these approaches to fit other types of cases to investigate.

Adapting Established HCI Methods for Security and Privacy

Often, Usable Security and Privacy lack the necessary tools for tackling the evolving types of problems that belong to the human factors area in Security and Privacy. It is not always easy to use existing methods from HCI because, usually, HCI deals with primary interaction goals – and security and privacy are secondary interaction goals [295]. Hence, adapting them to these secondary interaction goals is necessary. Chapter 6 and 7 describe example studies that adapted existing methods to Usable Security research questions.

The first study confronts online participants ($N = 91$) with five different contexts for software updates of smartphones and smart consumer devices, operating system updates, app updates, a smart dishwasher, a car's software system, and self-lacing smart shoes. For each context, the online survey inquires about participants' understanding of the situation and preferred design changes to the user interaction. Since software updates are not a daily occurrence, confronting participants directly with the situations allows for collecting participant responses at scale. The resulting implications for design demonstrate ways to communicate the importance of software updates to users while minimizing the distraction from the primary interaction goal.

The second study uses Autoethnography to highlight the social and cultural aspects of secure messengers' authentication ceremonies. Since this security mechanism is performed rarely, an extended study period and a focused approach to authenticating conversations help study the social and cultural context. The analysis suggests a significant sociotechnical gap [4] for authentication ceremonies, where messenger users must invest additional work to organize and plan meetings and integrate these ceremonies into their social life.

Designing Secure Experiences

Lastly, we need a method to design secure experiences that avoid security theater. Chapter 8 demonstrates how a user-centered security design process helped redesign the authentication ceremony of end-to-end-encrypted secure messengers. The study's design pipeline comprises collaborative design workshops with prospective users, an evaluation of the resulting concepts by a security expert, iterative storyboard prototyping to develop and fine-tune the prototypes, and an online evaluation. During the design stages, we apply Hassenzahl's hedonic/pragmatic model of user experience [161] to integrate the secure experience into the design process. During the iterative storyboard prototyping and the final evaluation, we ask participants quantitative and qualitative questions about their security perceptions (hedonic) and threat and security understanding (pragmatic). Ultimately, we evaluated three prototypes based on QR-code scanning, selfies, and a combination lock. Scanning QR codes increased the sense of a hedonic

secure experience without questioning the pragmatic side of security. Exchanging selfies between conversation partners also affected the hedonic experience but frequently came with doubts about the pragmatic effect on security. The combination lock was the only prototype that led to an increased pragmatic understanding of the threat models behind authentication ceremonies, and this increased understanding also affected the hedonic secure experience. This study demonstrates how integrating secure-experience-focused feedback questions into the security design process is useful for designing secure experiences that avoid becoming a security theater. Researchers can also apply the resulting user-centered security design process to other security issues after minor tailoring to each use case.

1.2. Thesis Structure

The findings of this thesis were published in peer-reviewed venues, ACM CHI, ACM CSCW, USENIX SOUPS, alt.chi at CHI, SPOSE Workshop at ESORICS. In this thesis, each of the papers has its chapter. The thesis structures the paper-based chapters according to the research questions they tackle.

Chapter 1: The introductory chapter motivates this work, introduces research questions it sought to answer, describes this thesis' contribution to existing knowledge, and explains its structure.

Chapter 2: This chapter explains the related work about secure experiences, security theater, security advice, and design methods.

Part I: Recognizing Instances of Security Theater

Chapter 3: This chapter provides an overview of behavioral-tracking-based surveillance capitalism on the web and distills three key issues from the corpus on related work on online behavioral advertisement and users' acceptance of it. A core message is that relying on consent for online behavioral advertising is infeasible because users are often unaware of what they are consenting to – which makes them unable to give informed consent. The chapter suggests different ways of improving privacy on the web without relying on individuals' consent and proposes future research directions to counter surveillance capitalism. The contents of this chapter have been published as part of the paper "Stop the Consent Theater" Matthias Fassl, Lea Gröber, and Katharina Krombholz. *In Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* [P1].

Chapter 4: This chapter describes a study that collects users' security perceptions of anti-stalkerware apps from app store reviews to contrast them with the results of a user interface walkthrough and a static and dynamic analysis of the anti-stalkerware apps' technical functionality. The chapter recommends different approaches to provide decision support for users trying to protect themselves from stalkerware and looking for adequate detection apps. The contents of this chapter were published as part of

the paper "Comparing User Perceptions of Anti-Stalkerware Apps with the Technical Reality" Matthias Fassl, Simon Anell, Sabine Houy, Martina Lindorfer, and Katharina Krombholz. *In Proceedings of the Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)* [P2].

Chapter 5: This chapter describes three studies that tried to understand the Tor over VPN phenomenon, i.e., that some users connect to a VPN before opening the Tor Browser even though the security benefits of the practice are unclear. Around 6% of Tor users connect via a VPN, of which most expect general security benefits from the combination, and a study of the online information sources on the topic indicate that normative beliefs, based on social proof, contribute to the spread of this type of combination practice. The contents of this chapter were published as part of the paper "Investigating Security Folklore: A Case Study on the Tor over VPN Phenomenon" Matthias Fassl, Alexander Ponticello, Adrian Dabrowski, and Katharina Krombholz. *In Proceedings ACM Human-Computer Interaction CSCW2 (Nov. 2023)* [P3].

Part II: Adapting Established HCI Methods for Security and Privacy

Chapter 6: This chapter describes two studies that aimed to understand how users transfer their previous experience with software updates to smart consumer devices and provide design recommendations for everyday consumer device software updates. The contents of this chapter were published as part of the paper "Transferring Update Behavior from Smartphones to Smart Consumer Devices" Matthias Fassl, Michaela Neumayr, Oliver Schedler, and Katharina Krombholz. *In Computer Security ESORICS 2021 International Workshops* [P4].

Chapter 7: This chapter describes an autoethnography study to understand why security-conscious users rarely authenticate their conversations on end-to-end-encrypted messengers. One of the reasons is that authentication ceremonies are a cooperative security mechanism, i.e., where two users need to work together and organize to increase their security. The chapter suggests a high-level method for designing cooperative security and privacy mechanisms based on transcultural encoding. The contents of this chapter were published as part of the paper "Why I Can't Authenticate—Understanding the Low Adoption of Authentication Ceremonies with Autoethnography" Matthias Fassl and Katharina Krombholz. *In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* [P5].

Part III: Matching Secure Experiences with Actual Security

Chapter 8: This chapter describes a series of studies that aimed to redesign authentication ceremonies for end-to-end-encrypted messengers. The research project comprised collaborative design workshops, security expert reviews of the participants' concepts, iterative storyboard prototyping, and a qualitative and quantitative online evaluation of the prototypes' resulting usability and user experience. The contents of this chapter were published as part of the paper "Exploring User-Centered Security Design for Usable

Authentication Ceremonies" Matthias Fassl, Lea Gröber, and Katharina Krombholz. *In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* [P6].

Part IV: Discussion and Conclusion

Chapter 9: This chapter summarizes the results of the papers described in Chapters 3 to 8. Based on the results, it outlines a systematic approach to discovering instances of security theater. It discusses how researchers may adapt existing design methods to security and privacy, the necessary conditions for matching secure experiences with actual security, and how security- and privacy-focused companies may exploit these mismatches.

Chapter 10: This chapter summarizes the outcomes of this thesis and suggests future work on the design of secure experiences.

Each of the papers this thesis is based on was a collaboration between two or more authors – which is also why most of the content of these chapters will use the pronoun "we". I am very grateful for my co-authors' support; their help made these research papers possible. However, to make my contribution abundantly clear, I added a disclaimer before the beginning of Chapters 3 to 8 that summarizes my co-authors' and my contributions to the work.

2 Background and Related Work

Since the Chapters in this thesis focus on different security and privacy topics, they contain specialized Sections on relevant related work. This Chapter covers the background and related work of the overarching topic of this thesis to avoid reiterating them in every chapter.

As such, the contents of this Chapter are distilled from published prior work [P1, P2, P3, P4, P5, P6]. It starts with the methodological development in the Usable Privacy and Security field and continues with related work on security advice and adoption, secure experience, security theater, and interaction design.

2.1. Methodological Development of UPS Research and Design

Each new research topic challenges earlier views, methodology, and methods. As Feyerabend highlighted in his book *Against Method* [81], accepting new methods in a field brings the opportunity to overcome long-standing research problems and may lead to an influx of new insights. Over the years the field that we now call Usable Privacy and Security has evolved drastically and it still continues to do so. Each development broadens the research challenges that aim to improve users' and organization's security, privacy, and safety.

HCI research has experienced two significant turns over the years that mirrored the *"changing relationship between people and the machines with which they work."* [22] These are generally thought of as three distinct waves of research approaches:

1st wave: Stemming from a human factors (North America) and software ergonomics (Europe) background, first-wave HCI sees humans as another system component with specific faults, such as limited attention span, impaired memory, or lack of knowledge that need to be factored into the system design. In general, this research approach focuses on the individual behavior of novice users in small controlled lab experiments, evaluating the overall performance of a system that includes defined human factors. [22].

2nd wave: Once computers became more widespread in the workplace, more people became discretionary users: They worked in a profession unrelated to computing but used a computer to make their job easier. From this time on, users

were seen as an active participant. They adapted their systems in unforeseen ways to make their jobs easier, becoming user-designers. Since research focused on the workplace, collaborative group work became more important than studying individual computer use. Consequently, research moved away from tightly controlled lab studies to in-situ studies on how computing systems integrated into the daily work and supported it. Acknowledging that workers had domain-expertise, studies also started considering domain experts in addition to novices. Importantly, second-wave HCI researchers began considering different design approaches to build systems that adhered to user-designers values. [22]

3rd wave: Computing devices became ubiquitous, personal computers became a common sight in homes and smartphones, tablets, and wearables started accompanying us everywhere we go. This changed (multi-)device use impacts peoples' everyday lives, social interactions, and culture – not just how work gets done. Consequently, HCI research turned towards technology use that transcends the workplace: studying user experience, emotions, technology-mediated social interactions, and culture – often with field methods that became necessary to collect ecologically valid data. [29]

The field of Usable Privacy and Security, which adopts many HCI research approaches, evolved analogously to these three waves. One of the first works that mention human factors in connection with a security mechanism were Saltzer and Schroeder [248] in 1975. Then, the focus was on the usability of authentication mechanisms. The focus on usability was underlined by Whitten and Tygar's [295] well-known usability analysis of PGP-encrypted email. In the nineties, researchers highlighted the importance of user-centered design for security [310] especially in the workplace context [7] – heralding 2nd wave HCI research in security. Around the same time, policy research in the context of usable privacy became more popular (e.g., Ackerman et al. [5]). More recently, system administrators and software developers have moved more into focus of research efforts because they have a significant role in relieving end users' security burden [117, 158].

The latest development in the field is the turn towards digital harms and safety. This turn takes user-centered security seriously and studies people's lived experience with ubiquitous interconnected technology. This change in focus on safety likely constitutes a shift in perspective towards third-wave usable security research. Strohmayer et al. [265] describe safety as the security communities grand challenge. Freed et al.'s seminal paper [90] – which received a best paper award at CHI'18 – on stalkers' use of smartphone apps in the context of intimate partner violence helped kick-start this development. Interestingly, security, cryptography, and privacy mechanisms might be tools to improve safety, but they are not necessarily the best way or the only way of achieving safety. Designing technology with safety in mind also requires considering potential abusers throughout the design process and conducting abuse testing [216]. As soon as we take the perspective of improving users safety, we must consider that one group's safety can have detrimental effects on another group's safety. Thus, these issues are not separable from existing power dynamics in society [177]. These kinds of policy issues must be considered during research, design, and evaluation – a challenging

endeavour.

Research approaches from all different waves co-exist in Usable Privacy and Security. It lies in the nature of an interdisciplinary research field that researchers with different backgrounds and experiences have different perspectives on the necessary research to improve security, privacy and safety. However, despite the advances in methodological approaches, a first-wave human-factors perspective on users – who have to be prodded and nudged to produce the "correct" behavior – is still prevalent in security research.

This thesis focuses on the kind of secure experiences people have when using security mechanisms. It adapts and develops necessary methods to evaluate and design them – to avoid misleading users about the effects and benefits they have. As such, this thesis work contributes to shifting the research perspectives towards third-wave usable security research.

2.2. Security Advice

Due to a lack of structured security education, users learn their security behaviors haphazardly from various sources. Media, negative experiences, family, peers, workplace, IT professionals, and service providers are common advice sources [239]. However, all these sources focus on different aspects of threats [226]. Hence, no single source is sufficient.

There is a large body of security advice from which users can choose. In a survey, security experts suggested a total of 152 different kinds of security advice for non-tech-savvy users [241]. When Redmiles et al. [240] created their corpus of security advice, they found 374 unique advice imperatives online. However, since people can only adhere to a limited amount of security advice (also called *compliance budget* [23]), a small set of high-quality security advice is needed that benefits a broad range of people. High-quality general advice should be effective, actionable, consistent, and concise [241]. Redmiles et al. [240] analyzed a corpus of collected security advice along three dimensions: comprehensibility, perceived efficacy, and perceived actionability. While many advice sources were hard to understand for the general public, the authors identified a crisis of advice prioritization: the professional security experts identified a total of 118 pieces of advice as their top five items they would recommend to users.

Social Support in Security Decisions

In the face of the overwhelming large body of available security advice [241, 240] and limited compliance budget [23], it is becoming increasingly difficult for users to pick security practices that fit their needs. Often, people resort to social support to inform their security choices.

For example, people learn informal security lessons from family members' and friends' stories about prior security incidents [227]. However, the contents, the location, and the storyteller influence how much effect these informal lessons have on the listeners' day-to-day security practices.

Online reviews are another common way to seek support in security decisions. For example, some users check app-store reviews for other people's experience with an

update before deciding to update themselves [269]. The quality of the review contents and the ranking affects consumer decisions more than the number of reviews and the sources' credibility [84].

Situations of abuse that involve intimate partner surveillance (IPS) are an especially sensitive area for security advice. Security advice that abruptly prevents attackers' access to data may worsen abusive situations [164]. Hence, anti-stalkerware apps that are not specifically adapted for use in abusive situations may only be safe to use in the life apart phase. Emms et al. [73] suggested approaches to improve survivors' ability to avoid traces in ongoing abuse situations. Since it might be hard to reach out to immediate peers, IPS victims and survivors seek help and support in online forums from other survivors [163]. However, forum users often lack appropriate technical know-how, making it hard to recommend safe and effective security practices or anti-stalkerware apps.

Security Advice Adoption and Abandonment

Understanding the circumstances that lead to user adoption of security and privacy measures is a matter of ongoing research. For the example of privacy-enhancing technology, Harborth et al. [127] found that perceived anonymity and trust influence perceived usefulness and, consequently, drive adoption. Security practices requiring reoccurring user interaction are less widely adopted than those requiring fewer interactions [308]. The last factor might benefit the adoption of more common types of privacy-enhancing technologies, such as VPN since they do not require a lot of user interaction compared to other security mechanisms.

Zou et al. [308] studied users' reasons for adopting and abandoning security and privacy behaviors. They found low adoption for recurring interaction practices and higher privacy practice adoption rates among low-income participants. Users abandoned security and privacy practices when they found them impractical, no longer saw their value, or perceived diminished risk. Similarly, users turn off protective measures such as firewalls when they find them complicated [229]. Personal negative experiences also influence future security decisions. Bad experiences with software updates discourage users from future updates, even if they would benefit security [281].

A general framework for understanding users' adoption and abandonment could be based on the theory of reasoned action [85]. It provides an overview of how background factors influence different beliefs about practices, how these beliefs affect user intention, and under which circumstances intention translates to user behavior. I applied the theory of reasoned action in Chapter 5 to understand the Tor over VPN practice in detail.

Social Influence on Security Practices

Social factors affect how people communicate about and adopt security practices. As Das et al. [50] discovered, people communicate about security processes to either warn each other about potential threats they observed or experienced or to gather information about security and privacy problems. A key factor to instigating conversations about security and privacy practices seems to be the *observability* of that practice. In a

quantitative study with 1.5 million Facebook users, Das et al. found that having fewer feature-adopting friends reduces the likelihood of adopting that feature. In contrast, having many feature-adopting friends increases the likelihood [52]. They also found that security features that are more observable by friends spread through the social network more effectively. This is an example of a *social proof* [44], where people unsure about what to do in a given situation look to their peers and copy their actions. This is a form of *social descriptive norm* that involves the perceptions of which behaviors are typically performed. In contrast, *social injunctive norms* are about the perception of which behaviors others approve or disapprove – helping to understand which behaviors are socially acceptable.

Not all security features are equally observable by others. However, when communicating, e.g., via mobile messenger, emails, or SMS, the adoption of secure messengers or the use of security features necessarily becomes observable. This starts with the (group) decision on the mode of communication. Luca et al. [53] identified peer pressure from friends as the main factor for secure instant messenger adoption. Abu-Salma et al. [3] found that friends' opinions about the security of a messenger are essential for adoption. In cases where the security feature (such as encryption) has to be used on a case-by-case basis, there are still social processes involved in deciding when to actually use that feature. For example, Gaw et al. [100] found that even employees of a security-concerned activist group considered the social implications before sending encrypted emails. Having the recipients invest extra effort to decrypt "unimportant" mail was considered rude.

Ignoring social aspects of the adoption and use of security features may dampen user adoption. Hence, they should be considered during the design and roll-out of security-relevant tools. Social aspects have been ignored before during the design phase of security-relevant features. For example, the security-relevant process of pairing different devices. These devices can belong to the same person or different people – a critical difference with implications on the design process. Uzun et al. [279] criticized in 2011 that (1) existing pairing methods have been devised by security professionals with little regard for their usability, and (2) it is not possible to reduce the problem of social pairing to personal pairing of devices. They argue that social pairing introduces an additional layer of necessary interaction and adds social context, such as potential embarrassment or discomfort to the problem.

2.3. Secure Experience

When we use technology in our daily life, we *experience* it [186]. According to McCarthy and Wright, this experience encompasses the sights, sounds, feelings, thoughts, and resulting actions of using technology in a specific life situation. This is a highly subjective process, i.e., using the designed technology at the moment transforms the technology into an experience [131]. Because experience is a subjective process, it can not be shaped entirely by the user experience designer; it is co-authored by the users. Hassenzahl et al. [130] used a pragmatic/hedonic model to understand and evaluate user experience. It comprises a pragmatic dimension, with task-related goals, such as usability and utility, and a hedonic dimension that concerns itself with the stimulating, beautiful, and

identity-creating features of an experience.

Experience with technology may also affect the users' security and privacy perceptions and, consequently, decisions. In 2006, Grossman [119] introduced the notion of secure user experiences. He recognized that in many instances, the functional aspects of security – like company firewalls, intrusion detection systems, and password expiration policies – are out of users' control. However, it is that sense of control that makes users feel secure. He argues that creating secure experiences is about respecting users, communicating security considerations effectively, and reducing anxiety around them. He advocated that the design of security features should consider users' emotional interpretation in addition to the functional tasks. Mathiasen and Bødker [179] criticized usable security research that focused on adjusting user behavior instead of taking secure experience into account, leading to entirely different analytical concerns. They observed that merely behaving securely does not necessarily result in a secure user experience – which could lead to adverse outcomes for users.

For example, modern secure messengers, such as WhatsApp and iMessage, made end-to-end encryption usable for the masses by hiding all the details from users and enabling it by default. However, this did not lead to a higher perception or valuation of them as secure technology [55]. Since the difference between unencrypted SMS and WhatsApp communication was not discernable to users, they did not perceive WhatsApp to be more secure. On the contrary, participants think that SMS is more secure, most likely because banks (a trusted entity) use them to communicate with their customers [3].

Side-effects of the Humans-out-of-the-loop Approach to Security

A common approach to tackling usability issues with security mechanisms is removing users from decisions whenever possible, i.e., keeping humans out of the loop [49]. Removing users' ability to make risky decisions can help protect them. Forget et al. [86] found that users who overestimate their security expertise make seemingly rational decisions that lead to ineffective security. However, while reducing the decisions that users face avoids overloading them and potentially risky decisions, it also removes at least some of their agency. While the human-out-of-the-loop approach is at least partially responsible for improved usability in many areas, it also led to widespread paternalistic automation of security mechanisms. As a side-effect, Ruoti et al. [245] showed that hiding security mechanisms can decrease users' trust in them.

Another side-effect of the humans-out-of-the-loop approach is that it likely impacts users' mental models of technology. Users use clues from user interaction to build mental models of technology to make sense of them [22] – automation and hiding security mechanisms make this more challenging. Redmiles et al. [237] found that software prompts are a form of advice source, which means that automating security decisions may reduce users' knowledge base. Users' mental models of potential attackers impact their adopted protection behavior [289] since each class of attacks calls for different protection mechanisms. Ultimately, removing users from security choices creates a problem when automation fails: users are ill-equipped to understand and cope with security decisions [70]. Wash et al. [290] found that users misunderstand their software

updating behavior. Future update decisions could be based on wrong assumptions, which improved education cannot fix. They argue that removing users from security decisions makes it difficult for them to make the remaining decisions intelligently.

In contrast, informative and reassuring secure experiences can positively affect users' security decisions. For example, users are comfortable enabling auto-updating for their apps if they consider them essential and trustworthy or if they are satisfied with them [181]. In contrast, negative experience with software updates reduces users' comfort with auto-updating. Distler et al. [63] found that including security-related information in an e-voting process improved users' secure experience. They discuss how quick and smooth security mechanisms may impede users' secure experience despite improved usability—an idea they extend on in a framework of security-enhancing friction [62].

Security Misconceptions

Users create mental models based on the user interaction [22]. Misleading mental models about security mechanisms can contribute to security misconceptions. This is documented in related work about, for example, private browsing [124, 98, 2], the security of electronic communication [3], secure password creation methods [278], or the anonymity of blockchain transactions [173]. Even when these security misconceptions have no direct negative influence on users' security, they can have side effects such as a security theater [253], i.e., when users feel more secure while the technology is not. This perception may lead to risk compensation behavior [243], i.e., users accepting more risks, believing they are secure. Users have a limited compliance budget [23], i.e., they can only consistently apply a few security practices. So, switching to higher-quality security practices is advisable.

2.4. Security Theater

In *Beyond Fear* [252], Schneier used the term *Security Theater* to describe procedures that provide people the feeling of security without mitigating actual risks. As an example, he named airport security: In the 1970s, armed guards were stationed in the boarding areas. While they did not decrease plane hijacking – a rare event – it did help calm the passengers. More closely related to IT security, he also mentioned modern encrypted cellphone networks. While voice data encryption on cellular networks is sometimes marketed as a privacy feature, its primary purpose is protecting network operators from fraud. Hence, Schneier concludes that this kind of security theater improves companies' market share while being cheaper than actual security measures.

Similarly, security theater can also be applied to IT security issues. It could be used anytime when a software or a user interaction's primary purpose is to make users feel reassured and in control – while not effectively mitigating threats. While the deceptive nature of security theater is an ethical issue, security theater also negatively impacts security behavior. First, it affects users' choice of secure practices. Users can only consistently apply a few security practices (compliance budget) [23]. So, they will inevitably need to prioritize one security practice over another. Since a security theater will skew users' perceptions of which practices provide security, they end up

with practices that feel secure even if they are not. Even though they could benefit more from other security practices. For example, users think SMS messages are more secure than WhatsApp [3]. This means users might choose an insecure communication medium for sensitive discussions, even though everyone involved might have access to an end-to-end-encrypted messenger. Second, wrongly believing in a security theater could increase users' risk-taking behavior. This is called a risk compensation effect [243]. The idea is that when users feel better protected from risk, they also feel more comfortable taking additional risks – because they are protected. While the risk compensation effect is already troubling in regular cases, security theater makes this effect even more dangerous.

However, considering the feeling of security is not always bad. In *Psychology of Security* [253], Schneier describes many psychological reasons for users' bad security trade-offs. Ultimately, he argues that users make the best security trade-offs when the feeling of security matches the reality of security. This means that the feeling of security needs to be considered during the design process of security mechanisms. This thesis lays the groundwork for this integration into the design process to reduce unintentional types of security theater.

2.5. Interaction Design

Dan Saffer [247] divides interaction design into four approaches: (1) *User-Centered Design*: designers translate users' needs and goals, (2) *Activity-Centered Design*: designers create tools for specific actions, (3) *Systems Design*: designers focus on components of a system, and (4) *Genius Design*: designers' skill and wisdom used to make a product.

A systems design perspective is a common approach to designing security mechanisms. This is why Zurko et al. [310] already advocated for a user-centered security design that focused more on users' needs in the 1990s. In a similar vein, Dodier-Lazaro et al. [64] criticized that many security "improvements" stem from a paternalistic mindset that tries to nudge people into "correct" behaviors while ignoring users' values. They advocated for a value-sensitive design approach to security and privacy. Slowly, a co-design approach is becoming more popular in security and privacy. In a co-design approach, potential users are more active in the design process. Ideally, security and user experience experts would help potential users to create security and privacy mechanisms that work for them. Weber et al. [291] used participatory design to draft SSL warning messages. Gorski et al. [115] used a similar approach to create warning messages for developers that use cryptographic libraries. Mathiasen et al. [180, 179] proposed an experience-centered design approach to enable secure experiences. They developed Prompted exploration workshops and Acting out security to target secure experiences in an iterative, participatory design process.

The design of authentication ceremonies in end-to-end-encrypted messaging is likely a mixture of systems design and genius design. The suggestions for improvements to these ceremonies above [283, 285, 298, 267] mostly employ activity-centered design. They all closely examine the authentication process and try to remove barriers, which is essential and effective work. However, this iterative, activity-centered approach overlooks more fundamental questions about users' needs and goals.

Part I.

Recognizing Instances of Security Theater

The contents of the following chapter were published as part of the publication *"Stop the Consent Theater"* (alt.CHI 2021) [P1]. This paper was produced in cooperation with my co-authors Lea Gröber and Katharina Krombholz. The following table describes the contributions of all authors to this paper:

Author	**Contribution**
Matthias Fassl	Apart from the initial idea for the topic of the paper, which was mine, Lea and I shared all work for this paper equally. We discussed the approach to use, collected and analyzed literature, discussed the conclusions from the related work, and ended up writing the paper together.
Lea Gröber	Lea and I shared all work for this paper equally.
Katharina Krombholz	As our academic advisor, Katharina was involved in the major decisions during this project. She gave feedback on the initial idea and advised us to use the approach we ended up taking. She reviewed the final paper before submission and discussed the conclusions with us.

Reference

Fassl, M., Gröber, L. T., & Krombholz, K. (2021). Stop the Consent Theater. Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems, 1–7. `https://doi.org/10.1145/3411763.3451230`

3 Stop the Consent Theater

Figure 3.1.: An artistic interpretation of an honest cookie consent notice.

The current web pesters visitors with consent notices that claim to "value" their privacy, thereby habituating them to accept all data practices. Users' lacking comprehension of these practices voids any claim of informed consent. Market forces specifically designed these consent notices in their favor to increase users' consent rates. Some sites even ignore users' decisions entirely, which results in a mere theatrical performance of consent procedures designed to appear as if it fulfills legal requirements.

Improving users' online privacy cannot rely on individuals' consent alone. We have to look for complementary approaches as well. Current online data practices are driven by powerful market forces whose interests oppose users' privacy expectations – making turnkey solutions difficult. Nevertheless, we provide a bird's-eye view on privacy-improving approaches beyond individuals' consent.

3.1. Those damn cookie banners!!

Cookie banners (i.e., consent notices) that randomly pop up to obstruct your view of something interesting on the web are, without question, annoying. You are probably just as guilty as we have been of clicking it away without reading what it said, thereby submitting to questionable privacy practices. The sheer number of consent notices users encounter while browsing the web can be aggravating. You might ask why this is even necessary. Consent notices are the naïve answer to the complex problem of informational self-determination on the web, where sites store so-called cookies in your browser all the time. Cookies' intended use was mostly centered around e-commerce features such as

virtual shopping carts and remembering logged in users. However, cookies turned out to be a useful tool to mark visitors of one site and track their behavior across other sites.

The intention behind consent notices is to grant users control over their data and thus enhance their privacy. This overall goal is becoming increasingly pressing with the rise of large-scale data gathering and accompanying surveillance. How users could effectively control if and how others profit from their personal data remains an open question. Admittedly, consent notices are doing a poor job at achieving this goal. They are supposably designed to obtain users' informed consent, yet nudge users to allow tracking behavior [6]. These nudges include default settings that allow all kinds of cookies or the deceiving visual presentation and phrasing of consent notices. In many cases, they do not comply with relevant privacy laws, such as GDPR or the ePrivacy directive [184, 249, 274, 191]. And to top it off, they often do not respect users' decisions, even if they opt-out [184]. The current situation supposedly exists for the visitors' privacy benefit but not all involved parties want to achieve that goal, resulting in a consent theater (similar to Schneier's *security theater* concept [252]). The ad-industry directs this theater, the content providers and their visitors are the performers on stage, and the regulators and policymakers are the audiences. This play aims to convince the audiences of the ongoing practices' legality while minimizing harm to the director's interest.

This paper offers a bird's-eye view on the issue of consent-notices, the accompanying data practices, and its different stakeholders. We reviewed and analyzed the relevant body of research in this area, touching on topics like tracking, online behavioral advertisements, and legal compliance. We found that in many papers, users' consent is the central point of investigation. However, this perspective is not sufficient to solve the underlying issues, as we have to deal with a complex ecosystem and conflicting interests. Building upon this, we outline future venues of research and discuss the responsibilities of different stakeholders. The goal of this work is to bring attention to and ultimately make progress towards stopping the consent theater.

3.2. How could it get so bad?

Internet sites have begun early on to finance their content with advertisements. Users understood and accepted that trade as consuming ads while getting content or functionality for free [277]. The nature of this accepted trade changed after the dot-com bubble burst. Google, which was under pressure to generate revenue, discovered that they could use the behavioral surplus (i.e., users' excess behavioral data that the service itself does not utilize) from their search engine to enhance the quality and value of online behavioral advertising (OBA). This triggered the discovery of surveillance capitalism's foundation: The endeavor to corner the market for big-data-based prediction products requires an arms race to collect and accumulate increasing amounts of behavioral surplus [309]. Since then, surveillance capitalists do not build new tech products for the benefit of their users but to increase their access to behavioral surplus.

Resulting from this development, website providers who use ads to monetize their content now also hand over their users' behavioral surplus (i.e., who viewed their content at which time) to advertisers. Figure 3.2 provides an overview of personal data flows

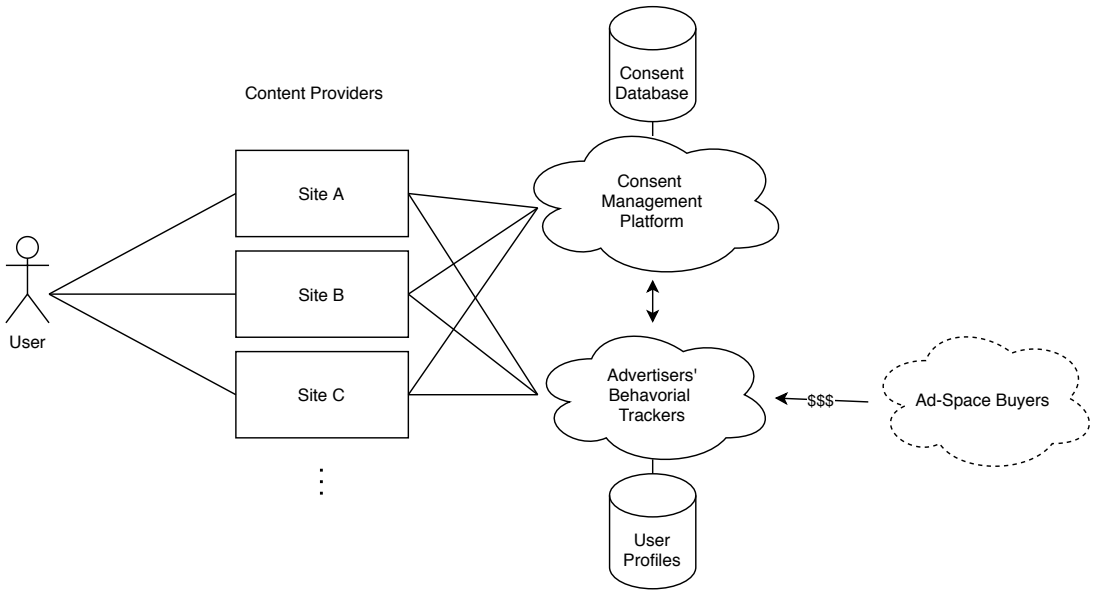


Figure 3.2.: Online behavioral advertising (OBA) and all involved parties

in online behavioral advertising. In all cases, the advertisers embed code into the content providers' site. This code asks for the users' consent (optionally using consent management platforms) and only then contacts the advertisers' behavioral trackers. The trackers identify the visitors across multiple websites and use this access information to build users' profiles. Increasingly detailed user profiles make the ad-space more valuable since trackers can use them to predict the advertisements' effectiveness reliably. Commonly, trackers use regular third-party cookies to identify visitors. However, the user tracking does not depend on them since other mechanisms such as cookie-synchronization, browser fingerprinting, super-cookies, or Verizon's Unique Identifier Header (UIDH) [286] achieve the same effect. After the behavioral tracker identifies a visitor, a real-time auction (based on the visitor's profile) determines which advertisement they will see. This type of online behavioral advertising (OBA) has been mostly invisible to users since browsers just accepted any cookies by default. However, the ePrivacy regulation and GDPR made these practices more visible since users need to consent to them.

The two main stakeholders in OBA are the content providers' sites and their visitors. Visitors want to consume content or access functionality, and sites want to monetize their content. Other stakeholders, apart from these two, are crucial for the business model of OBA. The ad-industry discovered that sufficiently large datasets on users' online behavior allow behavioral prediction and behavioral modification. This extraordinarily lucrative opportunity drives the ongoing efforts to collect increasingly more behavioral data since even small improvements in predictions are worth millions [309]. Regulators and policymakers have found that these data practices negatively impact users' right to self-determination [213, 76]. This is exemplified by the Internet users who are, despite the "consent" mechanisms, unaware of how their personal data is used [277, 193]. Currently, their response consists of a policy to increase awareness and repeatedly emphasize that users have the freedom of choice of being surveilled or not (in the form

of the ubiquitous consent notices). In light of the significant fines that GDPR allows, consent management platforms have become popular. They are used to ease advertisers' legal liability while simultaneously optimizing users' consent rates for profit.

The fundamental conflict of interest in OBA is that trackers base their entire business model on detailed user profiles that demand ever-increasing behavioral data. All the while visitors feel uncomfortable with that level of tracking and regulators want to reestablish reasonable ways of informational self-determination. This conflict of interest is difficult to untangle since many current internet sites' monetization depends on it.

3.3. Observations from prior security and privacy research

The introduction of the ePrivacy directive and GDPR reignited research interests in the topic of online consent mechanisms. Most of the research surrounds questions of regulatory effectiveness and users' perceptions of consent notices and data practices. We reviewed and analyzed prior work to see how researchers frame the continued problems, which systematic issues emerge, and which kinds of questions researchers pose. In this section, we present three key-issues we identified and take a closer look at solutions proposed in prior work.

3.3.1. Key Issue 1: Users largely do not understand that behavioral online ads require tracking

Research has invested considerable efforts into understanding the internet users' perceptions and behaviors concerning online tracking in general [39, 159, 193] and behavioral advertising in particular [277, 302, 188]. The findings paint a complicated picture and conclude that user decisions are highly individual and context-dependent, while at the same time exposing little understanding for the complexity of web tracking. Melicher et al. investigated users' contextual preferences for web tracking and found that users base their decisions to accept targeted ads mostly on concerns about functionality [193]. Complementing, people saw value in user-customized targeted ads [39, 193] and behavioural advertising [277]. However, acceptance of online behavioral advertising is highly complex and context-dependent [277]. As Melicher et al. found, while 74% of their participants preferred targeted ads, 60% considered targeted ads to be harmful in at least one tracking scenario [193]. Apart from the nuanced attitudes towards targeted online advertising, people often do not connect it with the concept of tracking. That is, McDonald et al. found that while their participants in principle approved the idea of advertisement funding free online content, they did not expect to be tracked in the process [188]. Research on users' perception of tracking found a general lack of information and understanding of the technologies and concepts involved [193, 159, 302]. For example, Kulyk et al. found that people do not understand how cookies work, what that meant for their life online, and what potential countermeasures look like [159]. If asked how tracking works, people reveal concerns about their ability to make sound decisions given their limited knowledge about the topic [193]. Yao et al. found that even tech-savvy people, such as web developers, did not know how online tracking works [302]. In their study, they identified four "folk models" about how online

behavioral advertisement works and found these models to be either incomplete or inaccurate. To name a few misconceptions, people thought trackers are hackers, viruses, or have access to local files on the computer. If people have mixed opinions about targeted advertisements and do not fully understand the workings and implications of tracking, which factors do influence their tracking preferences? Research identified external factors about the visited websites and the tracked data. On the one hand, trustworthiness of the webpage, familiarity to the user, and importance of the web page's contents [159]. On the other hand, the kind of data plays a role, such as personal or search information, correspondence, financial, educational, or health information [193]. Finally, Melicher et al. found that awareness and consent to tracking impacts people's comfort with such practices [193]. However, this is problematic and might promote a false sense of privacy. Also, it hints at the core of the tracking issue: you cannot consent to something you do not understand.

Key Issue 2: The ad-industry carefully crafts consent mechanisms to deceive users

In the opinion of Acquisti et al. [6], neither paternalistic (government regulation) nor strictly libertarian (self-regulatory) privacy rules guarantee desired privacy outcomes. Instead, Acquisti et al. suggest that soft-paternalistic approaches (i.e., nudges) may bridge the gap between both approaches. In 1997 advertisers fought bitterly against a proposal by the FTC that would have assigned users control over their personal information by default [309]. Instead, they formed the Network Advertising Initiative that promised to regulate the industry's behavior. Garlach et al. [99], studied the effects of NAI's self-regulatory notice-and-choice model and found that users did not see the notice and did not understand its purpose. They found it hard to reconcile that the industry has expertise in creating noticeable messages and chose not to create noticeable and informative privacy notices.

Suspiciously, this is a re-occurring pattern in related work. User-interface elements that would give users more options to opt-out of tracking use confusing and and inconspicuous design. Acquisti et al. [6] described different modes of soft-paternalistic nudges: *information*, *presentation*, *defaults*, *incentives*, *reversability*, and *timing*. Privacy-related research has found evidence of most of these nudging modes. Information: Machuletz et al. [172] recognized a proliferation of choice in consent notices and found that users disengage if more than a few data collection purposes are available. Hence, they warn that businesses use the flexibility in designing consent dialogues for their interest by maximizing data disclosure. Presentation: The ad choices icon' design uses unobtrusive colors and sizes and hard to understand phrasing [99]. Commercial software has habituated users to accept anything that resembles an end-user license agreement [31], we have to assume that this is similar for cookie consent notices. The most popular consent notice designs are not the most effective ones [191]. Defaults: Half of the websites do not have a reject all button, and 75% of the sites that do bury this option – requiring more clicks than the alternative. Additionally, 56% of sites pre-tick optional vendors [204]. Timing: Consent-management platforms sometimes ask users excessively for their consent decision (which they already have), bordering on "consent harassment" [184].

The ad-industry seems to use the promise of self-regulation and user consent to avoid stricter legal regulation while simultaneously using nudges and biases to continue their business model. Consequently, we cannot trust the advertisement industry's mechanisms that ask users for consent. On the contrary, we have to assume all parts of these mechanisms use meticulous design to confuse users and increase "consent" rates.

3.3.2. Key Issue 3: The advertisement industry tries to avoid litigation by providing consent mechanisms on the surface, but often disregards the choice in the implementation

The introduction of GDPR did not substantially impact web tracking. Although its intention was strengthening users' right to privacy, researchers found no decrease in web tracking in the European Union [56]. Degeling et al. argue that the supposed increase in transparency may even lead to a false sense of privacy and security [56]. On top of that, they found that few websites offer meaningful cookie choices to control tracking. Additionally, Mehrnezhad found inconsistencies in the presentation of privacy notices and banners within and across platforms [191]. In desktop and mobile browsers, as well as mobile applications, consent notices often do not comply with GDPR since they tie tracking to page load, not user consent. Several works found non-compliance to GDPR and violations of user consent [184, 249, 274]. Millett et al. identified four types of potential GDPR and the ePrivacy directive violations of European websites which are specific to consent notices: (1) storing consent before the users' choice, (2) not providing users a way to opt-out, (3) pre-selected cookie choices, and (4) not respecting the users' choice [184]. Similarly, Trevisan et al. conducted a large-scale measurement of 35.000 European websites to check the implementation of the EU cookie directive [274]. They found that half of the websites install tracking cookies before users provide their consent. On a global scale, Sanchez et al. investigated users' ability to opt-out of tracking [249]. According to them, opt-out mechanisms are often not correctly implemented and confront users with deceiving information. Additionally, long-lasting cookies are widely used, despite users having opted-out of tracking.

Research demonstrates that users have to face a heterogeneous ecosystem of consent notices that often does not respect their choices, even though they make one.

3.3.3. Proposed solutions to the flood of cookie decisions

Prior work has tried to find solutions for the flood of cookie decisions for 20 years now. Most of them have tried to make the individual users' cookie decisions easier or more understandable. However, a few also tried to tackle the problem not from an individualistic perspective but rather from collective enforcement of existing laws.

The Netscape browser implemented cookies first in 1994 to enable virtual shopping carts for e-commerce applications – resulting in an increasing burden on users to make cookie-related decisions. In the early 2000s, it was already clear that this burden is growing too large for them since they had little information about these cookies' purposes, and the sheer number of cookie decisions created habituation effects [197]. Friedman et al. [92] applied value-sensitive design to create a browser-integrated cookie information interface to help users to keep track and manage their cookies. Similarly, Kulyk et

al. [160] found that cookie-related browser settings confuse users. They designed an alternative settings interface to bridge the gap between the users' privacy preferences and the rather technical jargon of cookie settings. An alternative approach to the problem of increasing cookie decisions is automation. Shankar et al. [255] and Yue et al. [303] identified the problem that users face too many cookie decisions and suggested a tool that only accepts cookies required for a site's functionality. These kinds of solutions can be efficient by applying a pre-defined policy to all decisions. The authors' focus on functionality does not necessarily improve users' privacy, which opens up the question of suitable cookie decision policies. Such automation solutions provide an efficient way of dealing with large number of cookie decisions by moving from individual consent to generally applicable policies. This approach leaves room for a discussion about the types of policies that fit the users' requirements in their different contexts of use.

One thing all of these approaches have in common is that they come with an additional burden to users. They assume that it is the users' responsibility to understand the purposes of cookies and set them in a privacy-aware manner. Even automatic approaches to cookie management require users to handle edge-cases where pre-defined policies fail. However, other stakeholders such as the browser vendors, ISPs, websites, and law enforcement can combat the flood of cookies – by focusing less on individual users' behavior and more on the behavior of other stakeholders has the potential to improve the current status of privacy on the web substantially. An example of this kind of approach to improving cookie privacy is Trevisan et al.'s large-scale measurement [274]. They measured sites' cookie handling and found many configurations that do not respect the ePrivacy directive. Their audit tool, CookieCheck, automatically verifies legal violations and simplifies the enforcement of existing regulations on a large-scale, independent from individual users' behavior.

3.4. Questioning the ubiquitous paradigm of consent

In 1983 the German Federal Constitutional Court identified the fundamental right to informational self-determination in the context of modern data processing. The decision enshrined the users' right to decide if and how their personal data may be used. Since then, policymakers, privacy advocates, and privacy researchers alike focus on the issue of informed consent to data practices. Achieving informed consent requires that users (1) understand how their data will be used, and (2) agree explicitly and voluntarily to the use of their data. Both of these criteria have become increasingly difficult to achieve since the complexity of data practices has increased significantly, and privacy-infringing data practices have become so ubiquitous.

3.4.1. Informed consent to online data practices is difficult to achieve

Users have nuanced and context-dependent views on their acceptance of online behavioral advertising, understanding sites' need for monetization, and additional functionality enabled by tracking [39, 193]. However, they do not seem to fully understand that personalized ads rely on their tracked online behavior, how widespread these trackers are, which information these trackers can access, and what types of personal information

advertisers can derive from their behavior [277, 302]. Users can, by definition, not give their informed consent to data practices they do not understand.

According to GDPR rules, user consent legitimizes most forms of privacy-infringing data practices. Hence, achieving a high rate of consent is a goal with direct monetary value for advertisers. Advertisers achieve high rates of "consent" by employing several different kinds of dark patters in the user interfaces of online consent notices [172, 31, 99, 191, 204]. Optimizing consent in this manner protects advertisers' business model while keeping up the facade of lawful and effective informational self-determination. Efficiently optimizing consent rates opened up a new business opportunity for so-called consent management platforms (CMP). "OneTrust, on its webpage presenting its CMP solution, proposes publishers to 'maximize user opt-in with customizable publisher-specific cookie banners [...] to optimize consent collection' " [184].

Aside from these fundamental issues with consent to online data practices, there is also the issue of the sheer number of consent decisions. Since these habituate and disengage users from consent requests in general, it will be necessary to reduce their number significantly.

3.4.2. Reduce overall number of consent decisions

The current state of the web requires a high number of consent decisions from users. Aside from an increased cognitive load, they also habituate users to accept more cookies than they feel comfortable with. To avoid user disengagement, these consent decisions need to be reserved for impactful, non-trivial, and context-dependent situations as Boehme et al. noted: "a last resort to prevent habituation is economizing consent decisions and thus reserving users' scarce decision capacity for the really important choices." [31] Many consent notices unnecessarily ask users to accept data practices that are essential for the site's functionality (such as virtual shopping carts or remembering the login status). We interpret this as a "proliferation of choice" tactic to overwhelm users since these cases do not even require consent according to GDPR. Additionally, there the consent decisions regarding tracking practices that users almost universally decline if they do not see a clear benefit to them.

Some tools that minimize consent notices already exist [303, 255, 217, 156, 1]. These apply predefined policies and thereby move away from individual consent decisions. However, these predefined policies do not necessarily care for privacy concerns and instead focus on the sites' functionality, e.g., by agreeing to all data practices to hide the consent notice.

Even if predefined decision policies handle most of the consent notices, there remains a small set of consent decisions that are non-trivial and can be very personal.

For these context-dependent decisions, users might want to weigh the site's perceived trustworthiness and the benefits of accepting behavior tracking (maybe additional functionality or monetary value) with the privacy-infringing drawbacks. In case the behavior tracking has a direct effect on the website's functionality (e.g., in the case of personalized search), the website should not be allowed to deny service, and instead, provide a gracefully degraded level of service to users who reject tracking. In any case, these remaining consent notices need a unified presentation so that users can understand

and compare important points quickly [152].

3.4.3. Alternative approaches to improving privacy on the web

Achieving informed consent to data practices is not the turnkey solution to online privacy that we desire. The individualistic right to informational self-determination does not fight privacy-infringing behavior itself. Instead, it provides privileged, informed, and motivated individuals a seemingly neutral decision if and how they want to be under surveillance. As long as a sufficient number of users "consent" to surveillance, the underlying privacy-infringing business model continues. More regulation on how companies are not allowed to track users and ask users for consent could turn into an elaborate cat-and-mouse game. Hence, improving privacy on the web as a whole can not rely on informed consent alone.

One of the reasons for ubiquitous consent decisions on the web is that users' consent legalizes data practices that would otherwise undergo much more scrutiny. We can not rely on informed consent to improve privacy on the web. Hence, we need to make these privacy-infringing data practices more transparent, discuss the potential consequences, and use government regulation to keep them in check – independent of users' "consent" decisions. A core issue for regulators should be the market for predictive futures since it drives the urge to collect ever-increasing mounds of behavioral data.

3.5. How future research could counter surveillance capitalism

As we discussed in detail in Section 3.3, prior research has invested significant work into describing, measuring, and improving issues revolving around tracking, targeted advertisements, and cookie banners. These works built the foundation for further discussions and raised awareness for the complexity of the advertisement industry and its stakeholders' conflicting interests. While identifying the oppressive mechanisms of the status quo is difficult already, changing is even more so. As discussed by Keyes [153] and Asad [17], prefigurative politics is a viable method of engaging with power. It is a pragmatic and applied process-driven approach to producing counter-power and counter-structures in the here and now. This section concerns itself with smaller-scale future research directions that could scale up to engage with the power behind current privacy issues to realize online privacy eventually.

3.5.1. Shift Focus Away From Consent

Users' consent has been a focal point of prior research, not least because it is a central aspect of privacy laws. However, as Trevisan et al. pointed out, the concept of consent alone might not be sufficient to solve the issues around behavior tracking [274]. Though consent is a necessary aspect, research can benefit from looking at the problem from a different angle. For example, future work can focus more on restricting the privacy-infringing data practices themselves without depending on the users' consent.

3.5.2. Examine Understudied Parties

So far, research has focused on end-users, CMPs, and the technical implementation of consent notices on websites. This focus has avoided some of the stakeholders involved in online tracking and advertisement. Consequently, the perspective of the advertisement industry and the content providers remain understudied. Of these two, the advertisement industry seems to use insights from research to maximize their profits by optimizing users' consent rates. Hence, we do not deem it worthwhile to invest further resources into understanding their point of view. However, understanding the perspective of content providers might provide valuable information to improving users' privacy on the web. One potential area of research is a detailed understanding of the content providers' motivation behind incorporating tracking mechanisms. Additionally, we consider it urgent to examine privacy-friendly forms of monetization and provide incentives to content providers to shift to them.

3.5.3. Individual and Collective Solutions

In our opinion, resolving the issues around tracking and cookie banners requires a combination of individual and collective solutions that address different aspects of the problem. Individual solutions target end-users. These can be plugins that help users detect tracking or make monitoring visible. Additionally, they can remove the burden of trivial consent decisions from the user and only call them to action when it is needed. Nowadays, we already have such tools, usually third-party plugins or tools that proxy the users' decisions. From the perspective of the websites, consent management platforms (CMPs), which handle the consent management for website owners, are on the rise. The rise of CMPs is somewhat counterintuitive, as the sole idea behind GDPR and similar laws is to make the process of tracking and data collection more transparent. However, it perhaps opened up business opportunities for even more parties to get involved with user data. Before privacy regulations came into effect, advertisers gathered people's data silently without their awareness. While ubiquitous consent notices made the data collection very annoying nowadays, they did not seem to have curbed it significantly.

We are in desperate need of collective solutions that target the issue on a large scale. For example, legal authorities need to be equipped with suitable measures to detect and report violations. Trevisan et al.'s CookieCheck tool is a step in the right direction [274]. Additionally, we are convinced that the different technological parties need to get involved in solving the issue, such as browsers, search engines, and social networks. One of the biggest players, Google, has a special kind of responsibility since they have one of the most popular browsers and are also heavily invested in the ad industry. Past efforts of Google in cooperation with other tech giants have demonstrated a strong ability to transform the Internet's ecosystem. For example, joined efforts resulted in a switch from HTTP to HTTPS, which at the time was not widely adopted [266]. They could also significantly contribute to an improvement of the current tracking situation if they wanted to. However, calling them into action will undoubtedly be difficult as such endeavors counteract their business model.

3.5.4. About Science Communication and Asking the Right Questions

We have been conducting the same research year after year – and while it is undoubtedly important to measure the status quo, we have not been able to counteract these privacy-invasive tendencies. There were also early efforts to propose design guidelines, driven by value-sensitive design, which are still relevant today [92, 197]. Unfortunately, many of today's practices violate these guidelines. We see similar tendencies also in other fields of research. For example, in the field of email encryption, essentially the same research is conducted over and over again, but at the core, there has been little progress over the last 20 years. We, as a scientific community, must try to change our perspective. If research's efforts are not being heard or do not have the desired effects, we need to ask ourselves two questions: First, are we properly communicating our results to the public and relevant authorities to make changes? Especially when research involves practical applicability and can therefore be fast-moving, we need to pay special attention to science communication. And secondly, are we potentially asking the wrong questions to begin with? Implicit assumptions for which we have no evidence and which may not be correct could, for example, mislead us to ask the wrong questions. Change is possible, but it is usually an arduous process that brings together the combined efforts of research, jurisdictions, and industry.

3.6. Acknowledgement

We appreciate the fruitful discussion we had with Gabriel Grill (University of Michigan) on current and historic consent issues. He introduced us to the term "consent theater" which inspired this paper's title.

The contents of the following chapter were published as part of the publication *"Comparing User Perceptions of Anti-Stalkerware Apps with the Technical Reality"* (SOUPS 2022) [P2]. This paper was produced in cooperation with my co-authors Simon Anell, Sabine Houy, Martina Lindorfer, and Katharina Krombholz. The following table describes the contributions of all authors to this paper:

Author	**Contribution**
Matthias Fassl	I came up with the idea for this research project and chose the methods for investigating the research questions. I guided Sabine through her bachelor's thesis, analyzed the app-store reviews together with Simon, conducted the UI walkthroughs, and drafted the final version of this paper.
Simon Anell	Simon contributed to the qualitative analysis of the app-store reviews.
Sabine Houy	Sabine conducted her bachelor thesis under my guidance in this research project. The insights from her bachelor thesis are now part of the 'Anti-Stalkerware under the Hood' section of this chapter.
Martina Lindorfer	Martina contributed her expertise in mobile security and privacy and, in particular, malware analysis to this research project. She suggested missing literature for the related work section and revised the 'Anti-Stalkerware under the Hood' section of this chapter.
Katharina Krombholz	As my academic advisor, Katharina was involved in the major decisions during this research project. She gave me feedback on the initial idea and the choice of studies, including the methodological approach. She guided me through methodological and ethical questions, reviewed my interpretation of (intermediate) results, and helped me find an appropriate venue for this research project. She also gave feedback on draft versions and edited parts of the paper.

Reference

Fassl, M., Anell, S., Houy, S., Lindorfer, M., & Krombholz, K. (2022). Comparing User Perceptions of Anti-Stalkerware Apps with the Technical Reality. Proceedings of the Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022), 135–154. `https://www.usenix.org/system/files/soups2022-fassl.pdf`

4

Comparing User Perceptions of Anti-Stalkerware with the Technical Reality

Every year an increasing number of users face stalkerware on their phones [45]. Many of them are victims of intimate partner surveillance (IPS) who are unsure how to identify or remove stalkerware from their phones [163]. An intuitive approach would be to choose anti-stalkerware from the app store. However, a mismatch between user expectations and the technical capabilities can produce an illusion of security and risk compensation behavior (i.e., the Peltzmann effect).

We compare users' perceptions of anti-stalkerware with the technical reality. First, we applied thematic analysis to app reviews to analyze user perceptions. Then, we performed a cognitive walkthrough of two prominent anti-stalkerware apps available on the Google Play Store and reverse-engineered them to understand their detection features.

Our results suggest that users base their trust on the look and feel of the app, the number and type of alerts, and the apps' affordances. We also found that app capabilities do not correspond to the users' perceptions and expectations, impacting their practical effectiveness. We discuss different stakeholders' options to remedy these challenges and better align user perceptions with the technical reality.

4.1. Introduction

About one in five adults and even more young adults engage in snooping attacks on others' phones [176]. *Intimate partner surveillance* (IPS) is a specific subset of these attacks [25, 275]. Tool-based IPS often involves a type of spyware, called stalkerware (or surveillanceware), to collect live location data, contacts, call history, and text messages [41, 244]. According to the Coalition Against Stalkerware [45], 67,500 mobile users were confronted with stalkerware in 2019, a 67% increase compared to the year before. Randall et al. [233] estimated that at least 5,758 people in the US were targeted by overt stalkerware from March to May 2020. Two of the 22 apps they studied were available in the Google Play Store, the remainder were only available from third parties.

In October 2020, Google banned surveillance apps from their store [109] and now only allows surveillance in parental control and enterprise management apps if they do not hide or obfuscate their surveillance practices. Hence, stalkerware often rebrands itself as parental control apps or moves to third-party websites. Most stalkerware occurrences in *clinical computer security* [134] consultations comprise such "dual-use" apps [41].

An analysis of online domestic abuse forums and an assessment of the stalkerware application (app) industry identified that IPS survivors are unsure how to recognize and remove stalkerware [163, 214]. Installing anti-stalkerware apps from the Google Play Store is one possible approach. Users may choose from various apps, ranging from traditional anti-virus companies offering general mobile security solutions to specialized apps detecting stalkerware and other spyware. Prices vary widely, some are as cheap as € 5 (or $), but in-app purchase prices up to and beyond € 100 (or $) are not uncommon. However, these apps come with severe limitations on Android since they often operate with simple name-based blocklists, which stalkerware can circumvent easily [46]. More worryingly, there have also been instances of fake anti-virus apps in the Google Play Store with limited to no functionality at all [136, 305, 209, 65]. Thus, the marketed promise of identifying stalkerware is at odds with many of these apps' abilities, constituting an expectation-ability gap. This problem affects users' ability to make informed decisions. Survivors should be made aware of these problems to allow them to question their reliance on them.

We conduct an exploratory case study with two anti-stalkerware apps to understand this mismatch between expectations and abilities. We focus on the following research questions: (RQ1) *What are the differences between users' security perceptions and the anti-stalkerware apps' abilities?*; and (RQ2) *How could research and design begin to remedy this mismatch and foster users' anti-stalkerware decisions?* We apply thematic analysis to app-store reviews to study perceptions of these apps. We also perform a cognitive walkthrough of the respective apps and then reverse engineer them to understand how their detection mechanisms work. Hence, we elicit expectation-reality mismatches by combining qualitative user research with a reverse-engineering approach. Based on app reviews, we identified five user approaches to building confidence in their anti-stalkerware choice, all of them intuitive to apply and with some degree of legitimacy. However, contrasting these approaches with the cognitive walkthrough and reverse engineering results demonstrates that they fail to inform users about apps' abilities to mitigate violence, abuse, and harassment. Our work helps improve the current state of anti-stalkerware by suggesting design directions, proposing toolkit-supported user decisions, and discussing systemic, platform-level approaches to combating intimate partner surveillance.

4.2. Background and Related Work

This section describes background information and prior work on intimate partner surveillance and our methodology.

4.2.1. Intimate Partner Surveillance

Insiders, i.e., persons who are familiar to the victims, are a threat to smartphone users that security experts underestimated in the past [200]. Insiders' access to victims' devices varies significantly. However, according to one study in the US, 31% of participants looked through others' smartphones without their permission [176]. Surveillance among intimate partners, a specific insider attack, is usually technically unsophisticated and relies on UI-bound attacks or ready-made apps [90]. Bellini et al. [25] and Tseng et al. [275] analyzed stories on online forums about sexual infidelity. Abusers justify their surveillance with their suspicion of sexual infidelity. They want to collect evidence, understand behavior, and control behavior [25]. Bellini et al. [25] identified a four-stage abuse cycle: setting the abusers' expectations, attitude change, escalation, and reflection. Tseng et al. [275] categorized IPS attacks based on physical and non-physical access requirements. They found that online communities are a good source of IPS threat intelligence because their users collaborate to create new IPS attacks.

Chatterjee et al. [41] identified apps that are dangerous in the IPS context. They found explicit spyware apps and more subtle dual-use apps with legitimate use-cases (e.g., FindMyFriend). Often, anti-spyware does not identify the latter as a threat. Parental control apps, a classic example of dual-use, also suffer from other privacy issues, e.g., collecting sensitive data and distributing it to third parties without consent [78]. To understand the "creepware" ecosystem, Roundy et al. [244] developed the *creeprank* algorithm based on guilt by association. As a result, hundreds of apps were removed from official app stores and presumably moved to third-party repositories.

Based on survivors' stories, Matthews et al. [185] identified different phases of separation and technology use. Survivors' safety in the "life apart" phase depends on identifying stalkerware. Havron et al. [134] and Freed et al. [89] created a computer security clinic for IPS survivors who readily accepted support in this format. However, since anti-stalkerware apps have a low barrier for entry, survivors presumably also use them as part of their protection ensemble. Lee et al. [162] extended the theory of planned behavior to understand factors leading to anti-spyware software adoption.

4.2.2. Review Mining and Analysis

App-store reviews inform users about the apps' quality, but also developers about bugs and feature requests, as well as researchers to gain detailed insights about apps. In light of the sheer number, informality, and shortness of these reviews, researchers either mine reviews to get a broad overview or use thematic analysis to examine a subsample in rich detail.

The software engineering community explored automated ways to mine user reviews for actionable development feedback. Prior work discussed several different automatic approaches to identify informative complaints in app reviews [219, 42, 95, 171]. Khalid et al. [154] used manual qualitative analysis to identify complaints about iOS apps.

Others have focused on automatically retrieving feature requests from reviews [171, 144] using natural language processing, sentiment analysis, and LDA models. Automatic analysis of app reviews can also inform developers about usability and user experience issues [19, 190, 210, 135]. Gu et al. [120] and Guzman et al. [122] applied sentiment

analysis to understand how users feel about apps and individual features.

Researchers have also used reviews to study security- and privacy-related aspects of apps. Ha et al. [123] manually coded reviews to look for security and privacy complaints and found that about 1% of them concerned app permissions. Nguyen et al. [203] analyzed reviews for security- and privacy-related reports and traced 61% of security and privacy updates to corresponding user reviews. Voskobojnikov et al. [288] analyzed cryptocurrency wallets' reviews to understand security- and privacy-relevant UX issues. They identified a subsample of relevant reviews using machine learning and natural language processing and then applied thematic analysis. Gosh et al. [104, 103] qualitatively analyzed reviews of parental control apps to understand how children responded to them. They used a keyword search to filter children's reviews and applied thematic analysis. Children found the apps overly restrictive and privacy-invasive. They criticized their parents' reliance on these apps as a bad parenting technique.

4.2.3. Spyware Detection

In general, there are two basic approaches to detecting and analyzing malware, including stalkerware: static and dynamic analysis [9]. Static analysis is the understanding of a program at the syntactic source code or binary level [96]. Dynamic analysis focuses on an app's run-time behavior, including system calls and network traffic. For this purpose, researchers execute and observe apps in controlled environments [167].

Knowing the reliability of on-device anti-malware scanners (commonly referred to as anti-virus) is crucial for end users' safety. These scanners base their detection mechanisms on either static or dynamic analysis. However, compared to security solutions on desktop operating systems, mobile security apps have limited visibility into other apps due to extensive sandboxing, rendering behavioral heuristics unfeasible [166, 43, 221, 222]. Security solutions thus have to rely on signatures based on code-level characteristics or use machine learning [16, 166]. Related work has investigated in-depth how easy it is to evade those signatures [234, 306, 18, 126, 222]. Yet, no study so far compared the robustness of detection mechanisms to the trust users put into these security solutions.

4.3. Methodology

We explore the gap between users' expectations of the apps' functionality and the apps' technical abilities. Understanding this mismatch helps to improve users' protection against stalkerware. First, we apply thematic analysis [34] to app-store reviews of the two case-study apps to understand users' security perceptions and expectations. Based on the resulting themes, we perform cognitive walkthroughs of the apps and analyze them to understand how they detect stalkerware.

Selection of Anti-Stalkerware Apps

Spyware poses an increased danger to Android users compared to iPhone users [128, 214]. Apple's iOS claims tighter security controls [14] and does not allow apps with

"functionality it does not actually offer (e.g., iOS-based virus and malware scanners)" [15]. Hence, we focus on Android apps.

To cover a variety of app abilities and user expectations in our qualitative analysis, we base our selection on Chatterjee et al.'s anti-spyware list [41]. From the most-downloaded anti-stalkerware apps, we chose two to perform static analysis on: Mobile Security, Antivirus & Cleaner by Lookout[1] (100M+ installs) [114]. From the long-tail, we read app-store pages and chose a data-rich example suitable for further qualitative analysis: Anti Spy Mobile PRO[2] (100k+ installs) [113].

Fraudulent reviews and manipulated ratings plague free apps [228, 300, 301]. Therefore, we prefer to analyze reviews of paid apps. Lookout Mobile Security is free to download on the Google Play Store and uses an in-app subscription model. We can not differentiate between subscribed and unsubscribed users' reviews. Hence, we also analyzed reviews from unsubscribed users. Lookout Mobile Security is more extensive and complex than Anti Spy Mobile PRO. Lookout Mobile Security markets itself as a fully-fledged security solution, with anti-spyware as only one of its features. In contrast, Anti Spy Mobile is available as a free or paid version (€ 4.90 or $ 3.99) on the Google Play Store. The only difference is that the paid version has automatic daily background scans. We only analyzed the paid version's reviews.

The focus on these two apps affects the results twofold: First, their features are not representative of all security apps marketed as anti-stalkerware. Second, Lookout Mobile is pre-installed for some users, so the lack of choice may impact users' reviews. Hence, reviewers' sentiments from these two apps are not generalizable to all security apps that market themselves as anti-stalkerware.

Analysis of App-Store Reviews

To understand how users perceive our case study's anti-stalkerware apps and engender trust in them, we applied thematic analysis [34] to a sample of their app-store reviews.

We fetched all reviews from the Google Play Store.[3] We randomly sampled 200 comments from each app in German and English, languages all involved researchers understand well. To ensure the reviews had enough content, we only considered comments with at least ten words. We analyzed a total of 400 reviews for Lookout. Anti Spy Mobile PRO, had less than 200 reviews fulfilling our criteria, so we analyzed a total of 13 German and 102 English reviews for this app.

At the start of the thematic analysis, one researcher read all reviews and created an initial codebook. With it, both researchers coded the entire review sample. During the coding procedure, both researchers kept notes on potential themes in the data. This resulted in an inter-coder agreement of Krippendorff's alpha $\alpha = 0.86$, which suggests *excellent* agreement. Afterward, the researchers discussed all mismatches and the themes they identified. Vague reviews with multiple valid interpretations caused most of the disagreements. Resolving conflicts increased Krippendorff's alpha to $\alpha = 0.98$. Table B.1 in the Appendix presents the initial codebook.

[1] Version: 10.33-6652654, Downloaded: June 2020
[2] Version: 1.9.10.51, Downloaded: June 2020
[3] Anonymized JavaScript code: `https://pastebin.com/bRZ1v0XS`

The discussions led both researchers to agree on a focus on safety and security perceptions. We repeated the above procedure and constructed an additional codebook. Krippendorff's alpha was $\alpha = 0.78$ after the initial round of coding, suggesting *substantial* inter-coder agreement. Discussing all mismatches increased Krippendorff's alpha to $\alpha = 0.96$. At the start of the discussion, the researchers added a "time of experience" code and applied it whenever appropriate. Table B.2 in the Appendix presents the revised codebook. Afterward, both researchers discussed the identified themes and the presentation of the results.

4.3.1. Anti-Stalkerwares' Technical Abilities

After identifying security perceptions and expectations, we used *theoretical sampling* to understand these apps' technical abilities. Thus, we collected data about the user interface and the apps' internal detection mechanisms.

We conducted cognitive walkthroughs for both apps to improve our understanding of the reviews focusing on user experience. Based on the previously discovered themes, we focused on the following: (1) method of invoking scans (manual, scheduled, event-triggered), (2) type and amount of information in reports, (3) false positives in a general use scenario, (4) visible user interactions under regular usage. We screenshot these parts of the case study apps and deductively code them with the codebook from the review analysis.

Additionally, we reverse-engineered the case study apps to understand how they detect stalkerware. In both cases, we started with *static analysis*, i.e., decompiling and inspecting their source code. We used *dynamic analysis* to verify the results and to understand run-time behavior. This allowed us to observe and inspect the output of the apps' scanning and evaluation functions for potentially harmful behavior.

4.3.2. User Perceptions vs. App Capabilities

Finally, we juxtapose the trustworthiness and security perceptions with theoretical samples from each case-study app to point out mismatches between perceptions and technical reality. As far as possible, we embed the perceptions and theoretical samples into related work to provide an additional broader context. We evaluate the benefits and drawbacks of users' strategies for choosing anti-stalkerware.

4.3.3. Ethical and Legal Considerations

Using public data for research without explicit consent is an ethical challenge, especially concerning intimate partner abuse. Even though users can remove their public reviews, we handle all data with care to minimize potential harm. We omit usernames and rephrase quotes if they contain hints of abusive behavior, rendering identification difficult.

Reverse engineering is a legal grey area. In the US, good-faith security research is exempt from copyright law and the DMCA [212]. In the EU, decompilation is explicitly allowed to ensure interoperability with other software [287]. EU copyright law only protects the concrete expression of the source code, not the underlying ideas and

principles. We carefully reviewed our results to avoid publishing information that could be considered a concrete expression.

We want to minimize potential harm from publishing results of our technical analysis. After a careful review, we identified three types of potentially harmful information: (1) well-known stalkerware that apps do not identify correctly, (2) flawed general approaches to detecting stalkerware, and (3) specific implementation details about threat classification. We informed the app providers about well-known stalkerware their app did not identify before publication. The general flaws we identified are well-known; existing spyware and state-of-the-art anti-spyware already take them into account. Hence, publishing these general flaws does not introduce new harm. Specific implementation details on how apps classify threats are out of scope for this work. Since stalkerware could use these findings to evade detection, we refrain from publishing them. Our institution's ethical review board (ERB) approved this study.

4.4. Users' Perceptions of Anti-Stalkerware

To understand how users perceive the security of anti-stalkerware apps, we analyzed the app-store reviews of the two apps in our case study. We included a total of 518 reviews in our study and performed thematic analysis to find higher-level themes and patterns in the data. In the following, we report the results from this analysis, i.e., our findings on users' approaches to engendering trust in anti-stalkerware apps, general observations, and contradicting user expectations.

We identified five approaches users apply to convince others of anti-stalkerware apps' usefulness and trustworthiness.

Potentially harmful incidents. First-hand experience of an apps' protection is a popular way for users to establish trust. This approach to establishing trust covers a variety of different features. Amongst others, we have found praise for adware detection, e.g., *"has already found and removed adware three times." (R326)*, spyware detection, e.g., *"Someone had put a tracking app on my phone [...] I had it figured out in about 10 minutes!" (R425)*, and theft prevention, e.g., *"It [...] has saved me from losing my phone not once but twice to thifes." (R132)*. Interestingly, reviewers did not seem concerned about apps' potential shortcomings in other areas. One great first-hand experience may suffice to convince users of an app's general effectiveness.

However, we also observed this effect the other way around. As soon as users have negative experiences with core features, they lose confidence. In one case, the reviewer knew that an ex-partner spied on them, but the anti-stalkerware did not detect any malicious app: *"Never purchase this! My ex is still reads my messages - it's a disgrace" (R477)*. Similarly, this reviewer's trust vanished as soon as they realized they could not locate their stolen phone: *"The whole reason I have this app is in case I lose my phone." (R069)*.

While effective security apps must protect users in cases of attacks, a single thwarted attack is not a good indicator of a security app's effectiveness.

Reassuring user experience. Security apps' user experience influences the users' opinions about these apps. Frequent reminders of threats, updates, or scheduled scans keep users informed about the app's activity. Generally, attacks on users' security will be rare. So that these reminders of the ongoing protection effort can add a feeling of security for users: *"Get notified my phone is secure. That makes me feel better." (R165).*

Other users may see these reminders as a disruption of their regular phone use, e.g., *"the notification is permanently visible in the status bar. This is unsettling and annoying." (R202).*

For security use-cases, where apps might only rarely need to intervene, reassuring user experience is necessary to communicate that the app is still there and doing its job. However, reassuring user experience is independent of actual security. Hence, app developers may misuse this concept.

Building trust over time. Frequently, the history of app use influenced trust. Similar to human relationships, using the app over an extended period reassured users and increased their trust in the security app. We found three types of time references: establishing authority by stating experience, insufficient evidence of protection over time, and satisfaction with the absence of incidents.

In case of establishing authority, reviewers usually said they had used the app for years before telling us their verdict, e.g., *"Works as advertised have used it for years" (R173).* Some reviewers expected security apps to demonstrate their effectiveness. R476 assumed the app was a scam because they could not determine what it does: *"I cannot tell that this does anything for my phone so I think this is a rip off".* However, other reviewers were happy and felt safer when the security app did not find anything: *"Haven't found anything yet but thats a good thing!! Feeling alot more safe." (R475)*

These contradicting positions are interesting since they demonstrate two fundamental ways users think about apps' security. In the first one, users demand evidence of functionality, even if there is nothing wrong with their smartphone. The other approach assumes the security app's effectiveness without evidence. Even though both reviewers used the same app, they ended up with different trust assessments.

Testing app's abilities. Numerous users did not wait for incidents in their day-to-day life to establish trust. They decided to test the apps' abilities. They compared the abilities of different anti-stalkerware apps, e.g., *"This app missed two spyware apps that the others detected." (R470).* Some knew they had spyware installed and checked if a particular anti-stalkerware could remove it: *"Can't find the spyware that is obviously installed on my phone." (R512)* R291 reported using an EICAR test file to check if the security app would detect it: *"Garbage. Eicar test antivitus not detected" (R291).* In this case, the reviewer successfully tested the 'lost phone' feature: *"Locating/Alarm etc always worked when tested" (R344).*

In general, testing security features is a solid way to build trust. However, comprehensively testing apps' malware detection abilities is hard. Other security features, such as the 'lost phone' feature are easier to test than malware detection's effectiveness. Hence, reviewers could have a misleading impression of their app's abilities even after testing them.

Third-party recommendations. Reviews rarely referred to third-party resources to justify their trust in anti-stalkerware apps. In one case, a friend in IT security recommended an app: *"My friend who is in IT security suggested this app to me" (R131)* In another case, a reviewer referred to a study: *"saw a study that showed this had best spyware detection rate (but also false positives)" (R423).*

Users who got anti-stalkerware recommendations from third parties have delegated trust establishment. For them, the user experience of a security app is not as crucial as for other users – they are already confident in its security.

4.4.1. Observations

During our analysis, we also observed other noteworthy trends among the reviews: emotional language, assemblages of security tools, and cases of tracking family members.

We found that reviewers often used emotionally loaded language. Positive reviews, such as R145, described the protection app as a sort of guardian angel: *"It's a guardian keeping an eye on my stuff".* The name of one of the apps in our case study, i.e., Lookout, might explain why reviewers make this connection. Negative reviews often used strong language when talking about the apps' shortcomings. Such as R114, who complained about the app's malware detection ability: *"Pathetic virus support"*, or R014, who just wanted to remove the app altogether: *"take this Crappy off [my phone]".* However, since app-store reviews are voluntary, these observations could be due to self-selection bias, i.e., users who feel betrayed or well protected by the app submit more reviews.

Some reviewers did not evaluate the app independently from others. Instead, they considered how the app fits into their assemblage of security tools, e.g., *"Nice addition to any security set up." (R402)* or *"Lookout (Basic license) is good pair with Avast Mobile Security and CCleaner." (R098).* In such cases, users focus less on a specific tool's efficacy but rather on the feature set of the entire assemblage. However, some of these tools expect to be standalone tools, which may impact the resulting user experience.

One reviewer explicitly described their use-case for the app as tracking family members. *"We did not change anything but whenever I try locating my son there is an error." (R155)* We assume that parents such as these have only the best intentions for their children's safety. However, Gosh et al. [103] found that affected children perceive their parents' surveillance as overly restrictive and privacy-invasive. Our case also illustrates how users employ security apps to subvert their intended use-case.

4.4.2. Contradicting User Expectations

We found two approaches to trusting the apps in our case study: (1) trust, based on absent negative experiences with the app, and (2) no trust without proof that the app works as intended. Using the first approach increases trust in the security app the longer it runs without incidents. Users employing the second approach either wait until the app detects an issue or challenge the app to trigger an alert. R260 is exemplary for the first approach: *"I have had this app on all my devices over the years and no problems of any kind"* R215 is an example of the second approach: *"I've not had any positive hits from this yet, so it's difficult to say how good or bad the app is."*

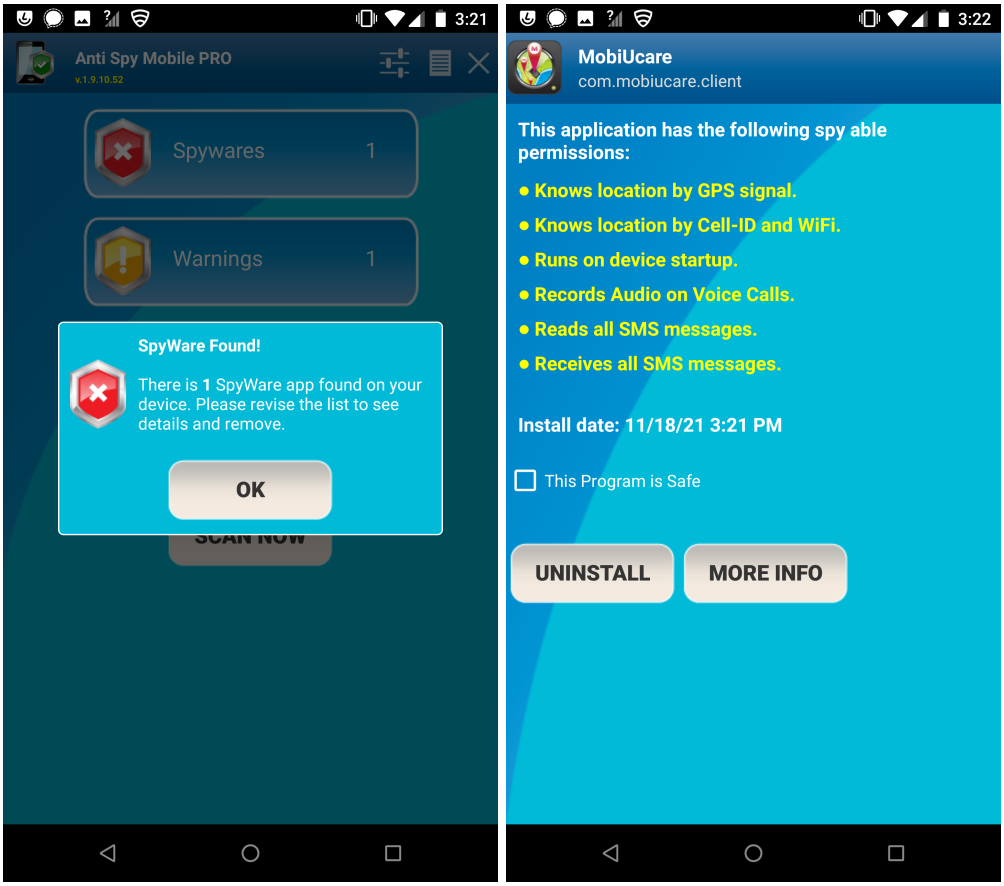


Figure 4.1.: Anti Spy Mobile PRO's response to a well-known spyware app.

The app's user interaction impacted users' trust in two contradictory ways: Some users thought the app was not doing anything when they could not observe any user interaction with it, i.e., they felt reassured by visible UI elements. Others interpreted the missing user interaction as a security indicator, expecting the app to respond only to security issues. R121 feels reassured when Lookout communicates that it is working: *"it lets me know they are working by updating me at various time intervals and pops up on your screen when you are not thinking about them" (R121)* R250 would feel more protected if Anti Spy Mobile were to indicate its ongoing operation: *"there should be an anti-spy guard for the icon on the home screen. That would enhance users to feel protected and safer" (R250)* In contrast, R065 is happy that the app stays silent and in the background: *"it silently keeps my phone in check from the behind the curtain" (R065)*

4.5. UI Walkthrough of Anti-Stalkerware

During our thematic analysis of the app-store reviews, we identified two approaches on how users establish trust with anti-stalkerware apps based on their user interface:

(1) Incidents with potential harm and experiencing how the app handles the situation builds users' trust; (2) Apart from potentially harmful incidents, users appreciate anti-stalkerware's reassuring security experience during everyday use.

This section reports the results of a cognitive walkthrough [294, 259] focused on these two trust establishment approaches. For the purpose of this walkthrough, we assumed that malicious parties may have had direct access to the phone before, but they no longer do at this point. When malicious parties still have direct access, removing electronic traces of anti-stalkerware usage afterwards is necessary to keep its users safe [73]. We simulated harmful incidents by installing several spyware apps on a smartphone that we reserved for this purpose. In the resulting user interactions, we document and inspect all the parts of the UI flow and answer guiding questions about the effect on users' trust. We simulated the day-to-day experience by using the smartphone with the installed case-study apps for 48 hours as our regular phone. We browse the web, download data, and install apps. We document and inspect user interaction and answer guiding questions about the effect on users' trust.

4.5.1. Potentially Harmful Incidents

Anti Spy Mobile PRO. Opening the app shows three different classifications of apps (as buttons): (1) *Spywares* for well-known blocklisted spyware apps; (2) *Warnings* for all suspicious apps not on the blocklist; (3) *All Applications* for all other apps.

Anti Spy Mobile automatically starts a scan when users open the app for the first time. Users may trigger a scan manually with the *Scan now* button or enable automatic daily scanning in the preferences (which is the default setting). After each scan, a dialog box presents the number of identified well-known spyware apps. If it did not find any, it presents the number of suspicious apps instead. Confirming the dialog box brings users to review the apps in question (as seen in Figure 4.1 and 4.2).

To test Anti Spy Mobile's reaction to a well-known spyware app, we installed MobiUcare (Phone Locator) on our test phone. Figure 4.1 shows the resulting *"SpyWare found"* dialog. After confirmation, Anti Spy Mobile shows the name, privacy-infringing permissions, and installations date of the detected spyware app. The *"More Info"* button would usually lead to the corresponding listing in the Google Play Store. However, this results in an error message since this app's removal from the store.

We installed two more spyware apps: mSpy Cellphone Tracker and SpyFone. The FTC banned the latter in September of 2021 [94]. Figure 4.2 shows that it does not consider them well-known spyware. Instead, it informs users about suspicious apps on their phones. The text describes the classification based on requested permissions and suggests how to deal with these apps: *"you should take a close look at them and uninstall them if you are not familiar with their existence"*.

Selecting suspicious apps reveals more detailed information about them (Figure 4.3), such as their name, suspicious permissions, and time of installation. This view offers users three responses. First, users may want more information about the app in question. However, the corresponding button leads to the Google Play Store website, which may not provide users with sufficient threat information. With MobiUcare, the button

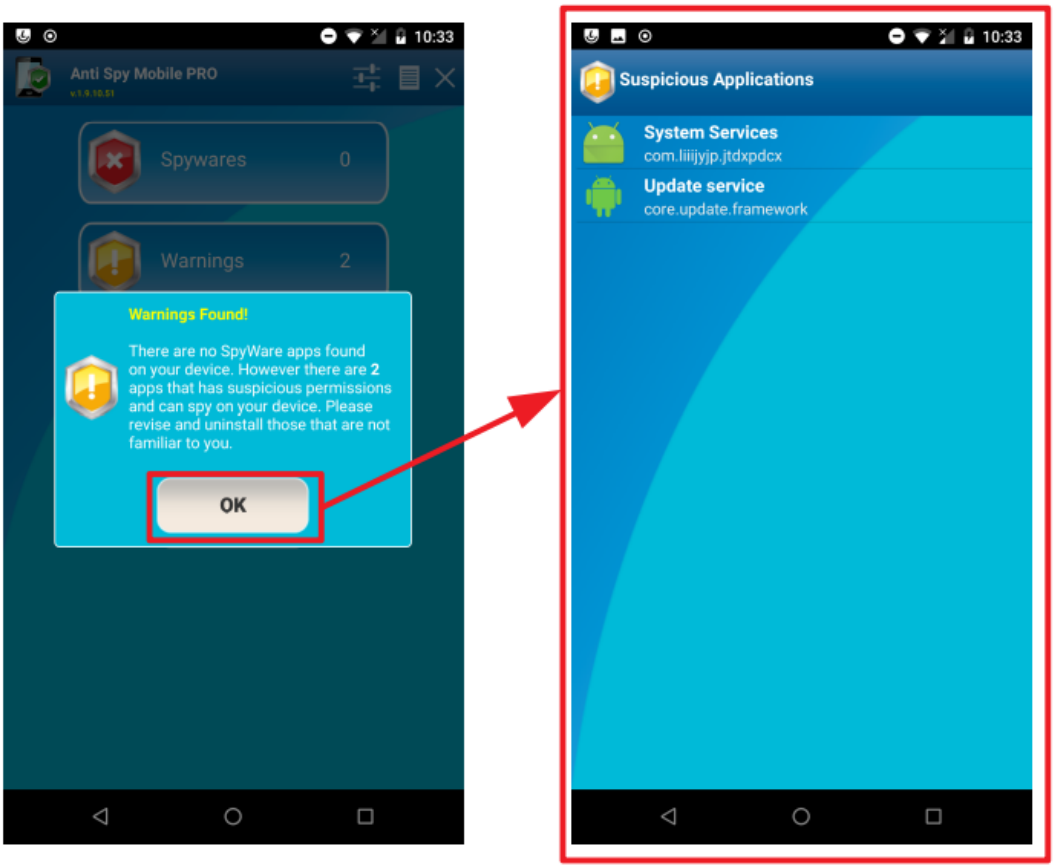


Figure 4.2.: Anti Spy Mobile PRO's response to spyware apps that are not on its list of well-known spyware.

generates an error since the app is no longer on the app store. Second, users can uninstall the app directly with a button click. However, if it concerns admin apps, this results in an error message: *"Uninstalling MobiUcare unsuccessful"*. The app does not provide any guidance in this case and acts as if the user never pressed the button in the first place. Third, if users do not want to take further action, they mark the app in question as *"safe"*. Then Anti Spy Mobile will stop notifying them about the app. Figure 4.3 shows that the threat response interface is independent of the identified threat. Anti Spy Mobile treats apps with merely suspicious permissions (the Signal messenger in this case) in the same way as apps on its list of well-known spyware.

Lookout Mobile Security. Lookout Mobile automatically scans all installed apps after installation. Users can start a scan manually at any time (see Figure 4.5).

To test Lookout's response to stalkerware, we installed MobiUcare, mSpy Cellphone Tracker, and SpyFone. Lookout Mobile correctly identified all three and classified them as *Surveillanceware*. In Figure 4.4 a pop-up window shows all identified apps with the option to either view details or set a reminder. The remind later option does not require users to specify a time that works better for them. Such commitment devices

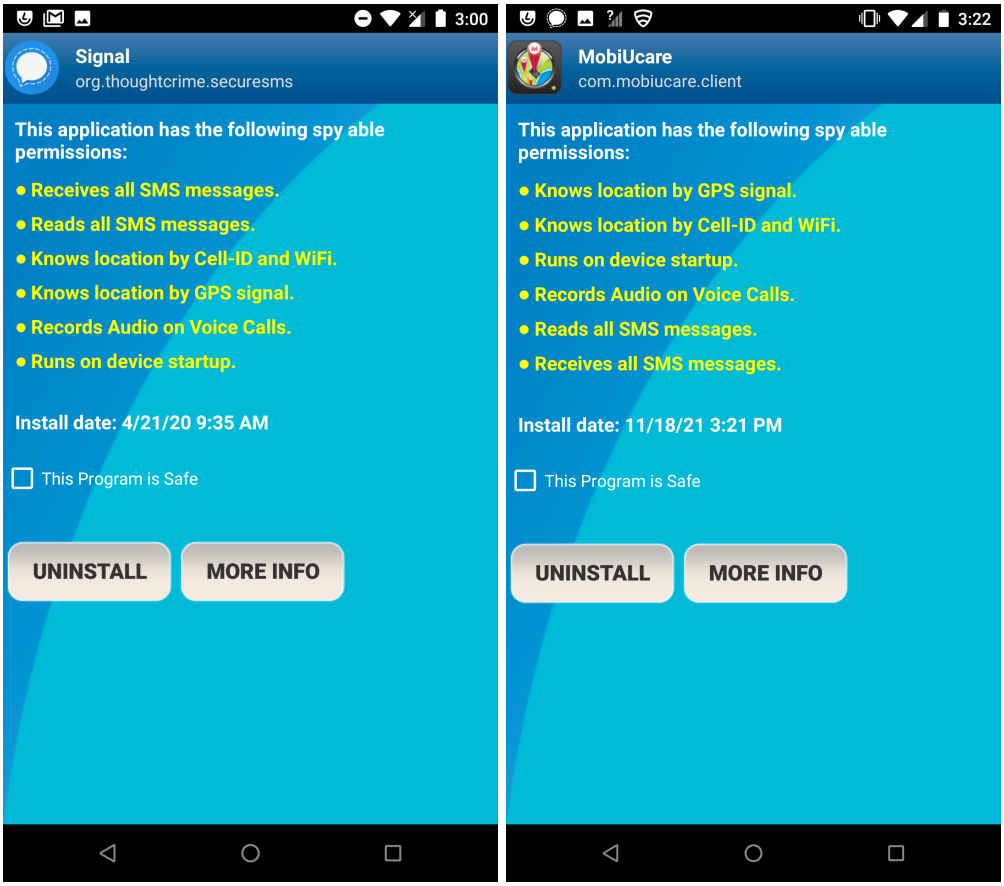


Figure 4.3.: Additional information provided by Anti Spy Mobile PRO on a suspicious app on the left and a well-known spyware app on the right – both apps request the same "spy able" permissions.

can increase security compliance [93].

In the detailed overview, Lookout shows a classification (e.g., Surveillanceware), logo, name, version, time of detection, and an app report for each identified threat. Reports comprise three parts: a statement if the app is a commercial surveillanceware (if applicable), a list of human-readable permissions, and a generic explanation about third parties monitoring user activity without consent. The only context-dependent information seems to be Lookout's analysis if the app in question is commercial surveillanceware.

Lookout affords users three responses for detected threats. First, users may click on App Info & options, leading them to the system's overview of the app in question. Second, a highlighted uninstall button. While Lookout does not explicitly suggest an appropriate response to the threat, the highlighted button strongly suggests uninstalling. Lastly, it offers the option to ignore threats. Lookout does not provide users an explicit discussion of these options, not even when it identifies commercial surveillanceware.

Additionally, users have access to the scan history (see Figure 4.5). Upon detection of surveillanceware, this view offers users to "learn more about surveillanceware", leading them to the built-in threat encyclopedia. The encyclopedia provides a general overview of

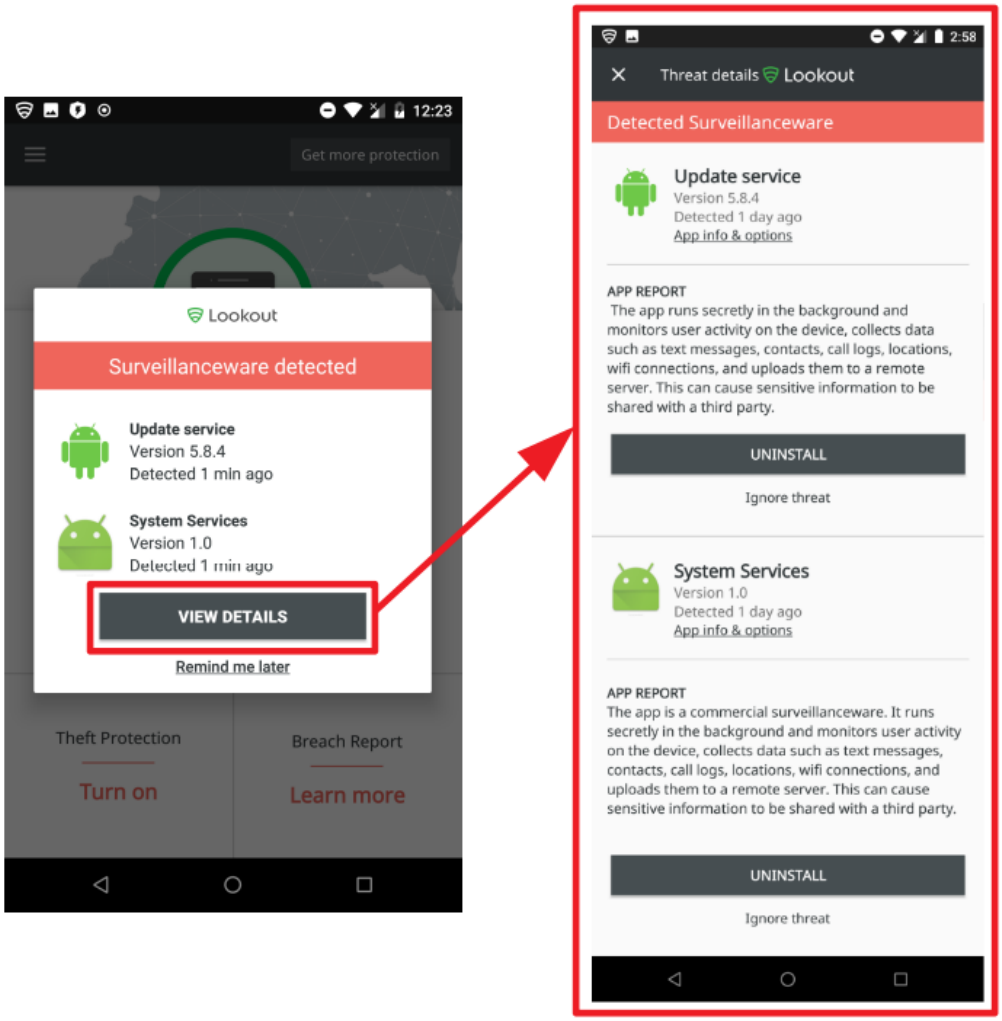


Figure 4.4.: Lookout Mobile Security's scan results identifying the spyware apps as surveillanceware.

surveillanceware abilities and only mentions a vague threat model, i.e., "Surveillanceware apps are typically installed directly by someone with physical access to the target device". The encyclopedia avoids discussing appropriate user responses.

4.5.2. Reassuring Everyday Experience

Anti Spy Mobile PRO. Apart from manual scans in the app itself, Anti Spy Mobile barely interacts with users. The paid version automatically scans all apps and notifies users about the results once per day (see Figure 4.6). This notification does not warn about suspicious apps. Anti Spy Mobile does not intervene during day-to-day activities, such as browsing the web, downloading files, or installing apps (from the Google Play Store or third-party repositories).

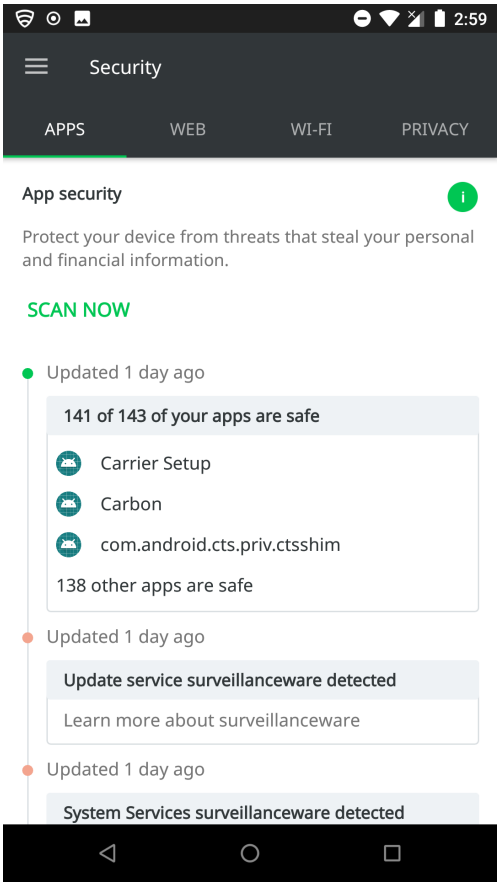


Figure 4.5.: Dashboard of Lookout Mobile Security with scan history and re-scan option.

Lookout Mobile Security. In general, Lookout Mobile focuses on reassuring user interaction. A sticky icon in the status bar and a permanent notification (shown in Figure 4.7) informs users that Lookout is active and that "everything is OK". Another aspect of Lookout's user interaction is its reactivity to the users' actions. It warns users about malicious files or apps immediately after downloading or installing them, respectively. Additionally, Lookout has a setting to notify users about a WiFi network's safety at connection time. Immediate responses improve users' mental models when the notification links causes and effects [260].

Enabling Lookout's VPN-based safe browsing feature did not affect the surfing experience. By default, Lookout analyzes downloaded files for threats (according to the description in the settings). Downloading regular files did not create a response from Lookout. However, it reacted when it detected spyware in a downloaded .apk file (Android Package, i.e., the Android app distribution format). Installing apps always created a response, regardless of the origin. Interestingly, Lookout considers the app Find My Kids safe (see Figure 4.7), while Anti Spy Mobile considers it well-known spyware.

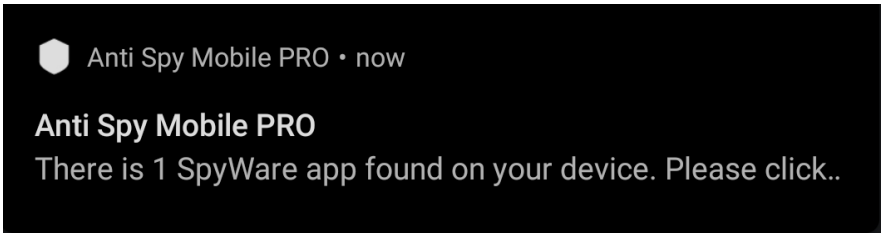


Figure 4.6.: Anti Spy Mobile PRO's daily scan notification.

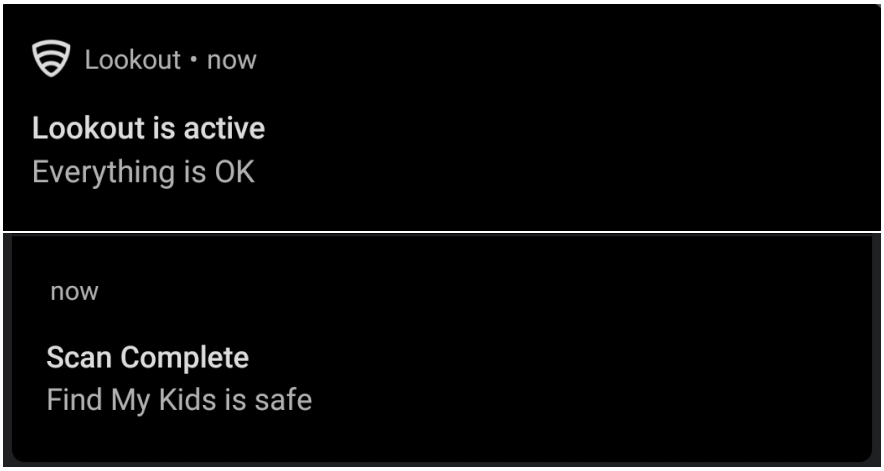


Figure 4.7.: Lookout Mobile Security's reassuring notifications.

4.6. Anti-Stalkerware under the Hood

Our thematic analysis identified two trust establishment approaches that users apply to anti-stalkerware. First, they build trust over time after seeing which threats the app caught and which it did not catch in time. Second, reviewers actively challenged the anti-stalkerware's abilities by installing known spyware on their phones. Both approaches are based on users' partially correct understanding of how to evaluate detection mechanisms.

To take a closer look at the detection mechanisms of our case-study apps and to understand how they determine which installed apps are threats, we performed static code analysis and dynamic run-time analysis. We follow established best practices (as outlined by OWASP [208]) for mobile app testing and rely on selected open-source tools. Android apps are typically written in Java, compiled to Dalvik bytecode, and then packaged as .apk files (essentially a zipped archive) [111]. A common first step is to transform this bytecode back into Java source code for easier comprehension. To do so, we use the Dalvik-to-Java decompiler `jadx` [147]. To monitor the run-time behavior of the case-study apps, we installed them on a Nexus 5 phone and instrumented them with Frida [91]. This tool allows reverse engineers to inject and execute JavaScript in the analyzed app. We use this feature to inspect the app's classes, methods, and data fields guided by the results of the static analysis. We further use the web proxy Fiddler [82] to intercept and inspect network traffic to the apps' backend server, if any.

Anti Spy Mobile PRO. We started by locating the main activity of the app, representing the UI shown to users when they first open an app. The class `AntiSpyActivity.java` represents this activity and loads the start screen defined in XML format (`/resources/res/layout/start.xml`). This screen contains the *Scan Now* button, which triggers the scanner activity (`ScannerService.java`). This activity implements the core functionality of Anti Spy Mobile PRO: it calls the Android PackageManager [112] to get the package names of all apps installed on the device and iterates over it.

The app distinguishes between two relevant types of installed apps: *SpyWare Applications* and *Suspicious Applications*. It identifies the first category by matching apps' package names against a list of well-known spyware apps. This blocklist of package names is embedded in the app as an XML file (*blackListPackagesDefs* in `resources/res/values/arrays.xml`). For the second category, Anti Spy Mobile PRO retrieves the apps' requested permissions to check for "spy able" permissions related to location, microphone, and SMS access. If the sum of these weighted permissions exceeds a certain threshold, it flags an app as suspicious.

The XML file that contains the blocklist also contains an allowlist of package names (*whiteListPackagesDefs*) of apps that presumably would trigger false positives based on their permissions. This list contains for example different browsers, but interestingly also security solutions such as Lookout Mobile Security. In its current version, the blocklist contains 494 entries, while the allowlist contains 146 entries, with 30 of these package names matching apps available on the Google Play Store, respectively.

We reverse-engineered the free version (Anti Spy Mobile Basic) and confirmed that the only difference is the option to schedule automatic background scans.

We further executed Anti Spy Mobile PRO to confirm our findings from the static code analysis and inspect its behavior during the actual scanning process. During this experiment, the app classified neither of the two spyware apps mSpy and SpyFone as SpyWare because its blocklist does not include them. However, it classified them as suspicious based on their permissions.

Lookout Mobile Security. This app is more complex than Anti Spy Mobile PRO, both in terms of code and UI. In this case, we started by looking for the *Scan Now* button in the dashboard UI (see Figure 4.5). This button triggers a SQL query for the already stored results of the previous scans. We then looked at the code populating this database, which is split across a number of different classes. We found that Lookout Mobile Security also collects information about each installed app from the Android PackageManager [112]. In addition, for apps classified as malicious, it also stores an assessment including the classification category, assessment ID, severity of the threat, and the response type.

The actual scanning mechanism is implemented both as a local and a cloud scan. In the case of a local scan, it checks for assessment in the `Policy.FLX`. This policy is distributed via over-the-air (OTA) updates, i.e., updates automatically pushed to the app without any active user interaction. For cloud scans, the app creates a request to `https://appintel.mobilethreat.net` with hashed information about the app under assessment.

Monitoring the network traffic of the app using Fiddler, we observed that during the first scan it received data from `https://ota.lookout.com`. We identified this as the

source of the OTA policies, but could not identify its format. Thus, using Frida, we injected JavaScript into the process to inspect the list of assessments read from this policy file. Most of the assessments seem to be in the form of signature-based detection methods, i.e., as a blocklist. Lookout detected both spyware apps (mSpy and SpyFone) as surveillanceware based on this blocklist.

Comparison of detection mechanisms. Both Anti Spy Mobile PRO and Lookout Mobile Security detect mSpy and SpyFone, the spyware apps. However, the first app merely classifies the two spyware apps as suspicious, while the second one accurately recognizes both as surveillanceware.

Anti Spy Mobile PRO mainly works with a block- and allowlist of package names. However, package names are weak identifiers of Android apps. The Google Play Store uses it to uniquely identify apps and recommends following Java package naming convention, i.e., to "use Internet domain ownership as the basis for package names (in reverse to avoid conflicts with other developers" [110]). Still, developers can choose arbitrary or conflicting package names for their apps, particularly when they are distributed via third-party repositories. Malware authors have been known to use the tactic of imitating package names of benign apps, or randomly generating package names to evade detection [167]. The package names of mSpy (`core.update.framework`) and SpyFone (`com.rzjzmlrm.vhqpmgzo`) seem to follow this pattern. Technically, stalkerware distributors could even automatically generate new package names for each customer.

Furthermore, these lists are part of the resources embedded in the .apk file, and the app does not implement any functionality to update this file. Thus, any changes in the blocklist need to be pushed as part of app updates through the Google Play Store—which users may or may not install [198, 269]. The update history indeed includes *[UPDATE] Spyware definitions update*, but updates have been sparse since 2018 [13].

In addition to the detection based on the package name, Anti Spy Mobile PRO also flags apps as suspicious if they request permissions that could be used for spying. Nevertheless, Anti Spy Mobile PRO does not provide more information about these apps than the requested permissions to the users and does not describe or explain what these apps do.

Lookout Mobile Security, on the other hand, dynamically fetches signature-based blocklists from the server and checks for newer versions during each launch. However, in this case, the scan is a "black box": we have no insights about the type of scans performed on Lookout's servers and the features they base their detection on.

4.7. Discussion

We compare our thematic analysis results, i.e., users' strategies for establishing trust in their installed anti-stalkerware, with our user interface walkthrough and reverse engineering results – highlighting the expectation-ability gap. Then we discuss different stakeholders' options to reduce this gap and improve users' anti-stalkerware decisions in the future.

4.7.1. Contrasting Users' Expectations with Actual Protection Capabilities

Potentially harmful incidents. One of the ways reviewers decided to trust anti-stalkerware apps depends on their incident response. This approach relies on apps' ability to detect incidents. Users' trust depends on the information and user agency that apps provide. Our walkthrough revealed that Anti Spy Mobile PRO's suspicious apps produced easily identifiable false positives – potentially decreasing users' trust. Also, we found inconsistent results: Anti Spy Mobile considered *Find my Kids* well-known spyware, while Lookout Mobile considered it safe. This mismatch highlights the need for context-sensitive classification, especially for dual-use apps. Neither app did a great job informing users about specific threats and providing context-appropriate user agency options. For example, Anti Spy Mobile PRO offers the same information and response options, whether it concerns well-known spyware or merely suspicious apps. Reverse engineering the apps showed that Anti Spy Mobile PRO uses a package name list of well-known spyware apps and a list of well-known benign apps. Updating these lists requires an app updating the app. Lookout Mobile checks apps against local OTA policies, regularly updated from Lookout's servers. Anti Spy Mobile PRO further uses a permission-based approach to identify suspicious apps not on the list of well-known apps, resulting in easily identifiable false positives. Hence, relying on potentially harmful incidents as a strategy to establish trust with anti-stalkerware apps comes with risks. It relies on users' ability to recognize harmful incidents to understand if the app should have detected and prevented them. Waiting for such moments is risky. Ideally, users trust their anti-stalkerware app before they face attacks. Lastly, awarding trust in this way may deceive users. One instance where the app protected them may lead users to overgeneralize the assumed protection.

Reassuring user experience. The analyzed reviews contained praise for reassuring user interaction in benign everyday scenarios. In addition to the regular alerts in case of threats, Lookout Mobile incorporates user interface elements that communicate the current positive security status, e.g., "everything is OK". Showing users the security mechanisms during threats as well as in benign situations helps build users' mental models [260]. Distler et al.'s study [63] suggests that visualizing security mechanisms improves user experience. Notably, in our case study, Lookout Mobile always seemed confident in its safety assessments. In contrast, Anti Spy Mobile depends on permissions-based classification — leading to false positives. In addition, Lookout Mobile was very reactive, immediately notifying users about their actions' safety consequences. The timing of privacy and security notices may affect users' decisions in general [6]. Observing links between cause and effect forms users' mental models, making this immediacy between action and response beneficial [260]. However, moderately delayed privacy feedback may be a compromise to minimize interruption [215]. Reassuring user experiences have benefits in benign situations. They improve users' mental models and appear to improve user experience overall. The immediate response to potential threats may improve users' mental models by linking cause and effect. The certainty of anti-stalkerware's verdicts, warranted or not, may heighten users' trust. Ultimately, reassuring user experiences do not make apps more secure. Hence, users who rely on this trust establishment approach are prone to deception.

Assumptions about apps' detection capabilities. Reviews contained two approaches based on assumptions of the anti-stalkerware's detection abilities. First, reviewers evaluated the app's abilities over time, building trust similar to a personal relationship. Second, reviewers explicitly tested and challenged the app's detection ability with selected spyware or test viruses. Both approaches are flawed. Using the first approach, users assume they can detect a threat when the app can not. Since they may not recognize when the app fails to detect threats, they may only be aware of incidents where the app protects them. Using the second approach, users generalize their test results from a single test to the apps' abilities to detect other malicious software, which might seriously mislead users. Even worse, since they tested the apps' ability personally, they put significant trust in their assessment.

Reliance on third-party evaluations. Some reviewers exclusively relied on third-party evaluations of anti-stalkerware apps. Depending on the third party may be the safest choice to establish trust. However, it also comes with drawbacks. First and foremost, trust in the third party is required — moving the issue of trust establishment from the app to the third party. Then, the third party has to have reviewed the users' chosen app. The effectiveness of this approach relies on reputable third parties. Ideally, trusted third parties are well-known for providing fair assessments. However, social effects may impact the choice of trusted third parties. Users rely on tech-savvy family members and friends even when they can not provide fair assessments. In any case, users can not influence and may not even know which aspects third parties consider for their reviews (e.g., usability, user agency, detection rate).

Relying on third-party reviews, users do not experience how the app reacts in case of an incident, which may affect their comfort, comprehension, and ultimately their safety.

4.7.2. Implications and Future Work

The thematic analysis results suggest that judging anti-stalkerware apps' efficacy is hard for users. In the current circumstances, their safest option is to rely on IPV-specific evaluation results of certified antivirus testing labs. In the future, we should try to support and improve users' existing evaluation approaches and give them more agency to safely build trust in anti-stalkerware apps. However, adapting apps and operating systems to make intimate partner surveillance difficult and less surreptitious would likely limit the proliferation of stalkerware and other abuse-enabling apps more effectively.

Reassuring experiences are useful (if done correctly) but cannot be trusted. One of the themes in our thematic analysis was that users felt reassured and well protected based on UI elements. The UI walkthrough confirmed that one of the apps relied on positive messaging to communicate to users about its work. Mathiasen et al. [179, 180] refer to this as *secure experiences*, which are not necessarily the same as security. According to them, users will base their security decisions on previous secure experiences. Spero et al. [260] argue that user interfaces that hide security mechanisms hinder users from building detailed mental models of security. Hence, security mechanisms should present users with model-building information, whether they face security risks or not. As an

example, Distler et al. [63] found that visualizing security mechanisms in an e-voting apps led to an increase in perceived security. While these kinds of reassuring and secure experiences may be understudied, they appear to provide several benefits: (1) they communicate to users that a security system is working, even when no security risk calls for action; (2) they may improve users mental model of security; and (3) they help improve users' security decisions later on. However, these kinds of secure experiences become a problem if they oversell the actual security, regardless of the intention. Therefore, simple reassurances that everything is safe may not be the best approach to building secure experiences. The anti-stalkerware apps in our case study probably use reassuring experiences to justify their existence to users. Without them, it may appear like anti-stalkerware apps do nothing of value, even when they work well. In summary, reassuring user experiences may improve users' mental models and security decisions, but users cannot rely on them alone to establish trust in security mechanisms.

Demonstrate stalkerware detection to users. In our thematic analysis, we found reviewers used several different (flawed) tactics to evaluate the detection efficacy of anti-stalkerware apps. Also, we found that the anti-stalkerware's response to stalkerware (user experience, information, and agency) affects users' trust. Hence, it would make sense to encourage and improve this kind of evaluation behavior. We suggest offering a toolkit for users to install on their phones. This toolkit should be able to install (and remove) a wide variety of stalkerware and dual-use software and track the anti-stalkerware's response. Such a toolkit would affect users in three ways: (1) all users would have the ability to safely and soundly evaluate their chosen tool's detection mechanism, (2) users could safely experience their tools response to malicious software, and (3) it would reduce the need to trust third-party reviews of anti-stalkerware apps. Similar to this approach, Parson et al. [214] suggest that a government body should track and evaluate anti-virus engines and publish public reviews. However, in contrast to our suggestion, users would then not experience their chosen app's response to threats.

Provide context-specific advice and give users agency. Detection ability is an important but not the only factor for users' safety. The type and amount of information apps present to users influence their response. Additionally, users' agency to respond to detected threats is crucial. Both information and agency need to be context-sensitive to the users' circumstances and the specific detected threats. For example, for IPS survivors safe responses to detected surveillance threats may be different before and after they have left their partner. This could include additional context-specific response options, e.g., generating fake location data or partially removing permissions without alerting the stalker. Without context-sensitive advice and user options, even an anti-stalkerware app with great detection ability may endanger users.

Leverage operating system's power to limit abuse. Improving anti-stalkerware apps and users' protection abilities is an individualistic approach to combating IPS. However, a systemic approach may be more effective in reducing IPS. Considering potential abuse in the design stage for operating systems, apps, and accessories may help fight IPS on a system level. Defensive design is a widely adopted approach across many disciplines.

However, it focuses on unintentional errors in programming code and resulting apps. Other general approaches take intentional abuse into account at every step of the design process to mitigate interpersonal harm [296, 216]. Levy and Schneier [164] offered design considerations to ameliorate intimate privacy risks. Slupska and Tanczer [257] suggested an approach to threat model intimate partner violence in the design process. Interestingly, the two most common smartphone platforms, iOS and Android, are not equally susceptible to stalkerware targeted at consumer audiences [128, 214]. Parsons et al. report on the stalkerware industry [214] and the limited options to install these stalkerware apps on iOS without jailbreaking. Consequently, most commercial stalkerware for iOS devices rely on the target's iCloud account. Reputable companies do not want to publicly support dedicated stalkerware, so these apps are not published in app stores—or are quickly removed. This may result in a proliferation of other abuse-enabling dual-use apps (such as parental control apps) and their legitimate use-cases make them harder to police. Since legitimate use-cases are here to stay, it is necessary to adapt the design of these apps and the operating systems to limit misuse. The authors report recommendations applicable to platform providers that may curb stalkerware. They call for prominent, ongoing, and meaningful consent notices. These make it harder to install stalkerware surreptitiously on others' smartphones. Additionally, they call for on-device platform heuristics that detect misuse of ostensible dual-use software. Platforms have the power to disable abuse-enabling apps entirely – which may protect users unable to manage apps on their device. Platform providers have significant power over the kind of software they allow to run and which kind of app activities they make visible to users. Using this power would be an effective measure against the current stalkerware ecosystem.

4.8. Conclusion

Choosing effective anti-stalkerware solutions is a struggle. This case study evaluated two anti-stalkerware apps from multiple perspectives to understand users' selection and trust strategies. We identified five approaches that users apply: two based on user interaction, two based on the assumed detection abilities, and one on trusted third parties. All approaches are intuitive to apply and have some degree of legitimacy. However, the cognitive walkthroughs and reverse engineering approaches revealed severe drawbacks. We found that users' strategies do not inform them sufficiently about these apps and their abilities to mitigate violence, abuse, and harassment.

Our work helps improve current anti-stalkerware by suggesting design directions that increase users' trust and safety. These design directions focus on reassuring user experience, context-sensitive advice, and risk-appropriate user agency. Also, we suggest a user-deployable, toolkit-supported approach to evaluate anti-stalkerware's detection abilities and user experience. Such a toolkit-based approach builds on and encourages existing user behavior while improving its efficacy and safety. Lastly, while our study focuses on individualistic responses to anti-stalkerware, we emphasize the need for a systemic, platform-level approach to effectively combat intimate partner surveillance.

Acknowledgements

We thank the reviewers for their feedback on improving our paper. In particular, we thank our anonymous shepherd for their responsive, helpful, and kind guidance. The first author conducted their work as part of the Saarbrücken Graduate School of Computer Science, Saarland University.

This research has received funding from the Vienna Science and Technology Fund (WWTF) through project ICT19-056, as well as SBA Research. SBA Research (SBA-K1) is a COMET Centre within the framework of COMET – Competence Centers for Excellent Technologies Programme and funded by BMK, BMDW, and the federal state of Vienna. The COMET Programme is managed by FFG.

The contents of the following chapter were published as part of the publication *"Investigating Security Folklore: A Case Study on the Tor over VPN Phenomenon"* (CSCW 2023) [P3]. This paper was produced in cooperation with my co-authors Alexander Ponticello, Adrian Dabrowski, and Katharina Krombholz. The following table describes the contributions of all authors to this paper:

Author	**Contribution**
Matthias Fassl	I had the idea to investigate Tor over VPN behavior, and I chose the three types of studies and the corresponding methods. I was the main person responsible for designing, conducting, and analyzing the results of all parts of the three studies. I conducted the parts of the studies that involved qualitative analysis together with my co-author Alexander. The initial draft version of the paper came almost entirely from me. Alexander initially drafted selected sections of the qualitative analysis parts.
Alexander Ponticello	Alexander contributed to the qualitative analysis of survey responses and the choice, collection, and analysis of the information sources. Additionally, he contributed by writing up parts of the qualitative results and retrieving appropriate quotes from our corpus to illustrate concepts.
Adrian Dabrowski	Adrian contributed the framing of this case study as an example of security folklore and helped me write the introduction. Additionally, he gave me helpful feedback on the entire paper for me to edit.
Katharina Krombholz	As my academic advisor, Katharina was involved in the major decisions during this research project. She gave me feedback on the initial idea and the choice of studies, including the methodological approach. She guided me through methodological and ethical questions, reviewed my interpretation of (intermediate) results, and helped me find an appropriate venue for this research project. She also gave feedback on draft versions and edited parts of the paper.

Reference

Matthias Fassl, Alexander Ponticello, Adrian Dabrowski, and Katharina Krombholz. 2023. Investigating Security Folklore: A Case Study on the Tor over VPN Phenomenon. Proc. ACM Hum.-Comput. Interact. 7, CSCW2, Article 344 (October 2023), 26 pages. `https://doi.org/10.1145/3610193`

5

The Curious Case of Tor over VPN

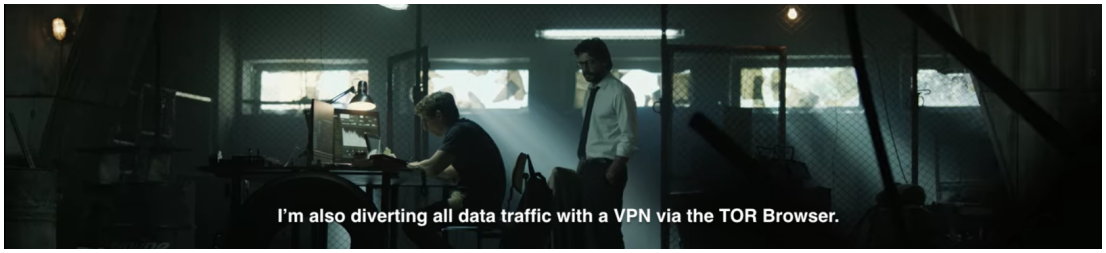


Figure 5.1.: Rio explains to the professor that he uses Tor over VPN to access the dark net (Money Heist S02E03, 2017 [220])

Users face security folklore in their daily lives in the form of security advice, myths, and word-of-mouth stories. Using a VPN to access the Tor network, i.e., Tor over VPN, is an interesting example of security folklore because of its inconclusive security benefits and its occurrence in pop-culture media.

Following the Theory of Reasoned Action, we investigated the phenomenon with three studies: (1) we quantified the behavior on real-world Tor traffic and measured a prevalence of 6.23%; (2) we surveyed users' intentions and beliefs, discovering that they try to protect themselves from the Tor network or increase their general security; and (3) we analyzed online information sources, suggesting that perceived norms and ease-of-use play a significant role while behavioral beliefs about the purpose and effect are less crucial in spreading security folklore. We discuss how to communicate security advice effectively and combat security misinformation and misconceptions.

5.1. Introduction

An increasing number of laypeople are concerned about their privacy and security online and want to minimize their risks. They are met with a conundrum of possible products, advice, myths, and word-of-mouth stories. For example, some users append symbols to their passwords assuming increased security [278], believe that (unencrypted)

SMS are more secure than email and encrypted instant messengers [3], assume that information on the blockchain is not readable by others despite it being a public ledger [173], or overestimate the privacy gains of private browsing features [124, 98, 2]. In these situations, users' folk theories [289, 225] about the purpose and effect of security behaviors do not reflect reality. People perpetuate *security folklore*, a collection of beliefs about security practices (discussed in more detail in Section 5.2.1) when they tell informal stories about their security experience to others [227]. At best, these behaviors do not impact privacy and security online. At worst, they give a false sense of security while negatively impacting security.

Currently, little is known about how security folklore becomes popular, spreads, and gets widely adopted. Understanding these mechanisms could facilitate the transport of accurate and up-to-date information about security practices. Of particular interest are factors that make them so easy to retell to others, intuitively understandable, feel so secure, and actionable in users' day-to-day life.

In this paper, we approach the broad topic of security folklore based on the peculiar case of *Tor over VPN*, where the user first connects to a VPN before connecting to Tor. Experts [273, 272] and special interest groups [236] doubt the general security benefits of this practice. The case of *Tor over VPN* is particularly interesting because its utility (or dangers) are actually not fully conclusive, it vividly illustrates the power of misconceptions, and it found entrance into popular culture (e.g., Figure 5.1). An in-depth understanding of the phenomenon, its mechanisms, and the forces at work may help us to start understanding the building blocks of security folklore.

We triangulate this phenomenon with three study parts that align with the *theory of reasoned action* [85]: (1) we measure the prevalence of the Tor over VPN practice on real-world Tor traffic, (2) we survey Tor over VPN users' beliefs about the practice using mixed qualitative/quantitative methods, and (3) we systematically analyze the online content of background information on Tor over VPN.

We found that Tor over VPN users, comprising 6.23% of daily Tor users, rarely articulated threat models and expected general security benefits. Users who gave reasons often wanted to protect themselves from the perceived dangers of using Tor. This attitude reflects how VPN providers tend to write about the practice, emphasizing and sometimes misrepresenting the risks of using Tor to recommend their VPN for accessing the dark net. Our document analysis suggests that background information establishes normative beliefs when users look to others for appropriate security behavior (i.e., social proof [44, 50, 51]), contributing to the spread of this kind of security folklore. The importance of normative beliefs, which do not rely on knowledge about the practice's purpose and effects, explains why many Tor over VPN users expect general security benefits. Establishing reliable behavioral beliefs about the effects of Tor over VPN is difficult since the information sources, mostly VPN providers and discussions in expert venues, often contradict each other. This work highlights that newspaper reports about security practices or demonstrated security behavior are specific forms of security advice – since they offer social proof for security behavior. Knowing the relevance of normative beliefs can improve the future dissemination of high-quality security advice. Reports about security practices must avoid unintentional misinformation and misconceptions. Always including the security practices' purpose and effects may help.

5.2. Background and Related Work

We discuss the basics of security folklore, i.e., folk theories, security misconceptions, and security advice, before covering prior work about users' perceptions and use of VPNs and Tor. We include short descriptions of different Tor and VPN combinations and experts' opinions on the security benefits.

5.2.1. Security Folklore and Misconceptions

According to Brunvand's definition from 1978, folklore is a part of a culture that *"encompasses all knowledge, understandings, values, attitudes, assumptions, feelings, and beliefs transmitted in traditional forms by word of mouth or by customary examples"* [36]. When we transfer this notion to security and modern communication, security folklore is a collection of beliefs about security practices that have become part of the culture and are transmitted informally through, for example, verbal tales, written social media posts, or demonstrated behavior. This informal and repetitive communication may also enhance the perceived legitimacy of these security beliefs, regardless of available empirical evidence. In contrast to the repeated informal stories about similar security incidents that people tell each other, official, e.g., password-creation guidelines would not constitute security folklore.

Thus, there are two approaches to studying security folklore: investigating the informal communication about security practices and understanding users' collective knowledge, assumptions, feelings, and beliefs about security practices. In the spirit of the first approach, Rader et al. [227] and Pfeffer et al. [218] investigated the stories that people tell each other about security and what kind of lessons they learned from them. They found that people remembered these stories for months or years after initially hearing them and that almost half of them retold them to others. Studying users' functional or structural mental models [151, 57] of security technology would be a way to use the second approach to researching security folklore. For example, Wash [289] found folk models of home computer security that influence which security software people use and which expert security advice they follow. Rader and Slaker [225] found that these folk theories about technology depend on how the technology represents its actions to the users. Similarly, technology may provide a secure experience [179] not based on actual security. In this work, we investigate a specific case of security folklore in both ways, i.e., investigating users' beliefs and assumptions about Tor over VPN and studying different kinds of informal communication about it.

5.2.2. The VPN Ecosystem and its Users

Commercial VPN providers are a 15-billion dollar industry that markets their services as a turnkey solution to users' privacy [155]. The industry invests heavily in marketing. Akgul et al. [10] investigated influencer marketing ads on YouTube for VPNs. They found misleading claims, including overpromises and exaggerations, that could negatively influence users' mental model of internet safety. Ads like these may increase users' trust in the services' offered privacy and lead to their adoption [127]. Ramesh et al. [231] found that security and privacy are people's main reasons for adopting VPNs, while around

40% of their participants have flawed mental models of them. They attribute aggressive and misleading VPN ad campaigns to the degradation of users' mental models.

Binkhorst et al. [27] found that non-experts and experts alike were unsure when to use a VPN. They partially explained this with both groups' unclear threat models. Users adopt and abandon VPNs for emotional and practical reasons [201], whereas university students adopt them primarily for practical reasons [67].

5.2.3. Combined Tor and VPN Use: Potential Benefits and Harms

In general, non-expert Tor users employ abstractions that hide essential operational aspects of Tor [97]. These abstractions may lead to behavior that compromises non-experts' anonymity.

For example, Biryukov and Pustogarov [28] discussed the anonymity effects of combining Tor with Bitcoin. Some developers endorsed Tor for Bitcoin to avoid IP address leaks. However, this only provides limited anonymity and introduces an additional attack surface for MitM attacks. Story et al. [264] studied the adoption of security tools and found that users with VPN and Tor browser experience were often confused about their protection. Several of their participants stated that the Tor browser must be used with a VPN for added security. There are several approaches to combining Tor with a VPN. The two common ways are Tor over VPN, i.e., connecting to a VPN before opening the Tor browser, and VPN providers' Tor mode, where VPN servers relay the users' traffic into the Tor network so that users do not need to install the Tor browser. Other niche approaches, such as using a VPN over Tor or elaborate combinations of different tunneling technologies, exist but are not the focus of this work.

While experts doubt the general security benefits of Tor over VPN, they agree on two specific cases where it is helpful: Circumventing censorship when access to the Tor network is blocked while access to VPNs is not, and, to some extent, hiding Tor traffic from ISPs [273, 272, 236]. However, Tor bridges may provide similar benefits in both cases [273].

Since VPN providers are Tor over VPN users' first hop forever, the same concerns as with potentially malicious Tor guard nodes apply: they are in the position for selective denial of service attacks or statistical profiling attacks [59]. Either could aid attackers in deanonymizing Tor users linking them to their online behavior. Hence, Tor over VPN users must trust their choice of VPN provider completely [58, 273]. However, as prior work [145, 230, 293] has noted, statements by VPN providers are not necessarily trustworthy and using VPN may do more harm than good.

Measuring traffic in the Tor network in a privacy-preserving manner has been an active research topic in the last few years, resulting in PrivCount [149] and PSC [79, 174]. Our measurement study builds on PrivCount to quantify VPN use in the Tor network.

5.3. Methodology

In this case study, we investigate the phenomenon of Tor over VPN practice to learn more about the mechanisms of security folklore. Hence, we are not only interested in the

efficacy of Tor over VPN (object) and the users who practice it (subject) but also in the *"terms, conditions, and situation of the interaction"* [121] between them. Initially, we want to know if Tor over VPN practices are widespread enough to consider them part of the culture. Then, we investigate this part of security folklore in two ways: first, by studying users' assumptions and beliefs about the practice, and second by investigating informal communication about the practice. Hence, we focus on the following three research questions:

RQ1: How widespread is the behavior of using Tor over VPN amongst Tor users?

RQ2: What are the users' reasons for employing Tor over VPN?

RQ3: Which communication mechanisms contribute to diffusing Tor over VPN information and behavior?

To guide our investigation of this security practice, we align our case study with the theory of reasoned action [85]. The theory of reasoned action postulates that human behavior, while not necessarily rational, follows basic patterns. The authors proposed this model to understand the determinants of behavior and, in the long run, design behavioral interventions. According to the theory, behavioral intention and actual control are strong indicators that individuals will carry out a behavior. These intentions are based on three belief types: behavioral, normative, and control.

Behavioral beliefs are about the expected positive or negative consequences of performing a certain behavior [85, p. 100]. The intention to perform the behavior is primarily influenced by possible outcomes that individuals perceive as probable and vital.

Normative beliefs, in general, are the foundation for "perceived social pressure to perform a given behavior" [85, p. 130]. Fishbein and Icek [85, p. 131] differentiate two types of norms: injunctive norms, which concern other individuals' or groups' moral judgment on behavior, and descriptive norms, which concern the perception of other individuals' or groups' behavior. Related Usable Security research [50, 51] also uses the term *social proof* for these descriptive norms to describe how people look to others to learn the appropriate behavior.

Control beliefs form the "perception that one has or does not have the ability to carry out the behavior (i.e., perceived behavioral control)" [85, p. 170]. These beliefs are based on several control factors, such as required resources, available opportunities, anticipated obstacles, and self-efficacy.

Combined, these three beliefs influence peoples' intention to perform the behavior in question. The last belief about self-efficacy moderates the intention to engage, i.e., people only plan to implement a behavior when they are confident that they have the necessary abilities. These beliefs may originate from diverse sources, e.g., formal education, prior experiences, newspapers, TV, other Internet media, or interactions with friends and family. Other background factors like personality and demographics can influence how people interpret and remember information about the considered behavior. In contrast to other behavioral theories, the theory of reasoned action also incorporates background factors that inform these behavioral, normative, and control beliefs. Since this theory provides a unified framework for social behavior, it helps us guide our investigation and find the determinants of the Tor over VPN practice.

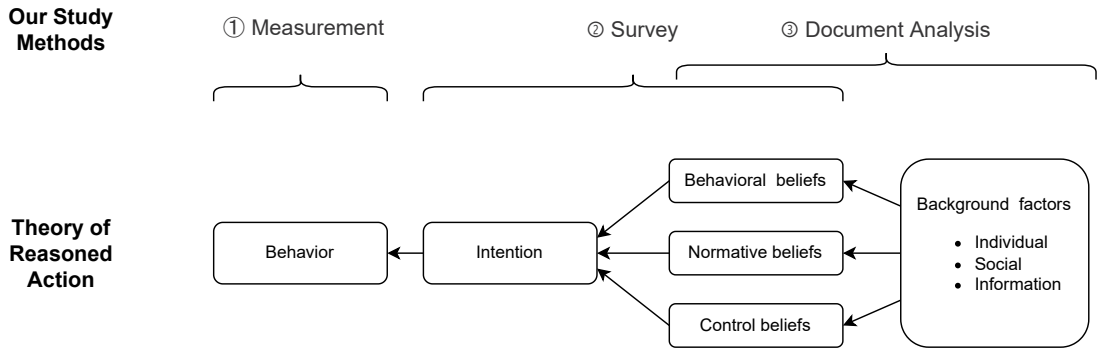


Figure 5.2.: Our study methods' alignment with the theory of reasoned action

Hence, we investigate the Tor over VPN phenomenon in three separate studies, each focusing on a different aspect of the theory of reasoned action (as shown in Figure 5.2):

1. we use a privacy-preserving method to *measure Tor over VPN behavior* in the Tor network to understand how widespread this practice is (see Section 5.4 and RQ1),
2. we use Amazon MTurk to *survey Tor over VPN users' beliefs* about the practice to understand more about why they choose to implement this practice (see Section 5.5and RQ2), and
3. we apply *document analysis to online information sources* covering Tor over VPN to understand on which information users may base their beliefs (see Section 5.6 and RQ3).

Since we also consider discussions on social media in our document analysis, we inadvertently also analyze user beliefs. Figure 5.2 illustrates this overlap between these two study parts with brackets. The upcoming sections describe these three studies in detail, including the methods, ethical considerations, and results.

5.4. Measurement of Tor over VPN Behavior

To study the prevalence of VPN use in the Tor network, we set up four guard nodes to measure the incoming traffic from detected VPN endpoints. Using this approach, we measure how human behavior impacts the entire Tor network in real-world conditions. However, we cannot directly conclude the number of VPN users in the Tor network from our measurement results since users might share the same VPN endpoints.

In the remainder of this section, we describe the measurement setup, the VPN decision procedure, counting Tor traffic data in a privacy-preserving manner, and analyzing the results.

5.4.1. Measurement Setup

We set up four Tor relays in a German data center. We started our measurement after the relays received their guard flag (showing that the relays had been around at least

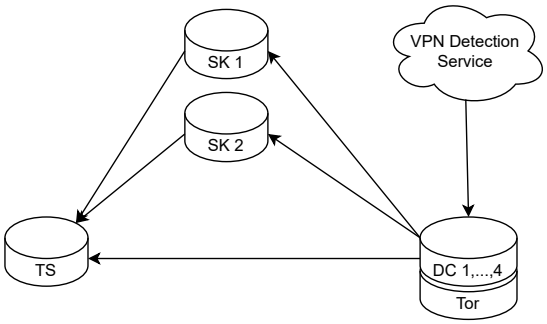


Figure 5.3.: Measurement setup containing four data collectors (DC 1-4), two share keepers (SK 1-2), and one tally server (TS). Each data collector receives events from a Tor guard node and retrieves information from a VPN detection service.

eight days[1]) and reached their bandwidth capacity. At this point, they advertised a combined total bandwidth of 1.46Gb/s. The probability of choosing one of our guard nodes was 0.6%, which made them the ninth-largest group of guards in the Tor network at the time, according to OrNetStats [205].

We use PrivCount [149], a system for safely measuring the Tor network, to count incoming connections' parameters. As Figure 5.3 shows, our PrivCount setup includes four data collectors (one for each guard node), two share keepers, and one tally server. Guard nodes ran a PrivCount-patched version of Tor, which notified the corresponding data collector of all connection events. Since the PrivCount patch was only available for an outdated version of Tor, we ported it to a supported version. Then, we adapted the data collectors to count the incoming connections from VPN providers.

5.4.2. VPN Decision Procedure

Commercial VPN services advertise their product as a way to access blocked content [155]. In response, content providers are interested in detecting VPN or proxy connections, which creates a market for so-called fraud protection services. Our initial market research uncovered the following services: proxycheck.io, IPtrooper, IPQualityScore, getipintel.net, IPHub.info, IPHunter, vpnblocker, IP2Location, Shodan, IPWarner, and MaxMind. All of them offer an online API service. Three of them, IPQualityScore, IP2Location, and Maxmind, also provide a privacy-preserving alternative: downloading a regularly updated database. We pre-filtered these services using basic quality criteria:

(1) *Specific responses*: a service's response must indicate if the IP address belongs to a VPN.

(2) *Rate limit:* services need to either support a minimum of 20,000 queries per day (i.e., the expected daily amount of connections per guard node) or provide an option to download a database for offline queries.

To select our VPN decision procedure, we evaluated the accuracy of the remaining services that fulfill these two criteria. We collected a ground-truth dataset consisting

[1] `https://blog.torproject.org/lifecycle-new-relay`

of known VPN and non-VPN addresses. For the set of known VPN IP addresses, we created accounts at four popular VPN providers (ExpressVPN, HideMyAss, NordVPN, and PIA), connected to all VPN endpoints accessible to us, and retrieved each endpoint's IP address (487 in total). For the set of known non-VPN IP addresses, we fetched a similar amount of sites from Alexa Top Sites and resolved them to IP addresses (485 in total). We consider these commercial sites suitable for testing a VPN decision procedure since VPN endpoints, similar to commercial sites, are commonly located in data centers rather than residential IP ranges. We included the complete dataset in the supplemental material[2] of the corresponding paper.

To compare the quality of these services, we calculated Recall, Precision, and the F1 score. The F1 score combines Recall and Precision equally and measures classifiers' accuracy. Calculating the precision and the F1 score requires careful consideration since they depend on the prevalence of the underlying category (VPNs), which is what we want to measure in the first place. Therefore, we assumed a prevalence of 0.05 for this part of the evaluation and adjusted our balanced dataset accordingly.

5.4.3. Differential Privacy

Like Jansen and Johnson [149], we decided on a 24-hour measurement period in all cases. Increased measurement periods do not reliably improve the accuracy since the differential privacy approach adds the corresponding amount of noise, and the set of guard users only changes slowly. Our data collectors receive information about each connection's byte count, circuit count, origin IP address, and the number of concurrent connections. Using this available data, we count the incoming connections, the amount of transferred data, and the number of circuits per connection, once for VPNs and once for all incoming traffic.

Our connection-based measurement approach does not allow differentiating between different users' connection data. Instead, we establish action bounds that define the privacy that our collection affords. We protect 24 hours of continuous use, including one new entry connection every other hour, six new circuits per hour (the typical circuit lifetime is 10 minutes), and two preemptive circuits. We also protect 10MiB of traffic in either direction. Since we did not know how many counts per hour we could expect, we conducted short pilot measurements and extrapolated these counts to 24 hours. Similar to Jansen and Johnson [149], we use the differential-privacy parameters of $\epsilon = 0.3$ and $\delta = 10^{-6}$ for our PrivCount deployment. We used the defined action bounds, the expected counts, and the differential-privacy parameters to calculate the appropriate noise levels with the noise computation script included in the PrivCount project.

5.4.4. Analysis

We analyze the measurement results by adjusting all the VPN traffic counts, calculating the 95% confidence intervals, and calculating the share that VPN traffic contributes to the overall counts.

[2] `https://publications.cispa.saarland/3996/`

Foreman [87] describes several methods that accurately count positives with inaccurate classifiers. One of the methods he describes is the "Adjusted Count" shown in Equation 5.1. This method only requires knowing the true-positive (TPR) and false-positive rates (FPR). These are independent of the prevalence, so we derived them using our measurements from Table 5.1.

$$adjusted = \frac{observed - FPR * total}{TPR - FPR} \tag{5.1}$$

Using this method, we adjust all measurements of VPN traffic before analyzing them further. We report both the observed and the adjusted measurements.

PrivCount adds noise to all counter values. This noise has a normal distribution, with a mean of zero and a pre-calculated standard deviation (see Section 2.3 in Jansen and Johnson [149]). We compute the 95% confidence interval to report the uncertainty that the added noise introduces.

When we calculated the VPN traffic's share of the overall counts, we used interval calculus operations to calculate the resulting intervals. These operations increase the range of the 95% confidence intervals for these measurements.

To ensure replicability and facilitate research cooperation, we will publish the scripts we used to evaluate the VPN detection services, the PrivCount-patched Tor version, our modified PrivCount project, and the evaluation scripts on Github.

We present the results of our measurement study in three parts: (a) the evaluation results of the VPN detection services in Section 5.4.6, (b) a comparison of VPN and non-VPN traffic patterns to check for systemic biases in Section 5.4.7, and (c) the measurement of the prevalence of VPN use in the Tor network in Section 5.4.8.

5.4.5. Ethical Considerations and Limitations

Before collecting data, we submitted a description of our motivation and methodology to the Tor Research Safety Board (TRSB). We adapted our method according to the feedback we received. Furthermore, our institution's ethical research board (ERB) approved this study.

In this measurement study, we face two ethical challenges: (1) Participants could not consent to the data collection since Tor clients choose guard nodes randomly (weighted by bandwidth). To mitigate this lack of information, the domain name of the Tor relays indicated their use for research purposes. In the Tor browser, clicking the lock icon in the address bar shows the currently used Tor circuit and addresses of the individual Tor nodes. Accessing the Tor node address with a web browser led to a website that explained the purpose of our research and how users could opt out of participation. However, since we did not log access to this website and had no feasible option to measure how many users opted out, we can not say how this affected the validity of our measurement. (2) The measurement procedure handled sensitive data, namely the IP address and access time of connecting users. The measurement setup did not store these data points but discarded them after increasing the corresponding counters. After careful evaluation, we checked for VPNs with a local version of IP2Location's database. Hence, the users' IP addresses did not leave the guard node. To ensure our

criteria for handling sensitive data, we only used Tor nodes under our control and did not collaborate with other established Tor guard families. We did not retain or share users' IP addresses and time of connection at any point. Neither the Tor daemon nor any other service logged the incoming connections on these Tor relays. Similar to prior work [149], we employed differential privacy to protect the data of individual users. Attackers cannot tie data that leaves the Tor measurement nodes to specific users as long as their traffic stays in our protected action bounds. The main risk associated with our data collection is an inference attack on the behavior of specific users of our Tor nodes. Our differential privacy approach mitigates this risk in case the users remain inside our defined action bounds.

All of our measurement nodes were located in Germany. If a relevant number of Tor users opted out of using Tor nodes in Germany, this may have impacted our sample's quality. However, Germany is the most common location for Tor nodes, with around a fifth of all Tor nodes, and the data center we used is the second most popular data center for Tor nodes. Hence, opting out of Tor nodes from the data center we used or Germany as a country would result in significantly fewer available Tor nodes, impacting the overall user experience. User behavior changes over time, and current events affect them. Hence, repeating the measurement of VPN use in the Tor network may yield different results. While we can measure behavior, the measurement data does not tell us about users' reasons for their behavior. Hence, we conduct a user survey and a document analysis to complement the measurement results (see Section 5.5).

5.4.6. Evaluation of VPN Detection Services

Our market research of existing VPN detection services resulted in five different candidates. We evaluated them with known VPN and non-VPN addresses as described in Section 5.4.2. We calculated recall, precision, and the F1 score for each service. We assumed a prevalence of 0.05 for the calculation of the latter two. Table 5.1 shows the results of this evaluation.

The low precision of most services in the list at a low prevalence would overestimate the amount of VPN traffic in our measurement. Only Shodan and IP2Location provide suitable precision. However, since Shodan relies on manually labeled datasets, the recall is very low. Even though we prioritize these services' precision, we also look at the F1 score, which combines recall and precision. IP2Location's service achieves our dataset's best F1 score (0.87). Therefore, we used the IP2Location database, which our measurement setup downloaded every time before starting a measurement. We improved its count accuracy with the adjustment method described by Foreman [87].

5.4.7. Validating (Non-)VPN Traffic Classification

We compared the VPN and non-VPN traffic in our measurement period to further validate our VPN classification procedure. We expect that VPN users' traffic follows the same patterns as non-VPN users' traffic. Significant differences in traffic patterns suggest a systemic bias in our classification procedure.

We collected the following information for each connection: lifetime, amount of circuits, concurrent connections, and the number of transferred bytes. The cumulative

Table 5.1.: Evaluation of selected VPN detection services

	PC	IPQS	Shodan	MM	IP2L
True Positive	456	474	106	482	404
False Negative	31	13	381	5	83
True Negative	353	94	485	306	483
False Positive	132	391	0	179	2
Recall	0.94	0.97	0.22	0.99	0.83
Precision*	0.15	0.06	1.00	0.12	0.91
F1*	0.26	0.11	0.36	0.22	0.87

* Assumes a VPN prevalence of 0.05
VPN detection services: Proxycheck (PC), IPQualityScore (IPQS), Shodan, Maxmind (MM), and IP2Location (IP2L)

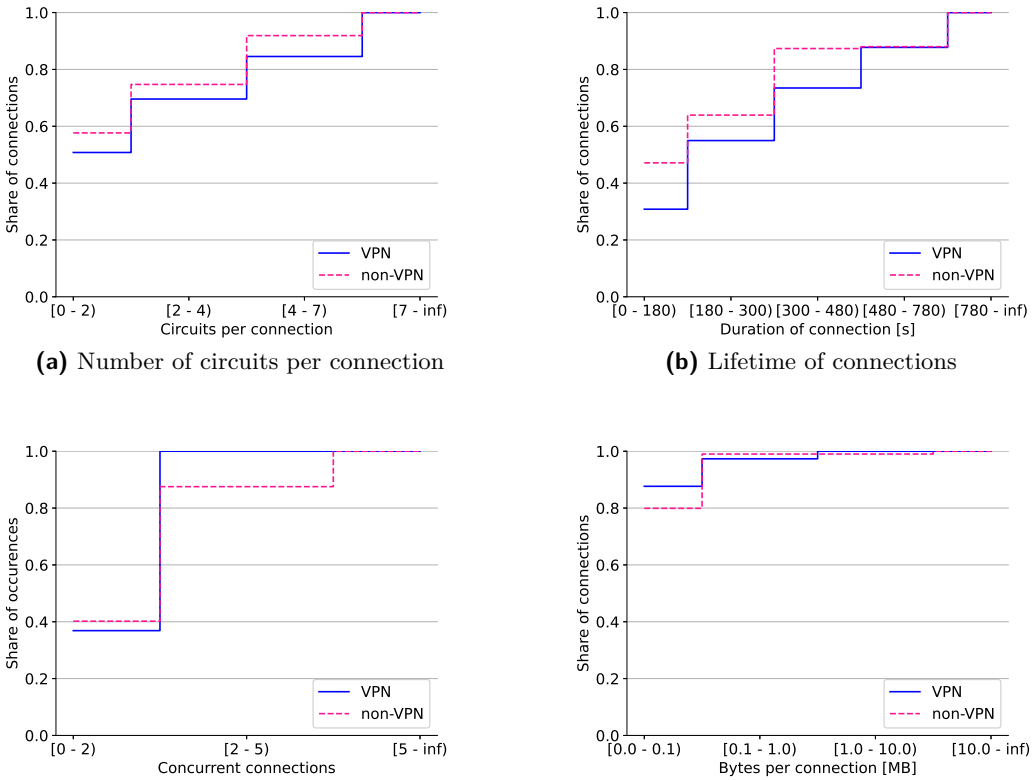
histograms in Figure 5.4 highlight the similarities and differences in these data points between the VPN users (solid blue lines) and non-VPN users (dashed red lines). Since the base rate of both groups differed in our dataset, we scaled the results along the y-axis for easier comparison. Subfigures (a), (c), and (d) in Figure 5.4 show similar VPN and non-VPN traffic in our measurement period. Increasing our confidence that the classification does not introduce major systematic bias.

Subfigure (b), a cumulative histogram of the connections' lifetime, is the only one that shows some differences. While the number of VPN connections monotonously decreases with increasing lifetime, the non-VPN connections' lifetime shows two spikes. The first spike is between 0 and 3 minutes, which is interesting since Tor should keep connections open for at least 3 minutes, even in cases without activity [60]. The second spike appears at five minutes. We have no conclusive explanation for these differences in connection lifetime between VPN and non-VPN traffic. Another interesting detail is that, for both VPN and non-VPN traffic, many connections only have one circuit. However, according to the Tor path specification [60], Tor clients aim to maintain two pre-built circuits for each recently seen port.

5.4.8. Measured Tor over VPN Prevalence

Using our setup, we measured three data types: the number of connections, circuits, and transferred data. We separate these data types into two bins, the total amount and the amount coming from detected VPN endpoints. Table 5.2 shows an overview of these measurement results. Next to each reported measurement, we added the 95% confidence interval in brackets. We report the adjusted counts [87] as well as the observed counts. Additionally, we calculate the ratio of VPN counts to total counts for all measurement variables. We used interval calculus operations to determine the resulting intervals for the calculated shares.

As Table 5.2 shows, our family of four guard nodes handled over 10 million connections, 3 TiB of data, and 43 million circuits in the 24-hour-long measurement period. A total of 6.23% of the connections, 6.59% of the transferred data, and 6.65% of the circuits



(a) Number of circuits per connection

(b) Lifetime of connections

(c) Number of concurrent connections

(d) Transferred data per connection

Figure 5.4.: Cumulative step histograms that compare VPN (solid blue line) and non-VPN (dashed red line) connections

originate from VPNs. The 95% confidence interval is $< 1\%$ in all three cases. These traffic measurements include all different ways of using a VPN to access the Tor network, i.e., regular Tor Browser users and users of VPN providers' Tor mode as described in Section 5.2. These results confirm that the practice of accessing the Tor network via a VPN provider is comparatively widespread, considering its doubtful benefits [273, 272, 236].

5.5. Survey of Users' Tor over VPN Intentions and Beliefs

In the context of the theory of reasoned action (see Figure 5.2), our measurement approach shed light on Tor users' behavior but did not help us understand their intentions, beliefs, or the background factors that led to them. For example, users may always use a VPN or explicitly apply Tor over VPN as a security practice regardless of context. Hence, we continue our investigation with a survey to understand Tor users' intentions and initial beliefs behind the behavior.

To investigate, we initially searched for prior work that provides insights into VPN usage patterns. According to the Global VPN Usage Report 2020 [271], 31% of internet

Table 5.2.: PrivCount measurement results for connections, transferred data, and circuits before and after adjusting the counts.

	Connections (×1000)	Transferred Data (GiB)	Circuits (×1000)
Total	10 186.9 [9 662.4, 10 711.4]	3 039.3 [2 984.7, 3 094.0]	42 952.5 [42 092.1, 43 812.8]
# VPN			
observed	565.6 [546.8, 5 843.9]	177.7 [174.6, 180.9]	2 533.4 [2 501.7, 2 565.2]
adjusted	634.3 [615.5, 653.1]	200.2 [197.0, 203.3]	2 854.6 [2 822.8, 2 886.3]
% VPN			
observed	5.55 [5.10, 6.05]	5.85 [5.64, 6.06]	5.90 [5.71, 6.09]
adjusted	6.23 [5.75, 6.76]	6.59 [6.37, 6.81]	6.65 [6.44, 6.86]

Note: The brackets next to each result include the 95% confidence interval.

users have used a VPN in the month before the survey. Of these VPN users, 36% reported using a VPN on a (nearly) daily basis. The report does not provide details on general VPN use that does not consider situation-specific context.However, the report states that accessing the Tor Browser is a reason for VPN use for 14% of VPN users. While this does not explain how much of our measurement results can be explained by these users, it confirms Story et al.'s [264] observation that some users intentionally apply 'Tor over VPN.' We conducted an online survey to help us interpret our measurement results and understand users' reasons for combining Tor with VPNs.

5.5.1. Method

We recruited $N = 119$ participants on Amazon MTurk. To learn more about Tor over VPN behavior and perceptions, we focused our recruitment on participants who had used both the Tor Browser and VPNs but not necessarily in combination. This results in a narrower sample than the previous measurement study, where we sampled real-world Tor users corresponding to our measurement nodes' bandwidth ratio.

We wanted to recruit participants with VPN and the Tor Browser experience. Only around one-third of Internet users seem to use VPNs at least occasionally [271] and definitive numbers for the prevalence of Tor Browser use are hard to find. Consequently, we were concerned about how many participants with that experience we would be able to recruit. As Kang et al. [150] showed, MTurk workers are more privacy-conscious than the U.S. public. Hence, we chose it for this survey because we hoped to reach our targeted population better on MTurk. However, as Tang et al. [268] showed after we conducted the survey, the same is true for Prolific. Since they also found that samples from Prolific or CloudResearch offer increased data quality, we would use one of them for future work on the Tor over VPN phenomenon.

Section A.1 of the Appendix contains the complete questionnaire. We asked if the participants used a VPN, whether they accessed VPNs for specific purposes or regardless of context, and whether they preferred using a VPN to access the Tor network or not. For the latter two questions, we included open-ended questions asking participants to explain their preferences.

We evaluated the open-ended questions using thematic analysis [34]. One author

performed an open-coding step on all responses to construct one codebook per question. Section A.2 contains all resulting codebooks. Another author used these codebooks to code the responses independently. During the coding process, both authors noted potential themes in the data. The initial inter-coder agreement Krippendorff's alpha was $\alpha = .59$ and $\alpha = .64$ for the first and second questions, respectively. Both authors met and discussed conflicts while updating the codebooks as necessary. The discussion increased the inter-coder agreement Krippendorff's alpha to $\alpha = .89$ and $\alpha = .94$. Finally, both authors discussed their notes and identified themes in the data. We present these themes in Section 5.5.4.

We required that participating workers maintain an approval rate of at least 95% for their past MTurk tasks. We received data from 190 MTurk workers. They spent an average of 3 minutes and 52 seconds on the survey, and we compensated all of them with USD 0.62. We excluded data from workers that have not used Tor (31) or failed the attention check question (40), resulting in a final dataset with responses from 119 participants. The median age of the included participants was 30 ($std = 7.95$). Overall, 27 (23%) women and 92 (77%) men participated. No participant selected one of the remaining response options, i.e., 'non-binary,' 'I prefer not to say,' or 'I prefer to self-describe.'

5.5.2. Ethical Considerations and Limitations

For our belief survey, we informed all participants about the study purpose and data handling practices on the first page of our survey. We compensated all survey participants and did not ask questions about their use case for the Tor browser. Per Amazon MTurk's guidelines, we did not collect any personally identifiable data. Our institution's ethical research board (ERB) approved this study.

Tor users prefer to stay anonymous. Recruiting Tor over VPN users for a survey that questions them about their observed behavior on the Tor network is ethically and technically challenging. For our survey on Tor over VPN beliefs, we recruit users on Amazon MTurk who self-report experience with Tor as well as VPNs.

5.5.3. (Un)intentional Use of Tor over VPN

Seventy-five (63%) of the 119 recruited Tor users had a VPN client installed on one of their devices. Table 5.3 shows these 75 participants preferred way of accessing the Tor network and VPN usage pattern. Sixty participants preferred to access the Tor network using a VPN. Of these, 42 (70%) reported using a VPN for specific situations, and 18 (30%) reported using a VPN regardless of context.

Fundamentally, there are two types of Tor over VPN users: (1) those who use a VPN regardless of context and happen to use Tor at the moment, and (2) those who specifically connect to a VPN before accessing the Tor network. Our survey results suggest that the latter type of intentional user is responsible for most of the VPN traffic in the Tor network.

Table 5.3.: Cross-tabulation of the participants' preferred way of accessing the Tor network and their VPN usage pattern.

		Tor Access Preference			
		VPN	no pref.	no VPN	$\sum$
VPN usage	Specific	42	3	9	54
	Usually	18	2	1	21
	$\sum$	60	5	10	75

5.5.4. Tor over VPN as a Security Practice

Most participants who used a VPN regardless of context, reported doing so for the general benefit of added security. They expected an extra layer of protection to guard their privacy online, with some stating that their VPN software provided them with a more secure experience. Participants connecting to VPN only occasionally named more diverse motivations for doing so: The most prominent reason was to bypass geo-blocking or otherwise hide their location from a service. Participants also used VPNs to hide other personally identifiable information (PII), such as IP addresses, from services. Others used VPN in untrusted networks, such as public WiFi, or due to work requirements. Finally, several participants explained that they connect to a VPN to hide specific activities from observers, most notably their internet provider. These include legal gray areas such as filesharing or torrenting. Three participants, who reported using VPNs for legal gray areas, explicitly named accessing the Tor network as a reason to connect to a VPN. Users seem to associate Tor with activities they would rather hide from internet providers and other observers.

We attributed users' reasons for combining Tor and VPN to behavioral, normative, or control beliefs. When participants mentioned concrete threat models or protection mechanisms, we treated them as behavioral beliefs. When trust issues and others' recommendations were prime concerns, we treated them as normative beliefs. When participants focused on the action itself, i.e., that it is one more thing they could do or how it made them feel, we classified this as a control belief.

Behavioral Beliefs. When investigating users' Tor access preferences (with or without a VPN), we found eleven participants who used a VPN to protect themselves from threats they identified within the Tor network. Some knew that guard nodes could collect their IP address and used a VPN to conceal this information. Participants associated the Tor network with unlawful activities, expected to find *"unsavory characters"* (P102), or suspected a trap by governmental agency: *"I have read some places that Tor was set up by the FBI."* (P95). Therefore, they employed extra protective measures to preserve their privacy. Other participants stated that, due to the open nature of the Tor network, i.e., everyone can contribute to the open-source project and run Tor nodes, malicious actors could easily infiltrate the Tor network and collect information about its users. Similarly, P37 pointed to past incidents affecting the Tor network to explain their distrust: *"There was a problem they did not know about, and the user's information may*

have been shared by someone else." Finally, some participants named uncertainty about the legality of Tor usage (in their country) as a reason to use a VPN before connecting to the guard node, hoping to evade prosecution. However, no participants brought up Tor bridges (in this context and throughout our study).

Normative Beliefs. Aside from the motivations mentioned above, we found that most participants, who preferred using a VPN when accessing Tor, referred to general security benefits as their main reason. These findings suggest that participants trust their VPN provider more than the Tor network. This difference in trust, even if rooted in wrong assumptions, can be hard to overcome. One possible explanation may be that users actively choose their VPN provider, while Tor randomly assigns guard nodes by default. Therefore, the observed trust might be related to choice-supportive bias, as described by Mather and Johnson [178]. Humans tend to attribute significantly more positive features to options they consciously chose and justify past decisions to themselves by perceiving their choice as superior to other options. Finally, a few participants explained that they adopted the practice of combining a VPN with Tor after getting recommendations from friends or other trusted sources. P74 disclosed: *"After conducting some research before starting to use it, I found it improves your privacy and security by using a VPN to access Tor."*

Control Beliefs. Participants often displayed a linear additive perception of security, where more measures equal improved security. P73 stated: *"If I'm trying to be private about what I'm doing I might as well set up as many safeguards as I can."* This assumption might hold poorly when combining Tor with a VPN, as we examined in Section 5.2. Other participants explained that combining these two technologies made them feel more secure. If such a perception does not coincide with actual security guarantees, users might take increased risks while navigating the web [243].

To summarize, we found that some (11) participants had concrete threat models for using Tor over VPN, such as surveillance by their internet provider or guard nodes, legal uncertainty, or mistrust in the Tor network. However, other participants employed Tor over VPN because they believed in linear additive security (i.e., the more, the better), followed others' recommendations, or simply because they felt more secure that way.

5.6. Analyzing Tor over VPN Information Sources

In the previous Section, we surveyed users' beliefs about the Tor over VPN practice. According to the theory of reasoned action, several background factors influence these beliefs. An important factor, especially for security and privacy-motivated users, is the available information on the topic. Thus, analyzing available information about Tor over VPN may allow us to understand how it affects users' beliefs.

According to Redmiles et al. [237], users often learn about security behavior from news articles (online, print, and TV), online forums, fictional narratives, digital service providers, and advertisements. Hence, we focused on these sources for our document analysis of available Tor over VPN information.

5.6.1. Method

We used a keyword search to acquire an initial corpus of documents, filtered search results to end up with topic-relevant corpus entries, and then analyzed their content.

Creating an initial corpus. After an initial discussion, we acquired documents from online newspapers, VPN service providers, and social media, i.e., Twitter, Reddit, and Stack Exchange. For newspapers, we focused on the 16 most-read online newspapers in the U.S. and Germany [202] since all involved researchers understand them well. We investigated the 20 most popular providers [199]. Since the market for VPN services is highly concentrated [155], these providers likely cover a significant portion of users.

We employed the keywords "Tor VPN" and "Onion VPN" to find relevant content. We used a site-specific Google search query (e.g., "site:mullvad.net Tor VPN") to find relevant newspaper articles and service provider sites. We used the built-in search feature on Reddit and Stack Exchange. For the online newspapers and VPN providers, we included all results on the first Google result page in the sample. For Reddit and Stack Exchange, we sampled the first 200 results for each search query. For Twitter, we used Atlas.ti's built-in Twitter import feature, which limits data collection to the previous seven days. This procedure resulted in an initial corpus of 1015 entries, consisting of 249 newspaper articles, 82 VPN provider sites, 257 threads on Reddit, 310 threads on Stack Exchange, and 117 tweets.

Filtering relevant information sources. The entries in the initial corpus contained material where the terms Tor and VPN just happen to coincide in the same document. However, we were only interested in entries that cover the practice of combining Tor with VPN. Hence, we inspected and filtered all search results. Two researchers independently classified the first 110 corpus entries (i.e., 10.8%), comprising 82 VPN providers pages and 28 threads on Stack Exchange. They made the identical exclusion decision in 92.7% of the cases, corresponding to a Krippendorff's alpha of $\alpha = .87$. Since both researchers agreed on how to apply the exclusion criteria, one of them applied the criteria to the rest of the corpus. This filtering operation resulted in a corpus of 389 entries, consisting of 42 newspaper articles, 47 VPN provider sites, 110 threads on Reddit, 133 threads on Stack Exchange, and 57 tweets. Some entries were ambiguous and allowed for different interpretations, e.g., it was unclear if the article authors recommended using Tor and VPNs separately for different use cases or the two in combination. Both researchers discussed the issue and decided that the readers' potential perception is more relevant than the authors' intentions for our analysis. Thus, the corpus should include these ambiguous cases. During the filtering process, the researchers kept notes on interesting observations and analytic thoughts, which they used as a starting point for the qualitative analysis.

Qualitative analysis. We combined deductive and inductive approaches to code the corpus entries. Because of our previous study part and our alignment with the theory of reasoned action, we deductively focused on information that may appeal to consumers' behavioral, normative, and control beliefs. Hence, explanations about the supposed

effects of combining Tor over VPN that include threat models would align with behavioral beliefs, statements about who is or should use Tor over VPN contribute to normative beliefs, and justifications about the ease of use align with control beliefs. However, we also applied inductive open coding to other types of information, e.g., conceptions of technology, when we found them relevant to the analysis of the Tor over VPN practice. As Ortloff et al. [S4] recommend for qualitative research in security, we describe the involved researchers' level of expertise: Two early-career researchers conducted this qualitative analysis of information sources. Both hold a Master's degree in computer science, have a strong background in security, and prior published research focusing on qualitative analysis. The two researchers started by cooperatively coding 48 documents, twelve each from VPN provider pages, newspaper articles, Reddit posts, and Stack Exchange questions, introducing codes as necessary. Afterward, they discussed the initial codebook and used categories to structure it. Using this initial codebook, both researchers independently coded the rest of the documents in three sessions, comprising 48, 50, and 188 documents, respectively. Both researchers kept notes about general observations and annotated quotes when they implied new codes or concepts. After each session, they discussed and resolved disagreements, adapting the codebook as necessary. In order to identify relevant areas and topics for discussion, the researchers calculated inter-coder agreement after each step, both across document groups and code categories [189]. The resulting inter-coder agreement values aided in adapting the coding process, if necessary. Most disagreements arose because the semantic meaning of the assigned codes was quite close or because the underlying data was ambiguous and allowed different interpretations by readers, including consumers seeking information. In the first individual coding session, we assigned codes to entire documents, which resulted in low inter-coder agreement values for detailed discussion threads on Stack Exchange and Reddit. Therefore, one researcher segmented these documents before the second and third individual coding sessions. In the second individual coding session, we found almost no new concepts, suggesting we had reached data saturation with our chosen information sources. Thus, we coded the rest of the documents in the third individual coding session. Since both researchers coded the entire corpus and agreed on the established concepts through discussion, it was optional to calculate an overall inter-coder agreement value [189]. Finally, the researchers discussed how the analyzed data answers the research questions and how to structure and present the findings. Section A.3 contains the final codebook and the supplemental material[3] of the corresponding paper contains the complete filtered corpus.

5.6.2. Ethical Considerations and Limitations

For our document analysis, we collected and analyzed users' comments on Reddit, Stack Exchange, and Twitter. We considered it infeasible to collect informed consent from all of them and filter the research data accordingly. In line with Gilbert et al. [105] recommendations, we try to adhere to the context-dependent norms and platform affordances of Twitter, Reddit, and StackExchange: We did not collect or use any limited-visibility communication and did not try to predict any identities from collected

[3] `https://publications.cispa.saarland/3996/`

site data. On the contrary, we protect site users (except for well-known public figures) by not analyzing, reporting, or quoting any information that could potentially identify them – which is in line with the recommendations by Fiesler and Proferes [83] on research with Twitter data. Our institution's ethical research board (ERB) approved this study.

Our document analysis considers document types ranging from newspapers to social media channels. However, information consumption behaviors vary among demographic groups, leading to different discussions and presented information. Hence, including further information sources may lead to different analysis outcomes.

5.6.3. How Tor over VPN Information Sources Affect Beliefs

Overview of the filtered corpus. We begin with an overview of the characteristics of data in our filtered corpus to provide context for our document analysis. Our corpus spans the period from 2010 to 2022. Renowned newspapers, e.g., The New York Times, The Washington Post, or NPR, write about Tor and VPNs in the context of current events. For example, ongoing censorship in Russia during the war in Ukraine, access to abortion care in the U.S. following the Roe v. Wade repeal, or the 2017 repeal of F.C.C. rules that limit internet providers from selling customers' online information. Some of these newspaper articles contain recommendations from security experts and describe the security behavior of people affected by these current events. Less renowned news venues, such as Fox News or the German Bild, mainly mentioned Tor and VPNs in listicles of generic security tips or VPN-provider-sponsored articles. Almost all VPN providers in our list had blogs, wikis, or Q&A platforms explaining the difference between Tor and VPN, discussing the use of Tor over VPN, advertising their VPN servers' Tor feature, or instructing readers how to use a VPN to access the dark net. On Twitter, we could search for tweets in the last seven days. Consequently, we found Tor and VPN mentioned in the context of current events, in our case (July 2022), mainly the Roe v. Wade repeal and safe access to abortion health care. Tweets are limited to 280 characters, so many contained only instructions or a short list of recommended security and privacy tools. On Reddit and Stack Exchange, we discovered threads that discuss combining Tor and VPNs in various communities. On Stack Exchange, most threads were in security, privacy, or system-administration-specific communities, but we also found discussions in cryptocurrency, fictional writing, and law-focused communities. All communities that we identified on Stack Exchange were also active on Reddit. However, we also found active discussions in communities related to online drug shopping, piracy, and conspiracy theories. We also found discussions about Tor and VPN use in location-specific, political, or religion-affiliated communities in the context of current events, such as protests or ongoing repression. On Reddit, we found some instances of (former) members of oppressive social communities discussing recommended safety and security precautions for leaving the community. We reference quotes from our documents in the format "<Venue><Document-ID>:<Quote-ID>" to convey the general source to readers while enabling us to trace the quote's origin in our corpus.

Behavioral beliefs. Behavioral beliefs are about the negative or positive consequences users might experience when following a specific behavior [85]. Information that informs these beliefs about Tor over VPN needs to cover the potential effects of the practice. Assessing the quality of security measures relies on threat models, i.e., a specific potential threat that a measure mitigates. Most documents in our corpus, especially news articles, and Twitter, contained only implicit, vague, or all-encompassing threat models, e.g., *"[w]hen you surf the internet, everyone is watching"* (News986:1).

VPN providers focused on the supposed dangers of Tor (*"There's a chance that your computer may be used as an end-relay for illegal activity. "* (VPN43:4)) and hiding activities from internet providers and the government. They often warned about the threat of untrustworthy VPNs (not themselves). VPN providers' articles also contradict other VPN providers or even their own statements about the effects of Tor over VPN. These contradictions usually depended on the effects of Tor over VPN and VPN over Tor they discussed. Sometimes they just blended the effects of both approaches into one. Unsurprisingly, most VPN providers who covered the topic recommended using a VPN to access the Tor network and advertised their VPN servers' Tor feature if available. Discussions about the effects of Tor over VPN were far more polarizing on Reddit and Stack Exchange. The range of opinions covered: absolutely necessary, beneficial, useless but harmless, and outright dangerous. On the more Tor-specific Reddit and Stack Exchange communities, the most common repeated answer to these discussions was that *"[t]he need for one, or both, depends on the user threat model"* (Reddit724:8) and is not needed for most users. However, they usually do not explicitly discuss which personal threat models warrant using Tor over VPN. While some threat models are relevant for all internet users, e.g., credential stuffing with leaked password data or malware attacks on users' outdated software, many threats do not affect internet users equally. Tor over VPN as a security practice appears to belong in the latter category. In general, interested readers looking for an authoritative answer to the question of what effect Tor over VPN has need to read discussions and choose an answer they like.

Normative beliefs. Injunctive normative beliefs concern others' assumed approval or disapproval of certain behaviors, which creates a social influence on behavior [85]. Descriptive normative beliefs concern peoples' perceptions of typically performed behavior based on observations. Related security research refers to this behavior as social proof [44, 50, 51]. Our analysis focused on these descriptive norms, i.e., what users would find in information sources when they look for social proof about the appropriate use of Tor. In many online discussions and fewer news articles, we found examples of people who combined Tor and VPNs and whom readers may perceive as experts. Their expertise expressed itself in often detailed technical questions that did not question the usefulness of Tor over VPN. Our teaser image in Figure 5.1 of a scene in the popular TV series Money Heist is a similar situation: Rio, a bona fide hacker, explains to the Professor that he accesses the dark net with a VPN and the Tor Browser. People who consume TV series may build up para-social relationships with the characters in them [106]. Hence, a scene like that, while not an explicit form of security advice, can create a perceived norm for people looking for social proof. Some news articles and VPN provider sites also expressed these perceived norms directly: For example, that people

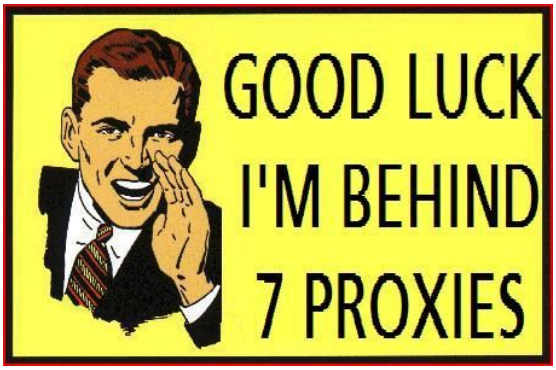


Figure 5.5.: The "Good Luck, I'm Behind 7 Proxies" meme, a sarcastic catchphrase that originated on 4chan in 2007 [194]

who are *"very serious about browsing anonymity"* (VPN43:2) or *"super security-focused"* (News913:2) can use Tor over VPN, and that Tor *"is popular among computer-savvy circles"* (News913:2). Consequently, some discussion participants assumed that using Tor over VPN is already a general practice, e.g., *"Pretty sure most Tor users do [use both at the same time]"* (Reddit708:6) or that using Tor by itself goes against best practice.

Perceived norms also manifest as jokes and memes about layering multiple security mechanisms on top of each other, e.g., *"OP is behind 3 VPNs, 2 reddit acounts [sic], a proxy, and also using TOR."* (Reddit708:2) or *"We can't find him! He's using 7 proxies in incognito mode!!"* (Reddit661:2). These comments seem to refer to the "Good Luck, I'm Behind 7 Proxies" meme, shown in Figure 5.5, which originated as a sarcastic retort on 4chan in 2007 [194].

While we sometimes had the impression that these references were sarcastic, discussion participants seemed to take them seriously more often than not, illustrated by this comment on Reddit: *"Seems the safest path is You -> Tor -> VPN -> Tor -> website."* (Reddit661:9); or an answer on Stack Exchange discussing the following setup: *"You –> VPN Provider (Probably listening by NSA) –> Tor –> Proxy (Any free proxy) –> VPN Provider(Another one listening by NSA) –> Tor –> Proxy (Any free proxy) –> Target"* (SE96:4).

Appeals to authority on the subject also influence perceived norms. Discussion participants on Reddit and Stack Exchange regularly point to Tor- or Tails[4]-affiliated sites to strengthen their arguments. To a similar effect, news articles also include quotes and recommendations from renowned security experts. Other factors also establish a norm of combining Tor and VPNs. People often recommend using Tor over VPN without justifying it any further or explaining its effects: *"Tor + VPN is a good combo"* (Reddit708:5). Several VPN providers officially support Tor so that customers do not require the Tor Browser to access onion links, which lends credibility to the idea of combining Tor with a VPN. Thus, we found several indicators that information sources influence normative beliefs and offer social proof for using Tor over VPN.

[4]Tails is a security-focused Linux distribution. It routes all data traffic through the Tor network.

Control beliefs. Control beliefs focus on personal and environmental aspects that support or hinder behavior [85]. Consequently, these beliefs impact people's perceived self-efficacy. Our corpus contained little information that could affect people's self-efficacy. We found specific step-by-step instructions on combining Tor and VPNs to achieve a particular goal, such as accessing the dark web safely or buying drugs online. These instructions usually did not explain the purpose or security of that practice. Achieving the primary goal is the only relevant aspect of the instructions. Other sources emphasize that using Tor over VPN or even the VPN-supported Tor feature is easy. That so few information sources cover these aspects could mean that VPNs and the Tor Browser are already easy to use or that the target demographic already has a high level of self-efficacy.

Misinterpretation of security advice involving Tor and VPNs. During the filtering process, we unexpectedly found that some sources contain ambiguous information on the combination of Tor and VPN, allowing for misinterpretation. Misinterpreting information sources may also impact users' normative beliefs about recommended practices. We included these ambiguous information sources in our corpus and labeled them with separate codes. We found two types of these information sources: First, we found lists of recommended security practices. While they do not directly recommend a combination of Tor and VPNs, these lists usually imply that users can combine all recommendations on that list and that each security advice provides additional security. We found these types of lists in news articles and social media posts. Second, we found ambiguous phrasing that may imply a combination of Tor and VPN, even when the author may have meant a separate use of these technologies. For example, a news article mentioned that internet users could use *"security tools like VPNs and Tor browser to bypass censorship"* (News970:1) and a Reddit user advised others to *"learn to how to use a VPN and Tor"* (Reddit435:2). This issue also affects statements from established security experts in well-known newspapers such as The New York Times, which said that *"many Russians were skilled at using it [Tor] and VPNs [for evading censorship]"* (News932:2). Journalists often look for short quotes from outside sources, making it difficult for experts to convey detailed information and recommendations. We found an example that illustrates how users misinterpret ambiguous information sources: In tweet Twitter1000:8, someone responded to the U.S. politician Alexandria Ocasio-Cortez, who disseminated a summarized version of EFF's digital safety tips for people seeking an abortion and providers of abortion support. The original message recommended *"steps like using a VPN, throwaway email addresses, private browsers"* and the commentator chimed in to say *"Yeah people get a tor and vpn and use duck duck go"*.

Misconceptions of the technology behind Tor and VPNs. User perception of the technology behind Tor and VPN may impact users' behavioral beliefs about Tor over VPN. Some people in online discussions saw the Tor browser as just one possible choice of privacy-preserving browser, listing it at the same level as Brave and the DuckDuckGo Privacy Browser, e.g., the Twitter user who recommended to *"Get a VPN, download tor or DuckDuckGo [Privacy Browser]"* (Twitter1000:33). For users who think of the Tor browser as just a browser with some extra features, it may seem intuitive that a VPN is

required to protect the traffic. A group of people views Tor as just a particular type of VPN, applying the same threat models to both. They may have been familiar with VPNs before learning about Tor and inadvertently transferred their conception from one technology to another. Story et al. [264] found that experience with VPNs and the Tor browser is associated with confusion about these tools' protection. Lastly, we also identified many arguments centering around the concept of layering security mechanisms, where discussion participants implicitly applied the Swiss cheese model [235] to minimize risks. While this is not a misconception, it is a specific conception of security tools related to users' control beliefs. Applying this concept of "cumulative act effect" too broadly may lead to the conclusion that it is always helpful to add additional security tools.

5.7. Discussion

5.7.1. Users Practice Tor over VPN with the Intention to Improve their General Security and Anonymity

We measured that 6.23% of connections to the Tor network originate from VPNs (see Section 5.4), corresponding to an estimated 140,000 daily users[5] at the time of writing. As our survey (see Section 5.5) indicates, most do not accidentally practice Tor over VPN. They intend to improve the security and start their VPN client specifically for their Tor session. They believe it increases their general security and protects them from unspecified dangers on the Tor network. Our document analysis of background sources (see Section 5.6) suggests that this can be explained by VPN providers who overemphasize or misrepresent the potential dangers of Tor in order to promote their products. We also found many recommendations to use Tor over VPN without describing the practice's purpose and effects, which explains why users expect general security benefits. This belief in security benefits is an issue since users expect more security and anonymity when not warranted. In the case of risk compensation behavior [243], this may result in net-negative security. Additionally, users have a limited compliance budget [23] for security practices. Hence, security practices with inconclusive benefits should be replaced by ones that provide general security benefits, such as using a password manager to create secure passwords or making backups to mitigate data loss.

5.7.2. Security Folklore Appears to Spread through Normative Beliefs

Results from our user survey (in Section 5.5) and document analysis (in Section 5.6) suggest that normative beliefs based on social proof contribute most to spreading security folklore. While we could find detailed discussions on the purpose and effects of Tor over VPN (behavioral beliefs) in expert communities on Reddit and Stack Exchange, we rarely found them in other places. These discussions included contradicting opinions and often had no clear outcome. However, VPN providers' websites gave some suggestions on Tor over VPN's purpose, i.e., hiding Tor use from internet providers and protecting

[5] `https://metrics.torproject.org/userstats-relay-country.html`

from the supposed dangers of Tor, which also came up again in the expert discussions on Reddit and Stack Exchange as well as our survey on users' intentions and beliefs.

In contrast, we found evidence for normative beliefs and quotes that provide social proof in all types of documents. Quotes that support normative beliefs come without threat models and descriptions of consequences, focusing instead on the people that implement the practice. Discussions that assume Tor over VPN is a general practice, news articles that say that "security-focused users" implement it, and pop-culture media mentions of the practice all contribute to establishing perceived norms of security practice. Even the security community is guilty of doing this: "Use Signal, use Tor" is a really common but criticized form of security advice without specific threat models. Even Edward Snowden tweeting the advice in 2016 [69] has firmly established it as a perceived norm. Today, memes, stickers, and t-shirts include this security advice. Hassoun et al. [133] found that Gen Zers' informational and social needs are inseparably entangled, using information to orient themselves socially and define their emerging identities. A similar entanglement of informational and social needs could explain some of our results on normative beliefs.

Hence, we argue that normative beliefs contribute to the spread of security folklore for social reasons when people look to others to decide which behavior is appropriately secure. Of course, this works better when recommended practices are easy to implement, i.e., when self-efficacy is not an issue (control beliefs). Lastly, the role of normative beliefs in spreading this Tor over VPN use is also underlined by the fact that most participants in the survey expected general security benefits, i.e., they did not associate any behavioral beliefs with the practice.

In Section 5.3, we described two types of normative beliefs: injunctive norms that focus on others' potential judgment of one's behavior and descriptive norms that stem from observing others' behavior. In general, injunctive norms have more influence than descriptive norms regarding actually swaying behaviors. While injunctive and descriptive normative beliefs likely play a role in the analyzed information sources, we deliberately focused on descriptive norms. Researching injunctive norms in an online setting is difficult since it is unclear how much discussion participants care about their peers' judgment – they never really have to reveal their actual behavior to them.

While the results of this work demonstrate which type of beliefs affect the spread of this particular security behavior, it raises new questions for future work about the exact origins of advice and beliefs and their spreading path across platforms.

5.7.3. Giving Effective Security Advice for Specific Use-Cases

Our case study illustrates some essential points for providing effective security advice: First, demonstrated security behavior or stories of security practices influence normative beliefs and should be treated as a form of (unintentional) security advice. We should try to communicate the minimal set of security advice with the most practicality and general security benefits [240]. However, it is not possible to only tell stories about security practices that have general security benefits. Hence, not only security advice [33, 101] but also stories about security practices, e.g., in news articles, need to include an explanation of their purpose.

Second, it is easy for readers to misinterpret security advice or stories about security practices, i.e., listicles that imply combination behavior, ambiguous phrasing, or lack of threat models. Therefore, checking for possible misinterpretations and providing extra space when necessary should be a required part of a responsible publication process on security topics.

Third, security is usually not the users' primary goal. Giving clear instructions on how to achieve the primary goal securely may be an effective method of giving security advice for exceptional use cases. VPN providers already apply this approach; many have articles instructing users to connect to a VPN before accessing the dark net.

Lastly, layering multiple security mechanisms seems to create a perceived sense of security, i.e., a secure experience. Avoiding these unwarranted secure experiences requires design approaches that take them into account. Designing the Tor browser differently so that it informs users who connect with a VPN of its security implications may combat an unwarranted perceived sense of security.

5.7.4. User Perceptions of Security Technology Shape their Behavior

In the user survey (Section 5.5) and the document analysis (Section 5.6), we found several cases where users' perceptions of the technology behind Tor and VPNs shaped their beliefs and self-reported behavior. We found that some users think of the Tor browser as a regular privacy-preserving browser that merely adds countermeasures at the application layer, e.g., protecting from behavioral online tracking by blocking third-party cookies. Such a belief would explain why these users think that protecting the network layer, i.e., the browsing traffic, with a VPN is beneficial for using the Tor browser securely – even though it is not.

Other users thought of Tor in the same way as VPNs and thus treated both in the same way, which confirms Story et al.'s [264] finding that users had trouble differentiating Tor and VPNs. We also found indications that users apply the Swiss cheese model of risk management [235] to justify layering security practices. Lastly, the responses to our user survey suggest that VPN providers have successfully created the perception that Tor is unsafe to use without a VPN. A part of the reason for these user perceptions and coping strategies is that the security technology behind Tor and VPNs and its effects are hard to communicate to laypeople.

While not all misconceptions are fixable, some general approaches tackle the misconception that the Tor browser is just a regular browser: names we use for software or features, visualizations that communicate inner workings, and user interaction that support users in building mental models. First, the name of the Tor software would ideally hint at its effect on the network layer while also making it more difficult to confuse with other types of privacy-preserving browsers. The potential effect on user perception did not seem to have been an important factor when renaming the "Tor Browser Bundle" to simply "Tor browser" nine years ago.[6] However, while a better name might help new users, it also has the potential to confuse the existing user base. Hence, future empirical user research must evaluate this option with care. Second, upon starting up the Tor browser, visualizing the ongoing connection attempts that are

[6] `https://gitlab.torproject.org/legacy/trac/-/issues/11193`

usually in the background may improve user understanding of the entire system. Third, upon detecting VPN use, user interaction in the form of prompts could explain that using a VPN in combination is potentially unnecessary. Nudging users in this way to question their beliefs and possibly change their behavior is a well-studied form of soft paternalism that helps users understand and choose appropriate options [6].

5.8. Conclusion

To understand how security folklore spreads and how it could facilitate useful security advice, we investigated a data-rich example, i.e., the practice of using Tor over VPN. Our case study consisted of three study parts aligned with the theory of reasoned action: We quantified behavior on real-world Tor traffic, surveyed users' intentions and beliefs that explain the behavior, and analyzed background information sources that affect user beliefs.

We found that the spread of security folklore relies on the normative beliefs of users looking for social proof [44, 50, 51] on how to use Tor securely. This effect appears especially important in cases such as Tor over VPN when threat models and consequences of security practices are hard to determine for laypeople. Thus, the security community may want to investigate how to effectively establish normative beliefs to spread useful security advice with generally applicable security benefits. However, to avoid the unintentional spread of security misinformation as normative beliefs, reports about security practices in newspaper articles and pop-culture media must explain their effects and accompanying threat models.

We also found that misinterpretations and misconceptions support erroneous beliefs about the security technology behind Tor and VPN and its effects.

Overall, our Tor-specific and generalizable findings suggest that careful phrasing is crucial for communicating information about security practices, and redesigning tools' secure experience through visualizations or UI friction may combat some misconceptions.

Part II.

Adapting Established HCI Methods to Security and Privacy

The contents of the following chapter were published as part of the publication *"Transferring Update Behavior from Smartphones to Smart Consumer Devices"* (SPOSE 2021) [P4]. This paper was produced in cooperation with my co-authors Michaela Neumayr, Oliver Schedler, and Katharina Krombholz. The following table describes the contributions of all authors to this paper:

Author	**Contribution**
Matthias Fassl	I had the original idea for this research project, designed the questionnaires with Michaela, recruited participants with Michaela, conducted the online survey, and analyzed the responses together with Michaela and Oliver. I wrote the draft versions of the iterations of the resulting paper.
Michaela Neumayr	Michaela provided her guidance for the design of the questionnaires, recruited participants and collected data together with me during the initial survey, and conducted the statistical analysis of the survey responses.
Oliver Schedler	Oliver contributed during the qualitative analysis of the open-ended survey responses.
Katharina Krombholz	As my academic advisor, Katharina was involved in the major decisions during this research project. She gave me feedback on the initial idea and the choice of studies, including the methodological approach. She guided me through methodological and ethical questions, helped me refocus the research project after the initial survey, reviewed my interpretation of (intermediate) results, and helped me find an appropriate venue for this research project. She also gave feedback on draft versions and edited parts of the paper.

Reference

Fassl, M., Neumayr, M., Schedler, O., & Krombholz, K. (2022). Transferring Update Behavior from Smartphones to Smart Consumer Devices. In S. Katsikas, C. Lambrinoudakis, N. Cuppens, J. Mylopoulos, C. Kalloniatis, W. Meng, S. Furnell, F. Pallas, J. Pohle, M. A. Sasse, H. Abie, S. Ranise, L. Verderame, E. Cambiaso, J. Maestre Vidal, & M. A. Sotelo Monge (Eds.), Computer Security. ESORICS 2021 International Workshops (Vol. 13106, pp. 357–383). Springer International Publishing. `https://doi.org/10.1007/978-3-030-95484-0_21`

6

Transferring Update Behavior from Smartphones to Smart Consumer Devices

Automatic updates are becoming increasingly common, which minimizes the amount of update decisions that users have to make. Rapidly deployed important updates have a major impact on security. However, automatic updates also reduce the users' opportunities to build useful mental models which makes decision-making harder on other consumer devices without automatic updates. Users generally transfer their understanding from domains that they know well (i.e. smartphones) to others. We investigate how well this transfer process works with respect to updates and if users with automatic updates fare worse than those with manual updates.

We conducted a formative field study ($N = 52$) to observe users' update settings on smartphones and examine reasons for their (de-)activation. Based on the results, we conducted an online survey ($N = 91$) to compare how users perceive update notifications for smartphones and smart consumer devices. One of our main findings is that update decisions based on *expected changes* do not apply well to these devices since participants do not expect meaningful and visual changes. We suggest naming updates for such devices 'maintenance' to move users' expectations from 'new features' to 'ensuring future functionality'.

6.1. Introduction

Keeping systems and software up to date is the most common expert advice for securing devices [241, 146]. Consequently, prior work extensively studied update attitudes and behavior [165, 269, 181, 290, 281, 270]. Vendors introduced partially or fully automatic updates since users often delay or skip updates. Windows 10 introduced intervention-less automatic update downloads and installation, Android and iOS introduced automatic updates, and Google Chrome started using silent automatic updates over ten years ago. Automatic updates improve the rate and speed of update deployment [66]. However, automatic updates create two potential pitfalls: (1) Users feel betrayed as soon as automated systems make choices that defy their expectations [70] and these incidents

will impact all future update decisions [281]; (2) Automated updates reduce users' understanding of what is happening on their computers [290]. These pitfalls diminish users' ability to make informed decisions when updates cannot be fully automated.

Update behaviors and attitudes on desktops and smartphones are well studied [66, 70, 77, 86, 165, 182, 183, 181, 269, 280, 281, 290, 270]. However, smart consumer devices with minimalistic user interfaces (UIs) and inconspicuous computing power became common after the Internet-of-Things emerged. Gartner predicted 20.8 billion IoT devices for 2020, thereof 13.5 billion consumer devices [196]. In contrast to communication and entertainment heavy smartphones, IoT devices control day-to-day life. Some smart consumer devices, e.g., dishwashers, have very minimal UIs, impacting how users perceive and handle updates. However, there is still little research on how users transfer update behavior to other application areas beyond smartphones and desktops. Automatic updates alleviate some update issues. However, sometimes they are neither practical nor safe, and maybe not even possible for devices with limited UIs – making questions on user understanding and engagement even more pressing [26, 256]. Since traditional computing devices move towards automatic updates, awareness of updates' effects and importance decreases. However, users may have to decide on updates again when handling smart consumer devices. It remains unclear how users make their update decisions on these devices and how they transfer their update-knowledge from traditional computing devices.

The aim of this work is (1) to study users' reasons for (de-)activating automatic updates, (2) to understand how users handle manual update decisions on smartphones, and (3) to evaluate if their update reasoning is transferable to smart consumer devices found in the IoT. We conducted an exploratory field study ($N = 52$) on users' reasons for deactivating automatic updates. We used a mixed-methods online survey ($N = 91$) to explore how automatic updates affect users' manual update decisions and how users transfer their update behavior smart consumer devices (in our study: dishwasher, self-lacing shoes, and a modern car). Our main contributions are: (1) we **observed an increased rate of automatic updates** for smartphone apps (compared to Tian et al. [269]) and provide **ranked lists of reasons for (de)activating automatic updates**; (2) we **describe the differences between users that activated automatic updates and those who did not** (3) we discuss how **transferring users' update behavior to smart consumer devices might fail** since two main strategies (evaluation by expected changes and evaluation by notification) are difficult to apply to smart consumer devices; (4) we provide **design implications** for smart consumer device updates.

6.2. Methodology

Guided by the following research questions, we study how automatic updates affect users' remaining update decisions on smartphones and how well these decisions transfer to smart consumer devices

RQ1: How common is deactivation of automatic updates and what are the users' reasons for it?

RQ2: How do users' update attitudes (information demand, perceived importance, and expected effects) transfer from smartphones to smart consumer devices?

Our *formative field study* establishes the share of users who (do not) use automatic updates and their (de)activation reasons. Based on these results, we designed an *online survey* which compares how participants make update decisions on smartphones and smart consumer devices. We explained the purpose, the procedure, and the type of questions to all study participants. We did not collect identifying information and instructed participants to provide screenshots without identifiers.

All participants gave us their informed consent. We compensated participants for their time, based on the minimum wage. Our university's ethical review board approved this study.

6.2.1. Formative Field Study

In the formative field study ($N = 52$), we collected the participants' OS version, the OS update settings (if applicable), and update settings for installed apps. We asked open-ended questions to understand their reasons for changing settings. Afterwards we used a questionnaire to collect demographic data. Section D.2.1 presents the entire questionnaire. We conducted a pre-study with 8 participants. For three days, we recruited participants with Android or Apple phones in front of our university's dining hall during lunch time. Table D.1 in the Appendix presents the demographics.

We analyzed the observed frequencies of smartphone OS settings. We used *open coding* to evaluate qualitative free-response data. Two researchers independently coded the responses and constructed two independent codebooks, then constructed a common codebook (see Section D.2.2) and resolved all disagreements.

6.2.2. Online Survey

The formative field study showed that participants like to maintain control over installed software. They preferred to update apps they considered important and influence the installation time to avoid bugs and data-loss, confirming previous work [280, 281, 183, 77].

Questionnaire. We used those results to construct an online survey on Amazon MTurk ($N = 91$) which exposed participants to five different update scenarios, two for mobile phones (system and app update) and three concerning smart consumer devices (dishwasher, shoes, car). Appendix D.3.2 shows the notifications that we used in the survey. We chose update scenarios that (1) concern devices with a low barrier to use – so most participants could imagine a use-case for them, and (2) includes an update decision that participants will not have faced before. Similarily, Fagan et al. [77] used fictional update notifications to understand users' update behaviors and attitudes. For each update notification, we asked participants to explain the update's importance, what kind of changes they expect, when they would prefer to install it, and how they would redesign the notification.

To evaluate users' responses in context we also asked for their update settings (phone OS version, screenshots of OS and app update settings), their potential update avoidance behavior (connected to WiFi and charging habits), and their 5-point Likert evaluation of (de)activation reasons. Since prior work [116] suggests that update behavior depends on technology-savyness, sense of autonomy, and personality, we added appropriate psychometric scales (Affinity for Technology Interaction (ATI) [88], Reactance to Autonomy [141], and Big Five Inventory (BFI-K) [232]). We asked for general demographic information such as gender, occupation, educational background, and household income and added three attention check questions throughout the survey. Section D.3.1 in the Appendix presents the full questionnaire (translated into English).

Evaluation. We used a repeated-measures ANOVA to find significant differences between perceived importance of the five notifications. We evaluated the open-ended responses to the five notifications with thematic analysis [34]. Two researchers used *open coding* to independently assign initial codes to their part of the data. They used the other's initial codebook to independently code the remaining data, resulting in an inter-coder reliability (ICA) of Brennan and Prediger's $\kappa = 0.63$. In an iterative approach, the two researchers discussed the categories with the most mismatches, renamed or merged codes, and revised the segments in questions, resulting in an inter-coder reliability (ICA) of Brennan and Prediger's $\kappa = 0.83$. During the last session they used *axial coding* to restructured the entire codebook and identify themes. Section D.3.3 in the Appendix contains the final codebook (containing 8 categories with a total of 70 codes).

To understand how well update decisions transfer to smart consumer devices, we qualitatively compare users responses according to their update preferences (automatic vs. manual) and the type of notification they responded to (smartphone vs. smart consumer device). We report differences between those groups if: (1) codes are not included in both groups, (2) the most frequently assigned codes are different, or (3) if a code was assigned three times more often in one group.

Recruitment and Participants. After conducting a pilot study ($N = 3$), we recruited Amazon MTurk workers from Germany with an approval rate of 99.0% and compensated them with USD 5.60. We excluded five of 96 participants, either because the GeoIP results showed that they were not in Germany or two researchers independently agreed that their provided answers did not answer the open questions. Table D.4 in the Appendix presents the demographics.

6.3. Results

We report the prevalence of automatic updates that we observed in our formative field study and our online survey in Section 6.3.1. Using that information we evaluate (in subsection 6.3.2) how activated automatic updates influence the participants' responses to the shown update notifications. In subsection 6.3.3 we describe how participants decide if and when they would like to install updates and how well this decision-process transfers from smartphones to smart consumer devices. During the evaluation we found several contradicting user requirements which we present in subsection 6.3.4.

6.3.1. Automatic Update Settings and Reasons for (De)activation

Most of the participants in the formative field study did not change default update settings. Almost all Android users had operating system updates enabled and used the "WiFi only" option for application updates (the default). Most iOS users had activated OS updates, but more than a third of them deactivated automatic application updates. Table D.2 in the Appendix shows a summary of the observed update settings. Table D.3 compares update settings of users with high (≥ 4) and low (< 4) self-efficacy scores. Participants most commonly mentioned three types of security-relevant practices that they did on a regular basis: *authentication*, *privacy settings*, and *abstention* from potentially useful products or features. Even though our study procedure primed all participants on updates, only four participants mentioned that they regularly apply updates to keep their mobile secure. In the online survey 63 (69%) participants had an Android phone, whereas 28 (31%) had an iPhone. By default, Android enables automatic OS updates, and iOS will ask during the initial setup. 52 participants (57%) had automatic OS updates enabled, 17 (19%) had them disabled, and 22 (24%) did not submit a suitable screenshot. By default, both Android and iOS enable automatic app updates. 79 participants (87%) had enabled automatic app updates, 10 (11%) disabled them, and 2 (2%) did not submit a suitable screenshot.

In the formative field study, the two most common reasons for deactivating updates were the wish to maintain control over installed software or concerns about data usage. Two aspects of maintaining control came up: (1) participants only wanted increased agency over updates for apps they perceived as important enough, and (2) they would like to decide when to install an update since they know from experience that new updates may have bugs and could lead to data-loss. In the formative field study the two most common reasons for participants to activate automatic updates were convenience and the general desire to be up to date. The online survey asked participants to rate these reasons for (de)activation of automatic updates on a 7-point Likert scale (see Table D.5 in the Appendix).

6.3.2. Automatic Updates and their Effect on Update Decisions

We assumed activated automatic updates would influence users in two ways: (1) that some of the users that are unhappy with automatic updates would try to avoid triggering the installation criteria for them (thereby delaying or skipping updates). This would increase participants' agency in deciding the installation time without deactivating automatic updates. (2) that users that are happy with automatic updates would slowly lose the ability to make update decisions over time and factor in fewer potential problems before deciding. In order to find evidence for these assumptions we added two sections to our online survey.

Avoidance behavior. On Android and iOS, automatic updates are performed by default when the phone charges and is connected to a WiFi network. For 80% of the participants the time of day that they most often charge their phone coincides with a time of day that they are usually connected to WiFi. That means that most participants are able to receive their automatic updates during the course of 24 hours and do not show signs

of update avoidance. Table D.6 in the Appendix presents the participants' complete responses.

Effects of automatic update settings. We compared the qualitative answers of participants that activated automatic updates with the answers of participants who favored manual updates. We found no qualitative differences between these groups regarding their preferred installation time and their suggested changes to the update notification. Participants who activated automatic updates mainly mentioned three concepts: (1) updates are necessary for maintenance, (2) updates are necessary for security, and (3) updates can be important without having visible effects. Only participants that favored manual updates stated that they would like to wait for experience reports from other users.

6.3.3. Transferring Update Behavior to Smart Consumer Devices

In an effort to understand how well the users' update behavior transfers to smart consumer or IoT devices, we start by reporting general results on the responses to update notifications shown in the online study. We present our results according to three of the six update stages discovered by Vaniea et al. [280]: deciding, preparation, and deployment. Afterwards, we elaborate on the participants' different attitudes to smartphone and smart consumer device update notifications.

Deciding. Our formative field study indicated that the participants' perception of a manual update's importance influences their decision to install them. Therefore, we asked participants to rate the importance of the presented manual update notifications on a 5-point Likert scale and provide a qualitative explanation. We present the participants' ranking of importance before going into more detail with the qualitative evaluation of the response.

Participants considered system updates the most important type of update (m = 3.69, sd = 1.09), followed by updates for cars (m = 3.1, sd = 1.20), phone apps (m = 2.41, sd = 1.1), and dishwashers (m = 2.29, sd = 1.28). Updates for shoes were considered least important (m = 1.9, sd = 1.03) of all five update notifications. Figure 6.1 provides an overview of the resulting scores and which group comparisons revealed significant differences. We used a one-way repeated measures ANOVA to compare the mean importance scores of the update notifications. Shapiro-Wilk's test indicated that we cannot assume a normal distribution. However, a repeated-measures ANOVA is robust against such a violation. Mauchly's test indicated that the assumption of sphericity had been violated, therefore we report Greenhouse-Geisser corrected tests. Mean scores for the perceived importance of the update situation were statistically different ($F(3.39, 331.94) = 44.25$, $p < .001$,$\eta^2 = .33$). Table 6.1 shows notification comparisons with differences according to the post-hoc tests. The resulting ranking of importance indicates that participants might view smart consumer devices (that are not evidently safety-critical) to be less important than other kind of updates.

The evaluation of the open-ended questions for each update notification resulted in different themes covering the *decision* stage. Many participants reported possible

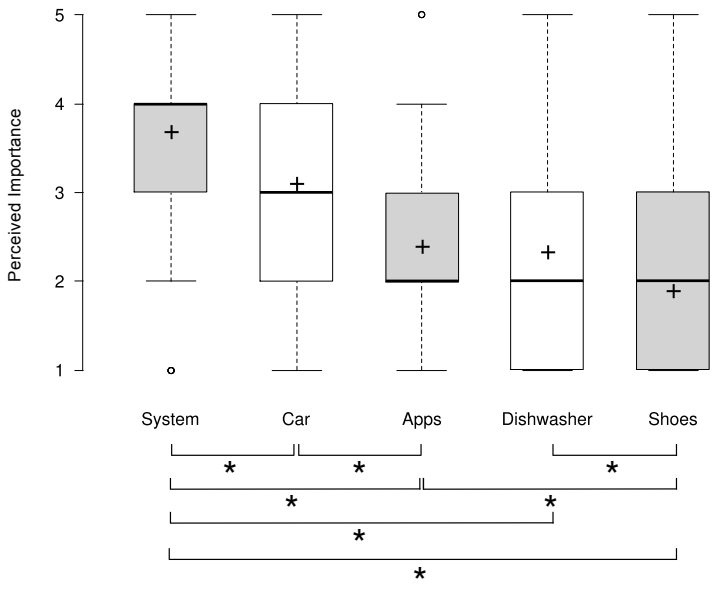


Figure 6.1.: Ranking of updates according to perceived importance (* marks pairwise significant differences)

positive or negative effects that they considered before updating. Amongst others, participants named new features, performance, stability, and usability improvements as potentially positive effects. Almost all of the reported negative effects were based on personal experience: participants reported that some updates removed features, introduced bugs, led to loss of personal data, and that they took too much time.

Amongst participants with negative experience were also some who did not have any expectations from updates, but were happy if they did not impede the functionality: "if it still works afterwards, then it's fine" (P96). In the qualitative data we found three different strategies that participants used to evaluate the importance of updates:

1. *By expected changes.* Expected changes can increase or decrease an update's perceived importance. Device maintenance, new features, and security increased the perceived importance, except in cases of minor bug fixes: "probably just some bug fixes" (P74).

2. *By the presentation and content of the update notification.* Some participants scrutinized update notifications to understand the updates' importance. Participants concluded that notifications without information are not important: "the green color is a sign that it [the update] is not important" (P20) or "it did not appear to be important" (P80).

3. *By principle.* Often, participants used a general principle such as "software updates are always important" (P66) to guide their update-decisions. However, participants sometimes based their principles on the type of device, e.g., smart consumer device updates were not important, and smartphone system updates were important.

Table 6.1.: Significant differences in importance between update notifications

Comparison	Mean Diff.	Sign.	
System & Apps	1.29	$< .001$	***
System & Dishwasher	1.41	$< .001$	***
System & Shoes	1.79	$< .001$	***
System & Car	0.59	$< .001$	***
Apps & Shoes	0.51	.02	*
Apps & Car	-0.69	$< .001$	***
Dishwasher & Car	-0.81	$< .001$	***
Shoes & Car	-1.20	$< .001$	***

Sign. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 6.2.: Participants' preferred update timing.

	At Notification	**Later**	**Never**	**No Opinion**
Phone (System)	35 (38.%5)	52 (57.1%)	4 (4.4%)	-
Car (System)	40 (44.0%)	47 (51.6%)	4 (4.4%)	-
Phone (App)	32 (35.2%)	49 (53.8%)	10 (11.0%)	-
Dishwasher (System)	39 (42.8%)	24 (26.4%)	19 (20.9%)	9 (9.9%)
Shoe (System)	-	-	37 (40.7%)	54 (59.3%)

Many participants could not imagine what a smart consumer device update would change, e.g., "I can not imagine what advantages an updated dishwasher could offer" (P35). Hence, evaluating *by expected changes* might not work well with smart consumer devices. *Evaluating by notification* could work in case update notifications provide the necessary information. However, the only approach that transfers well to smart consumer devices is the last one, *by principle*. Participants applied this approach to smartphones and other smart consumer devices alike.

Preparation. Answers from the *Preparation* stage mainly concerned the update procedures' timing: delay updates in general, inconvenient update time, waiting for specific resources (power or WiFi access), or create backups before update. Participants commonly waited until bed time to install updates: "I prefer updating just before bedtime. Since I don't need a smartphone during that time." (P23).

Deployment. For the *Deployment* stage, participants wanted to decide the updates' installation time and demanded detailed information in notifications. As P62 put it: "I like having the option to decide for myself when something will be installed". More users preferred to postpone smartphone and car updates, although between 35% and 44% would update right away. The majority of participants would perform dishwasher updates right away, even though they did not regard it as especially important (ranked

fourth in Figure 6.1). Participants either had no opinion on the preferred time of installation or would like to skip installing the self-lacing shoes update altogether, suggesting that users do not see any benefit of updating self-lacing shoes. Table 6.2 shows the preferred time to perform updates.

In some cases participants did not care about small and unimportant changes and wanted to install them automatically, while still keeping the agency for important updates. In contrast, other participants preferred automatic installation of important updates: "Special updates should be installed automatically" (P65).

Comparing update behavior for smartphones and smart consumer devices. While participants focused on security benefits of smartphone updates (by principle), participants did not consider smart consumer device updates important by principle, probably because they did not see the point of them. Compared to smartphone updates, they focused more on safety aspects and maintenance, e.g., "Ensures that the system runs correctly" (P69 regarding car software updates). Participants focused on potential security benefits and privacy-infringements with smartphone updates, which was not a concern with smart consumer device. Most participants did not expect any visible changes to smart consumer devices after updates. Commonly, participants preferred to install smartphones updates "Instantly if WiFi is available and the battery is sufficiently charged" (P5). Additionally, some participants install smartphone updates because they are curious about potential changes, which they did not do with smart consumer devices. Participants suspected that they could not use smart consumer devices during the update process, which is why they preferred to delay updates. Participants demanded similar changes to update notification for smartphone and smart consumer device updates. However, more participants did not have design suggestions for smart consumer device update notifications, probably because they did not deem updates necessary for these types of devices.

6.3.4. Contradicting User Requirements

During the evaluation, we uncovered the following five different contradictions of user requirements:

CR1: Installation time. Some participants thought updates that take a long time to install are important because they change a lot. Other participants delay these updates because they fear disrupting their regular activities. Resulting in a small conundrum: small, quickly installed, security patches may reduce the perceived importance to users – while bundling them in large updates keeps users vulnerable who defer them. This contradicting requirement is a problem for systems in immediate use such as cars or even self-lacing basketball shoes, while it is not an issue for asynchronously used smart consumer devices, e.g., dishwashers.

CR2: Amount of information. Some participants demanded detailed update notifications that explain its purpose and affected software parts. They carefully vet updates to avoid specific negative consequences. Others did not care about information, preferred influencing the installation time, or did not want any agency. Systems may accommodate all these user types by asking them about their policy preference and

adapting to their update behavior. Detailed information in update notifications is crucial for smart consumer devices since participants had difficulties understanding their purpose and effect.

CR3: UI and changes. A few participants disliked updates that changed UIs, they claimed that older UI versions worked better and did not confuse them. Others enthusiastically looked forward to using new UIs. Hence, everyone demands information about UI changes, even though users' reaction can vary. This contradicting requirement only applies to smart devices with malleable user interaction, such as car's touchscreens or voice interfaces. It does not apply to smart consumer devices with fixed interaction, such as the basketball shoes (only two buttons) or dishwashers.

CR4: Time of notification. Participants could not agree on appropriate times for update notifications. Several factors influenced the appropriate installation time: (1) necessary resources (remaining battery life or internet access), (2) necessary preparations (reading the installation notes or creating a backup), and (3) when they are planning on using the device. While smartphones consider the first point, the second and third are highly context-dependent or specific to the users' update attitudes. Smart consumer device users are concerned with (3) since they want to immediately use their device (such as the basketball shoes or the car) – for these devices notification should arrive at the end of a usage session or offer to delay the installation accordingly.

CR5: Automating updates. For several participants it was important to control updates for applications they considered important, but would even welcome automatic updates for all other applications. Other participants' approach was exactly opposite, they wanted to automatically install important updates, because they felt their decision was not necessary or beneficial in those cases. While still maintaining control for update decisions that were not critical. This contradicting user requirement applies to all IoT devices and smartphones. This issue warrants closer inspection in future work to see if those are actually opposite requirements or if participants thought about different levels of importance. Different levels of importance would result in three categories: (1) critical: automatic updates, (2) important for personal use: manual update decisions, and (3) others: automatic updates.

Interestingly, participants reported being annoyed by manual and automatic updates. Some said that update notifications requiring their decision annoyed them, which they resolved by enabling automatic updates. Others felt that updates slowed down the system or reduced the available download speed, which they resolved by disabling them. Some of those contradictions result from a fixed security policy and could be remedied by dynamic policies that are adaptable to the individual user, as suggested by Edwards et al. [70].

6.4. Discussion

Like all other study designs, this work and its results come with limitations. The results from our formative field study have an age bias (Table D.1), our online survey' participants felt more comfortable with technology than the average population (Table D.4), and both datasets have a gender bias to men. However, Amazon MTurk is more representative of the U.S. than the census-representative panel responses [238]. In

the foreseeable future, the average (target) users of smart consumer and IoT devices will be older than today. Hence, more research on the security of smart consumer devices with an older population will be necessary.

Given the nature of an online survey, we collected self-reported data about update notifications that participants did not experience on their own devices. However, we were primarily interested in the participants' update thought-process, which we could not have researched without self-reported data, even if participants experienced a real update situation.

6.4.1. Automatic Update Settings

The push for automatic updates by default has been effective at increasing the amount of users that keep automatic updates enabled. Previous work by Tian et al. [269] concluded that 47.7% updated their apps automatically, which has increased to 86.8% according to our results. We observed that iOS users more commonly deactivated automatic updates than Android users: 33% of them disabled automatic system updates (compared to 18% of Android users) and 16% of them disabled automatic app updates (compared to 14% of Android users). We assume that the reason for this difference is grounded in the UI: on iOS, the options to deactivate updates are in the general settings menu, whereas they are harder to find on Android. Prior to our work, we assumed that users change their update settings at most once. However, four (7.6%) participants of the formative field study stated that they had changed their update settings multiple times, indicating that the available options do not fit the participants' needs. For those users a more dynamic, context-sensitive security policy might be important [70].

We analyzed the participants' answers according to their update settings to find possible effects of those settings on the remaining manual update decisions. Participants with automatic update settings more commonly referenced concepts such as maintenance, security, and the invisibility of software-changes. We assume the reason for this difference is that users who think of the necessary but invisible changes included in updates are generally more comfortable with the idea of automatic updates. Additionally, we found that only users with manual updates wait for experience reports from other users before updating themselves. A possible explanation for this difference is that users with negative update experiences in the past are more risk-averse when installing updates. This would also explain why they deactivate automatic updates in the first place.

6.4.2. Transferring Update Behavior

Not all IoT device updates are automatable and some of them have minimalistic UIs, so we have to know how users will handle update decisions. Prior work [280] and our formative field study identify an update's perceived importance as a decision factor. Participants ranked the importance of the five update notifications as follows: operating system updates, car, apps, dishwasher, shoes. Indicating that users might think IoT device updates are less important than other kinds of updates, except for safety-relevant IoT devices.

In our qualitative data, we found three different approaches to evaluate the importance of updates: by the expected changes, by the presentation or content of the update

notification, or by principle. We discovered that participants in our study could often not imagine what kind of changes updates for IoT devices might entail. However, some users judge the importance of an update by evaluating the expected changes, impacting their install decision.

One of the root causes for this could be the analogical transfer of update behavior based on the term 'update'. An 'update' often implies new and improved software, which either scares or excites users. In recent years, we saw how major updates with invisible changes are bundled with minor visible changes, such as dark mode[1], or a new set of emojis, to communicate the update's importance. Analogical transfer of update decisions from smartphones to IoT devices may cause similar expectations. Son et al. [258] discussed how words influence the analogical transfer of concepts. Renaming 'updates' for IoT devices may avoid expectations of visible changes. Participants often mentioned 'maintenance', e.g. P43 "So that I can use the device without problems". We suggest this term for IoT updates that (1) do not contain visible changes and (2) cannot lead to loss of data. We do not recommend a name change for other updates to avoid undermining users' trust. Separating the terms "updates" and "maintenance" could eliminate unwarranted expectations of visible changes and reduce the fear of unexpected functionality or user interface changes.

More than half of the participants would have delayed the updates for the smartphone or the car which is in stark contrast to the update for the dishwasher or shoes. We think that this is a sign of risk-aversion, since the participants heavily rely on the functionality of those devices. While this study focused on smart consumer devices used by individuals, there is also communal use of smart consumer and IoT devices. While we expect increased risk-aversion in these cases, future research would be valuable to get a more complete picture of users' update preferences. The most popular option for the dishwasher update was to install it at the time of the notification, presumably as the distraction from the main task was perceived as less severe and participants had no issues with postponing an unattended task. Regarding the updates for shoes, participants either had no opinion about their preferred time of installation or did not want to install them at all. We interpret this as a sign that participants did not see the point of self-lacing shoes in general and did not want to maintain them in a working condition.

Comparing participants' perspective on updating smartphones and IoT devices also warrants a discussion about differences between devices and applications: (1) if the device has a fixed user interface or a reconfigurable one, (2) if the device is for a single purpose or for multiple purposes, (3) the type and amount of available resources such as Internet connection and power supply, and (4) how frequently people use them in everyday life. Comparing along these categories suggests that smartphones and IoT devices have different usage patterns - an exception being multi-purpose IoT devices with a malleable user interface such as voice assistants. If we instead compare specific IoT and smartphone applications it makes sense to classify according to user-centered themes from the qualitative analysis: urgency of use, importance of continued functionality, importance of specific feature-set, and importance of personal data associated with application. Especially the first two themes are important factors for smartphones as

[1] Dark mode changes the UI to a darker color palette to reduce strain on the eyes in low ambient light.

well as IoT devices and they will shape users' update decisions. However, we should not overestimate the usage patterns specific to applications or devices, since decisions for new devices or applications are often based on prior experience with other applications [281].

User interaction design is a tool for communication between users and the underlying technology. It should take the users' mental models into account and translate them as well as possible to corresponding mechanisms. Our findings can serve as a basis to understand user-specific constraints on update procedures. This gives several design indications which we will present in the next section. The technical goal of broadly deployed updates for security and maintenance purposes does not seem far-fetched and does not necessarily contradict users' values.

6.4.3. Implications for Design

Users consider several types of information before installing updates: how important they perceive the update, if they update interferes with their current primary task, and if they can live with the expected changes. How users perceive those factors can be influenced to some degree by design mechanisms. In the following, we provide a series of design implications based on the open-ended questions in our online survey to lay foundations for future work. In future work, we plan to expand and validate these recommendations.

ID1: Store information about users' software or IoT device usage and use this data to adapt update procedures to them. A common sentiment among participants was that they only consider apps that they frequently use as important and worthy of updates. This allows auto-updating IoT devices (if possible) or smartphones apps that user do not consider important without infringing on their sense of control.

ID2: Reduce the amount of update notifications as much as possible. Participants considered frequently occurring updates as not interesting and unimportant. In contrast, they perceive rare updates as special and probably important enough to warrant their attention. This applies to IoT devices and smartphones equally.

ID3: Important updates should take longer to install than unimportant ones. Participants perceived large updates that take longer to install as more important than quickly installed updates. Hence, the duration of the installation should reflect the update's importance. In most cases, developers should consider an update important if it reflects the users' values of important updates (this requires some feedback from individual users). However, systems should be able to (if possible) install critical updates that do not impact user experience without user-interaction. This applies equally to IoT devices and smartphones. However, immediate use is important for some types of IoT devices (e.g., the car, shoes, TVs, ...) the timing of these longer updates is critical.

ID4: Restrict install options for important updates to convey importance. The interface options available in the update notification also communicate how important an update is. As one participant phrased it: "Since I can delay the update, it is apparently not an important update" (P53). This type of modifications has even more effect on IoT devices with immediate use requirements, since an update has to be very important to force active waiting until the update is finished. Other devices, such as a dishwasher or apps that are not used often, will not be as affected as much by such a

design change.

ID5: Clearly communicate possible consequences of an update. The fear of data loss made our participants delay an update until they create a backup of their data. Informing users if their personal data will be affected by the update and creating automatic backups could reduce the users' fear of updating. Participants were worried that an update could take longer than expected and prevent them from completing their primary task. Therefore, it is important that the update notification conveys these aspects ahead of time. This is important for software and devices that users depend on for regular activities that cannot be arbitrarily delayed: software required for work related tasks, and mobility devices (cars, shoes).

ID6: Provide context-dependent options to delay installation time. One of the major suggestions of improvement for update notifications was that users want more agency to select the time of installation. Some participants proposed a "later" button, some wanted to select a certain time, and one suggested an option "install after current task". We suggest decoupling the decision time from installation time in a context-sensitive way to provide a user-centered installation time. All software and devices that users immediately require and that are task-centered would benefit from such an option. Distinguishing tasks might be easier for IoT devices even (as in all our presented IoT devices), since they are often only used for a single purpose.

ID7: Changes of the User Interface should remain optional wherever possible. Updated user interfaces were considered unimportant by most participants. Some considered UI changes a burden, others thought they had the potential to make them feel as if they had gotten a new device. Since UI updates could be a barrier to updating, those should be separated from the rest and remain optional for users. This design would probably not affect IoT devices as much, because many of them do not have a malleable user interface in the first place. However, this could be a necessary option for devices that are controlled by a touch screen or voice.

ID8: Let users decide if software that they consider important should update automatically. Some participants were annoyed by the amount of updates that they considered unimportant: they wanted to have these automated but still manually update apps they consider important. Other participants thought it did not make sense for them to be able to decline important updates, instead they were willing to decide upon less important updates. Especially for smartphones a choice like this could severely improve the amount of update notifications that users see, while increasing the relevance of these notifications. This distinctions is less important for IoT devices, because they receive a smaller amount of updates in general.

6.5. Related Work

Reasons for (not) updating. Vaniea et al. [280] found that participants who always installed updates or believed in an update's importance readily installed updates, whereas participants who were satisfied with current versions delayed updates. Satisfaction with the current software version, undesired UI changes, the perceived lack of purpose of software updates, and negative prior update experiences hinder participants from updating [281]. Mathur et al. [183] found that 40.5% of participants thought about the

costs of updating, 29.2% considered the necessity of updates before installing, and 7.5% were concerned about the potential risk of updating. We extend Mathur et al.'s work with a focus on smartphones and smart consumer devices.

Users' main source of information about updates is the notification that they see. Users often misunderstand how updates change their system, which frustrates them and about 16% of them refuse to apply updates [77]. Tian et al. [269] found that 42.6% participants regretted updating a smartphone app in the past because of bugs, "bad" UIs, and privacy-invasive practices. Since participants relied on reviews for their update-decisions, the authors introduced a review-based support system. Mathur et al.'s [182] formative study found that users want know about an update's purpose and that trust in vendors, expected compatibility issues, user interface changes, social influences, and installation time affected update decisions. They built a prototype of a corresponding OS update process which satisfied half of the participants because it decreased interruptions.

Updating IoT devices. Fernandes et al. [80] found that over 55% of existing SmartApps on SmartThings are over-privileged and have inadequate security controls. Zeng et al. [304] used an exploratory design study to understand users' requirements for access control in multi-user smart home designs. In 2006, Bellissimo et al. [26] found that secure updates for IoT devices face challenges such as untrusted infrastructure, sporadic network connectivity, or limited local resources. Simpson et al. [256] discuss usability challenges of applying updates on IoT devices, specifically update notification and predicting convenient update times.

6.6. Conclusion and Future Work

Among other things, we found that the prevalence of automatic updates for applications on mobile phones has increased to 86.8% (in comparison to 47.7% [269]) and that 18.7% of participants deactivated automatic system updates. Our results suggest that iOS users deactivate automatic updates more often than Android users. We hypothesize that easy access to the relevant option in the UI explains most of that difference (see Section D.1). Users explained their deactivated automatic updates with a fear that updates might introduce flaws and agency in update decisions, fear of compatibility issues, and a limited or expensive data plan. The most important reasons to activate automatic updates were staying up to date, convenience, and security. We expected to find evidence of avoidance behavior amongst participants (i.e., avoiding charging their phone while connected to WiFi), but our results do not support this. Participants who enabled automatic updates approached update decisions similar to those with manual updates. However, three concepts were more important to them: the idea that updates are necessary for maintenance, for security, and that updates could be important even if they do not have any visible effects. Additionally, participants who favored automatic updates were not interested in other users' experience reports.

Prior work [280] and our formative field study provide evidence that the perceived importance of an update is a decisive factor for installing it. Our results indicate that users perceive updates for smart consumer devices as less important than regular

updates, except for safety-relevant devices. Our contribution includes a classification of how users evaluate the importance of updates: by expected changes, by the presentation and content of the notification, and by principle. Participants in our study could not imagine meaningful changes for smart consumer devices; the corresponding notification lacked information. Therefore, the evaluation by principle is the only method that led participants to conclude that updates for these devices are important and that they would install them soon or immediately after receiving the notification. Prior work [258] indicates that a concept's name promotes analogical transfer: An 'update' might imply new features or at least focus on visible changes. However, in the case of IoT device updates, participants mentioned the concept of 'maintenance' more often. We hypothesize that using the word 'maintenance' to describe updates without visible changes might increase users' willingness to install them. We provide a list of areas of tension based on conflicting motivations and themes from our data. These open up new directions for designing update solutions that work well for everyone.

At the workshop, participants discussed areas of future research with us. The first idea was identifying and developing a fine-granular terminology to describe the exact nature of updates. Using such terminology, developers could easily communicate updates' effects to end users – simplifying their update decision. The second idea concerned how the companies who create software updates manage this process. Understanding the rationale for changing user interfaces, deprecating features, packaging update bundles, and automating update decisions may help in improving end users' update experiences.

The contents of the following chapter were published as part of the publication *"Why I Can't Authenticate — Understanding the Low Adoption of Authentication Ceremonies with Autoethnography"* (CHI 2023) [P5]. This paper was produced in cooperation with my co-author Katharina Krombholz. The following table describes the contributions of all authors to this paper:

Author	**Contribution**
Matthias Fassl	I had the initial idea for this research project, including the methodological approach. I conducted the 5-month diary study, analyzed the results, and wrote the paper.
Katharina Krombholz	Katharina gave me feedback on the initial idea and the methodological approach, reviewed my analysis of my diary entries, and helped me find an appropriate venue for this research project. She also gave feedback on draft versions, edited parts of the paper, and guided me during the reviewing process.

Reference

Fassl, M. & Krombholz, K. (2023). Why I Can't Authenticate — Understanding the Low Adoption of Authentication Ceremonies with Autoethnography. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), 1–15. `https://doi.org/10.1145/3544548.3581508`

7

Understanding the Low Adoption of Authentication Ceremonies with Autoethnography

Figure 7.1.: Two people authenticating their secure messaging conversation

Authentication ceremonies detect and mitigate Man-in-the-Middle (MitM) attacks on end-to-end encrypted messengers, such as Signal, WhatsApp, or Threema. However, prior work found that adoption remains low as non-expert users have difficulties using them correctly. Anecdotal evidence suggests that security researchers also have trouble authenticating others. Since their issues are probably unrelated to user comprehension or usability, the root causes may lie deeper.

This work explores these root causes using autoethnography. The first author kept a five-month research diary of their experience with authentication ceremonies. The results uncover points of failure while planning and conducting authentication ceremonies. They include cognitive load, forgetfulness, social awkwardness, and explanations required by a communication partner. Additionally, this work identifies and discusses how sociocultural aspects affect authentication ceremonies. Lastly, this work discusses a design approach for cooperative security that employs cultural transcoding to improve sociocultural aspects of security by design.

7.1. Introduction

Successful MitM attackers are able to read and also fake messages in end-to-end encrypted instant messaging conversations while their active attack is ongoing. Authentication ceremonies, which require verifying cryptographic keys with conversation partners, detect and mitigate these attacks. However, users have several issues with these ceremonies: They are unaware of them, do not understand their purpose, and have trouble finding and conducting them [139]. But not only end users have trouble with authentication ceremonies. Anecdotal evidence suggests that even security experts, such as we researchers, have problems keeping contacts authenticated – even though this is our area of expertise! Hence, we suspect that authentication ceremonies suffer from deeper problems than merely a lack of UI usability or user comprehension.

Previous work on authentication ceremonies focused on user behavior in lab settings [137, 284, 283, 285], some of which simulated MitM attacks [254, 298]. Lab settings are great environments to study usability issues in a controlled way. And studying user behavior during attacks is crucial since users require protection in these exact moments. However, real-world conditions for authentication are more complex. First, users are likely not in proximity when they recognize the need to authenticate or notice key-reset notifications. Second, since key resets usually happen for benign reasons, users will not necessarily be in a hurry to conduct a new authentication ceremony. Hence, users will most likely plan a future authentication ceremony with their contact. Authentication ceremonies are social cybersecurity [299] mechanisms, which depend on conversation partners to meet up and cooperate to improve their security equally. As such, social and cultural expectations and practices affect their success, as Uzun et al. [279] suggested in their work.

RQ1: What problems do knowledgeable and motivated users encounter when planning and conducting authentication ceremonies?

RQ2: How do social and cultural factors impact authentication ceremonies between conversation partners?

To understand the potential issues around planning authentication ceremonies and navigating social and cultural issues, we conduct an autoethnography based on a five-month research diary documenting the first author's authentication experience. While autoethnographic approaches are uncommon in the field of Usable Privacy and Security (UPS) — Turner et al.'s work [276] being a rare example — researchers of the closely related HCI field regularly use it to gain a deeper understanding of how technology affects users' lives [38, 142, 148, 169, 170, 206, 261, 263]. In contrast to other types of diary studies, self-inquiries can easily be long-term, are deeply introspective, and combine the analysis step with data collection – adapting the data collection method according to preliminary analysis results. Autoethnographies focus their analysis on social interactions and the cultural rules that govern them [40]. Since security experts rarely report their own failures to cope with security systematically, this work may provide valuable insights as an autoethnography. While strict generalizability is not a meaningful goal for this kind of qualitative research [35], we improve the transferability

of our results by describing the first author's background and all the situations in as much detail as possible.

The first author aims to authenticate as many contacts as possible in naturalistic settings for this study. Prior work identified the issues of lacking user comprehension and lacking usability of the user interfaces [139]. However, since the first author is a security researcher who understands the underlying issues and knows how to conduct authentication ceremonies correctly, these potential barriers should not apply. Autoethnography enables the study of authentication ceremonies in infrequent and naturalistic contexts, which are not easily accessible with other study designs or participants.

This paper contributes: (1) a phase model of planned authentication ceremonies, comprising need recognition, planning, meeting, convincing, and authenticating; (2) the identification of failure points in the planning process, such as forgetting, social awkwardness, and necessary explanations; (3) an account of subjective emotions of guilt and frustration that constituted the first author's authentication experience; and (4) a discussion about sociocultural barriers and facilitators of authentication.

7.2. Background and Related Work

First, we provide some background on the security benefits and process of authentication ceremonies in end-to-end encrypted messengers. Then, this section presents the theoretical framework of autoethnography, data collection methods, and work from the related HCI field. Afterward, this section briefly introduces the concept of social cybersecurity, explaining why autoethnography is a suitable approach to research this area. Finally, this section elaborates the research on authentication ceremonies, discussing prior work's chosen methods and the identified issues.

7.2.1. Authentication Ceremonies in End-to-End Encrypted Messengers

To communicate with a conversation partner, end-to-end-encrypted messengers need to know the recipient's public key. In most modern messengers, this public key comes from a central key server and is trusted without further verification. MitM attackers impersonate this key server, either by directly taking control of the key server or with an active attack at the network level. With the powers of the key server, MitM attackers can convince the clients of both conversation partners to use a different encryption key (which is in the attacker's possession) for communicating (see Figure 7.2). As a consequence, successful MitM attackers can read and manipulate all messages in a conversation as long as the active attack continues. In 2018, the Dutch police appears to have used such an attack to intercept messages sent via IronChat [108]. To detect and mitigate these MitM attacks, end users need to authenticate the key material, i.e., verify that they are using the correct encryption key to communicate with their conversation partner. Most end-to-end encrypted messengers offer a dedicated authentication ceremony for this purpose.

There are many different versions of authentication ceremonies, most are intended for in-person authentication (e.g., in Signal, WhatsApp, or Telegram's secret chats) but some are also designed for remote use. During calls in Viber, conversation partners

compare a secret identification key; in Telegram calls, conversation partners compare a set of emojis; and for the Wire messenger, users record a video of themselves announcing the short authentication strings (SAS) that correspond to their key. However, meeting in person is usually preferable because (a) it is a secure out-of-band communication channel and (b) scanning others' QR codes is simpler and less exhausting than remotely reading and comparing safety numbers [284]. Here, we describe the steps to authenticate a conversation in Signal and WhatsApp in person with QR codes. Figure 7.3 shows the corresponding user interface to each of the steps.

1. Both conversation partners open up the shared conversation in their messenger.
2. They open the authentication interface by clicking on their contact's name and selecting "View Safety Number".
3. The conversation partners take turns showing their QR code and letting the others scan it. Users scan by tapping the QR code or pressing the corresponding button, depending on the operating system and messenger.
4. If the scan is successful, the conversation can be marked as verified. If the check fails, the conversation is under attack and must not be used for communication.
5. The verified marker vanishes when either end reinstalls the messenger, one of the conversation partners gets a new phone, or when the conversation is under attack. Repeating the authentication ceremony is required in all cases.

If users are aware of authentication ceremonies, they will need to decide which conversation they want to protect since this will require expending additional effort to avoid MitM attacks. In general, even unauthenticated conversations are still protected from various passive eavesdropping attacks and are more secure than regular email or SMS conversations.

Fassl et al. [P6] found that end users mostly considered authenticating messenger conversations with friends, partners, and family members. However, the perceived overhead of arranging meetings might have impacted participants' responses. Also, other types of contacts, such as tax advisors, lawyers, or business partners, might be sensitive conversation partners outside the immediate circle of friends, partners, and family members.

7.2.2. Autoethnography

According to Chang [40], autoethnography is a self-reflective form of cultural analysis. The term describes the method and final product alike. It assumes that the self learns and upholds values, norms, and customs to become part of a cultural group. Thus, we can learn about cultural factors by understanding the relationship between the self and others. Self-narrative reports are the basis for autoethnographies. Additional explanations connect these personal experiences to the cultural environment. According to Ellis and Bochner [72], autoethnographies comprise three parts: research process (graphy), culture (ethno), and the self (auto). However, the focus on these parts varies widely among researchers. Some emphasize personal experience, while others give

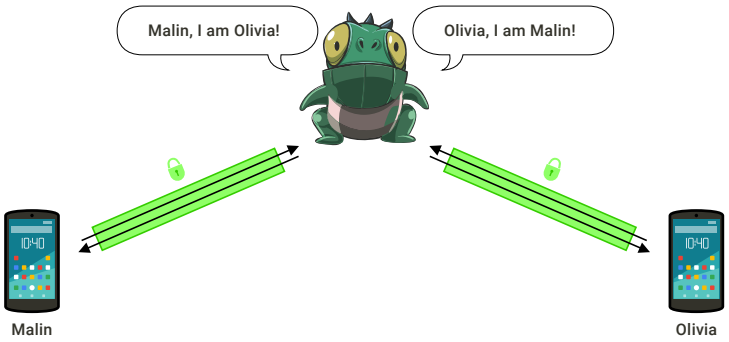
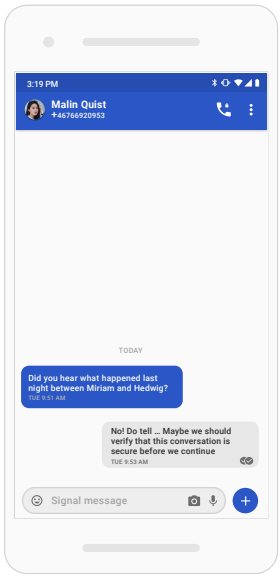


Figure 7.2.: The MitM attacker (the monster) convinces the messenger clients (of Malin and Olivia) that they should use their encryption key instead of their conversation partner's. During an ongoing attack, the monster will have to continuously forward Olivia and Malin's messages.

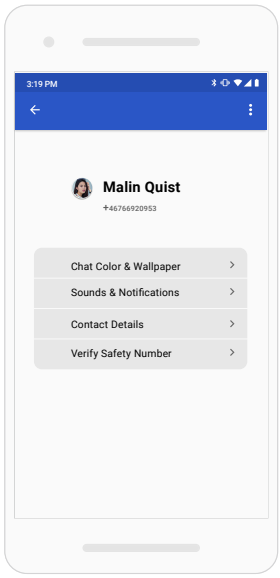
more space to cultural explanations. For researchers, autoethnography is a practical approach for investigating human relations in their cultural context. It improves the cultural understanding of the self and others while invoking self-reflection and self-examination among its readers [40]. According to Chang [40], autoethnographers collect data by chronicling their past, i.e., memories, and recording field data. Usually, they avoid mixing these two. Recorded field data includes routines, rituals, celebrations, or cultural artifacts. The focus lies on experiences, stories, and objects with sociocultural significance. In parallel to data collection, autoethnographers use inventorying to evaluate and organize their data, i.e., they select, prioritize, rank, and categorize collected data.

Autoethnography, narrative inquiries, and self-studies all privilege the self in the research design. According to Hamilton [125], narrative inquiry shares stories of the researchers' experiences so that others may learn from them. Self-studies are systematic inquires into the researchers' own practices to gain knowledge about them and improve them. In contrast, autoethnography highlights the changing perspectives of the self and puts experiences in their broader cultural context. Commonly, researchers write in the first person when they use these approaches.

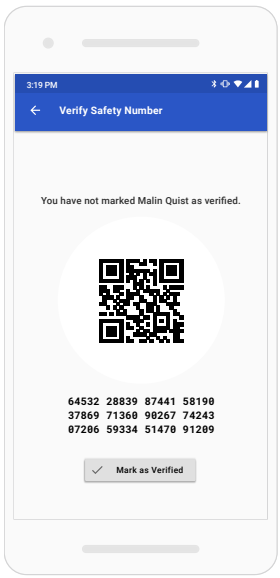
Autoethnography is a rarely used method in Usable Privacy and Security (UPS) research, e.g., Turner et al. [276]'s work on smart home cyber security practices in their family. However, researchers in Human-Computer Interaction (HCI) regularly apply autoethnography to study how technology affects users. Spiel [261] used it to demonstrate how technical infrastructures reinforce binary gender ideology by documenting their experience with systems not allowing them to register their gender correctly. Jain et al. [148] used it to highlight tensions and nuances during the travels of a hard-of-hearing individual, focusing on difficult social conversations, navigation problems, and personal assistive technologies. Stephen et al. [263] provide an autoethnographic account of the barriers an independent blind traveler faced during her 28-day cruise, focusing on the use of technology to access visual information and maps. Chamberlain et al. [38] explore autoethnography as a method to self-design personal heritage soundscapes.



(a) Open the messenger conversation.

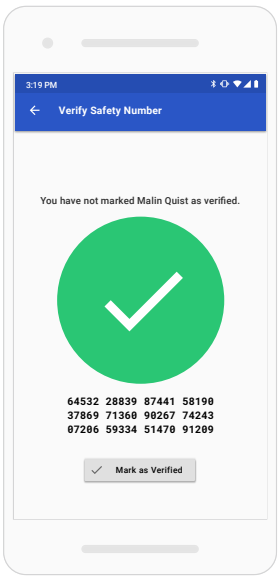


(b) Select "Verify Safety Number" in the conversation menu.



(c) Show QR code to conversation partner.

(d) Scan conversation partner's QR code.



(e) If check is successful, mark conversation as verified.

Figure 7.3.: The step-by-step procedure to authenticate conversations in Signal and WhatsApp. The procedure is similar in other messengers that offer an authentication ceremony.

Lockton et al. [169] designed a series of research probes that enable autoethnographic exploration, investigating students' bedtime routines, sleep patterns, personal time scheduling strategies, and sleep in non-traditional places. Lucero [170] conducted an autoethnography to understand how living without a mobile phone affects their life. O'Kane et al. [206] used autoethnography to evaluate a wrist blood pressure monitor. They argue that it enabled them to empathize with users in contexts that are otherwise hard to investigate using traditional in-situ studies.

Assessing Methodological Fit

We assessed the methodological fit of autoethnography to our research problem along three criteria: The value that the first author's vantage point adds to answering the research question, the required level of introspection and reflection, and the appropriateness of a socio-cultural analysis lens for the research questions.

1. *Vantage point:* The first author's vantage point is primarily that of a security expert and offers two benefits: They can collect more data on all parts of the authentication process since their background knowledge of authentication ceremonies helps them complete ceremonies even when others can not. Similarly, Turner et al. [276] thought their increased cybersecurity awareness as experts may have generated more data for their autoethnographic work on smart home security practices. Also, with their experience in Usable Privacy and Security, the first author can provide explanations at different levels of complexity to engage and educate people with varying technical know-how.

2. *Introspection:* Self-studies provide a unique ability to researchers: introspection of the feelings and thought processes behind one's actions and adapting and refining the used research methods based on them. These abilities are useful for studying authentication ceremonies because ceremonies incorporate social interactions that come with social expectations, boundaries, feelings, and context-dependent norms. Direct access to the experience and sufficient time for reflecting on them makes an in-depth analysis easier for researchers. Understanding these aspects with a diary study with multiple participants can be more difficult. Participants know that researchers will read the diary entries, which limits the amount and type of information in them. To get a comparable level of analysis to an autoethnography, researchers would need to question participants repeatedly about each (potentially embarrassing) experience during the ongoing analysis process — resulting in a resource-intensive hybrid between diary and interview study.

3. *Sociocultural analysis:* Authentication ceremonies require that two communication partners cooperate to increase their security. Thus, they are a sociotechnical system that researchers can analyze through a sociocultural lens. While the sociocultural analysis of ethnographic approaches is not strictly necessary, it does provide a lens to understand the underlying social and technical issues of using authentication ceremonies more completely.

While we could use a regular diary study or ethnography to learn more about others' difficulties with authentication ceremonies in their daily life, an autoethnography gives us a more in-depth understanding of one person's experience.

7.2.3. Social Cybersecurity

Social cybersecurity research recognizes that security mechanisms are often social in nature [299]. However, security tools are still built primarily with an individual user in mind, resulting in a social-technical gap [4] between technology's abilities and what it requires socially from users. Exploring this social-technical gap works well with autoethnography because it connects individuals' security-tool experience with its sociocultural context. Also, studying social cybersecurity tool usage is challenging because of the naturalistic settings and infrequent use. And as O'Kane observed in prior work on non-routine use of technology [206], autoethnography is valuable to study usage under these constraints.

Wu et al. [299] structure the research on social cybersecurity into four categories: negotiating access to shared resources, shared and social authentication, managing self-representation, and influencing others' security and privacy behaviors. While not explicitly mentioned, authentication ceremonies in secure messengers are shared and social authentication mechanisms: Two messenger users cooperate to detect and mitigate MitM attacks against their conversation.

Prior work extensively studied the usability issues of authentication ceremonies in a lab. Herzberg et al. [137] studied the usability of WhatsApp, Viber, Telegram, and Signal. They found that participants were unaware of the need to authenticate and that 56.5% of them could not do so after being asked. Schröder et al. [254] found that the majority of participating CS students failed to detect and mitigate MitM attacks using Signal's authentication ceremony. Vaziripour et al. [284] compared authentication ceremonies of Viber, WhatsApp, and Facebook Messenger. They found that only 14% of their study participants successfully verified the key material without further explanation. When Vaziripour et al. [282] surveyed Iranian Telegram users, they found that only 29.6% had ever used the authentication ceremony in text conversations.

Researchers proposed several usability improvements to authentication ceremonies. Vaziripour et al. [285] streamlined Signal's ceremony by providing easy access and additional guidance. Wu et al. [298] produced new visual indicators, new notification dialogs, and a simplified notification flow. Vaziripour et al. [283] partially removed users from the loop by using Keybase for an authentication method based on social media. Fassl et al. [P6] redesigned authentication ceremonies from the ground up and found that user-centered prototypes can increase users' comprehension of the security implications. These design changes that prior work suggested may solve some of the problems associated with authentication ceremonies. However, since their implementations are not (yet) widespread, it is hard to understand their effects on the users' daily life, i.e., in infrequent and naturalistic scenarios.

In their paper "Secure Messaging Authentication Ceremonies are Broken", Herzberg et al. [139] summarize the current state of research: Users do not understand encryption, cryptographic keys, and the concept of MitM attacks. When facing a MitM attack

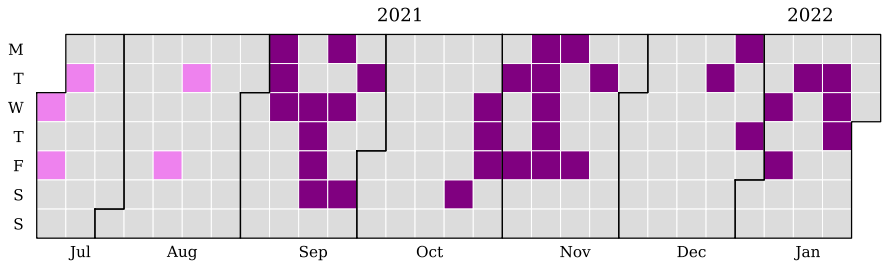


Figure 7.4.: Calendar overview of diary entries. Dark purple days mark regular diary entries, while light pink days mark memories added at a later time.

in a lab environment, they have significant issues finding and completing ceremonies successfully. In contrast, this work applies autoethnography to identify how sociocultural factors unrelated to either comprehension or usability affect the use of authentication ceremonies.

7.3. Methodology

Authentication ceremonies are social cybersecurity [299] mechanisms, which depend on conversation partners to meet up and improve their security equally. In this work, we try to understand how a user who fulfills all prerequisites, i.e., knowledgeable about security of messengers, experienced with the messenger interface, and sufficiently motivated, copes with authentication ceremonies in their daily life and how social and cultural factors impact these interactions.

Autoethnography, a self-reflective form of cultural analysis [40], is a powerful tool for understanding these ceremonies in-depth. We decided to use an autoethnographic approach based on three criteria: the usefulness of the first author's vantage point, the required level of introspection, and the necessity of sociocultural analysis. Section 7.2.2 contains the complete deliberation. In summary, we found that: (1) The *vantage point* as a usable security expert adds value to the analysis by collecting more data points and empathizing with people at different levels of skill and interest — instrumental to answering RQ1. In Turner et al.'s autoethnography [276], their cybersecurity expertise helped to reflect on their family's problems with security mechanisms in the smart home. (2) The *introspection* and reflection on the experienced authentication ceremonies help understand the scope of problems with authentication ceremonies and the depth of sociocultural issues. Achieving the same level of analysis with regular diary studies may be difficult. (3) The *sociocultural analysis* is necessary to answer RQ2.

Compared to lab studies, which use artificial scenarios and interactions, autoethnographies offer insights into more naturalistic experiences in the field. Long-term diaries are the appropriate basis for understanding these infrequently occurring authentication ceremonies. However, as discussed in Section 7.3.2, autoethnographies can not offer insights into entirely uninhibited natural interactions since the research question always influences the behavior to some extent when the researcher is also the participant.

In our systematic autoethnography approach, the first author's subjective experiences influence data collection, analysis, and results. The upcoming sections use the first

person to communicate this influence honestly.

7.3.1. Procedure and Analysis

The autoethnography's data source is a research diary covering the five-month study period. For these five months, my goal was to authenticate as many secure messaging conversations as possible with any secure messenger installed on my phone (Signal, WhatsApp, and Telegram). While I would probably not try to authenticate every messenger conversation outside of this study — because not every conversation is that important — I did not want to limit myself to a specific subset of contacts for this study. This way, I could understand how different types of contacts and their social-context impact the corresponding authentication ceremony. In the spirit of Glaser and Strauss [107], I did not want to limit my data collection with preconceived theories. Hence, I tried to document all my unfiltered thoughts on authentication ceremonies. These thoughts can come up at any time, so the study required the possibility of creating diary entries on any nearby personal device. Day One, a personal journaling app, fulfills this requirement. Initially, I followed a trigger-based approach to collect data, creating diary entries whenever thinking about authentication ceremonies and describing as many details as possible. Over time, I developed more specific documentation patterns, closing in on reoccurring phenomena. In general, diary entries focused on either planning an authentication, the authentication ceremonies themselves, memories of past events concerning authentication, and introspection or analysis. Additionally, entries described immediate surroundings and what triggered thoughts about authentication. Later, I decided to add a periodic diary reminder, forcing me to reflect on overlooked authentication opportunities. Section C.1 contains the guidelines I used for these diary entries. The length of all diary entries ranged from 30 to 400 words. However, most entries were between 50 and 90 words long. Sometimes I started shorter diary entries on the smartphone, only noting keywords, and expanded them later on my desktop computer.

Due to the nature of autoethnographic approaches, i.e., researching one's own experiences, data collection and analysis are parallel processes. I made this transparent by documenting ongoing analytic thoughts in the research diary. While these introspective moments provided some initial insights during the data collection phase, I also applied a structured approach to analyzing diary entries. To gather an overview of the concepts in the diary, I started with an open coding procedure on all entries with the help of qualitative data analysis software (ATLAS.ti 9). Based on this initial overview, I focused further analysis on planning phases, subjective emotional experience, and social and cultural aspects. For each diary entry, I consulted a Feeling Wheel [297] to help articulate corresponding emotions accurately. Autoethnography analyzes researchers' personal experiences in addition to the literal diary entries. Hence, calculating inter-rater reliability or discussing agreement among several researchers is methodologically not appropriate [189], since other researchers may read the diary but can hardly reflect on the experience. However, I shared the diary in confidence with co-authors to identify overlooked aspects. That they did not find any that may be explained by our similar cultural background and opinions on the research topic.

In November 2021, two months after beginning the study, I conducted a preliminary structured analysis. In the collected data, I identified groups of entries that repeatedly centered around different parts of the process for planning and conducting authentication ceremonies. Based on this observation, I modeled the phases of planned authentication ceremonies (see Figure 7.5). Using this model, I placed all diary events in their appropriate context. Additionally, I used an explicitly abductive approach to identify areas of interest that were up to this point missing from the collected data. For example, during the study, I often forgot about planned authentication ceremonies. Forgetfulness impacted my trigger-based data collection strategy because it relied on me thinking about authentication. Hence, I added a periodic reminder to my diary software to identify and document missed authentication opportunities. At the end of January 2021, I conducted my final structured analysis. Based on the preliminary analysis of the data entries, I focused on three areas of interest: (1) potential barriers to planned authentication ceremonies and their location in the phase model, (2) subjective emotions and insecurities during the process of planning and conducting authentication ceremonies, and (3) the social and cultural aspects that impact authentication ceremonies. Section C.2 of the Appendix includes the entire translated codebook.

7.3.2. Limitations

As with any methodological choice, autoethnographies come with limitations. The primary one is a lack of generality. Since all results stem from one person's experience, it is difficult to claim that they generalize. However, we claim that the results are transferable to other similar situations and people. See Section 9.5 for a detailed discussion of the concept of transferability. To improve the transferability of this work, we provided a detailed description of the first author's background in Section 7.3.4 and the circumstances and context of each interaction in as much detail as possible without compromising others' identities. Depending on the context, identified issues transfer to non-experts, albeit in a more severe manner. Put simply, when security mechanisms pose a problem to security experts we can not expect laypeople to do much better, regardless of user education and motivation. The lack of generality seems acceptable in order to surface issues that may affect users who are attentive to the need for authentication.

The second limitation is how the research affects the described behavior and experience. In an autoethnography, the researcher and participant are the same, making it hard to delineate naturally-occurring behavior from research-influenced behavior. However, to some extent, this is a common issue with many research approaches, e.g., the artifical scenarios in lab studies affect participant behavior and certain interview questions or environments will influence responses to some degree. To mitigate the effects of this issue, this work transparently communicates the ways in which the research approach may have influenced the reported results.

7.3.3. Overview of Collected Diary Data

From Sept. 3rd 2021 until Jan. 26nd 2022, I collected 69 entries in the research diary. In total, the diary contains 17 successful authentication attempts. While it is challenging

to define clearly when an authentication attempt failed, I identified seven missed opportunities for authentication. I authenticated conversation partners in Austria, Germany, and Japan – almost all of them in person. I authenticated two contacts during a video call, one currently living in Japan and a work colleague I rarely see in person. In neither case, the video quality was good enough to scan the QR codes. Instead, we compared the safety number verbally. Figure 7.4 shows the distribution of created diary entries during the study period, whereby some days have multiple entries. Entries before Sept. 3rd represent memories I added after the start of the study.

7.3.4. My Cultural, Educational, and Security Background

Researchers' positionality affects all their research. Autoethnography, which focuses on their experiences, amplifies that effect. This section provides context about my environment and prior experience so that readers may judge the results in the appropriate context.

I am a white European man who grew up in the suburbs of a large metropolitan area in central Europe. I attended a high school specializing in applied Information and Communication Technologies (ICT). During that time, I was fascinated by the local hackspace and its security and privacy-minded members, which later led to my first part-time job at an NGO focused on data protection and privacy.

My first exposure to authentication ceremonies was PGP key signing. At that time, my circle of friends haphazardly authenticated PGP keys even though we never encrypted any emails. Using PGP and authenticating PGP keys seemed like a way of signaling that we belonged to the security and privacy-aware community.

During my master's in computer engineering, I became more interested in the then up-and-coming secure mobile messengers. I explained their purpose and corresponding threat models to countless users while demonstrating how to conduct them. During my Ph.D., I designed alternative ceremonies in the hope of improving usability and user comprehension through improved design. During this research, I was intrigued because even security researchers in my surroundings rarely authenticated any of their messenger contacts.

Since authentication ceremonies are social in nature, it also makes sense to tell you about my usual social demeanor. While I am critical of the introversion-extroversion dichotomy, I consider myself more an Introvert. I carefully balance my need for time alone with my need for social interaction. Disliking large parties, I prefer meeting one or two people at a time for in-depth conversations about personal problems, wishes, and dreams. Meeting new people or people I do not know well can be a draining experience, mostly because I try to learn others' potential boundaries that I do not want to overstep so early after meeting.

7.3.5. Ethical Considerations

As this study collects subjective experiences with authentication ceremonies, it is not subject to our institution's ethical review process. However, studying experiences with authentication ceremonies necessarily involves other people – who can not meaningfully consent to their involvement in the study. I tackle this ethical issue by carefully

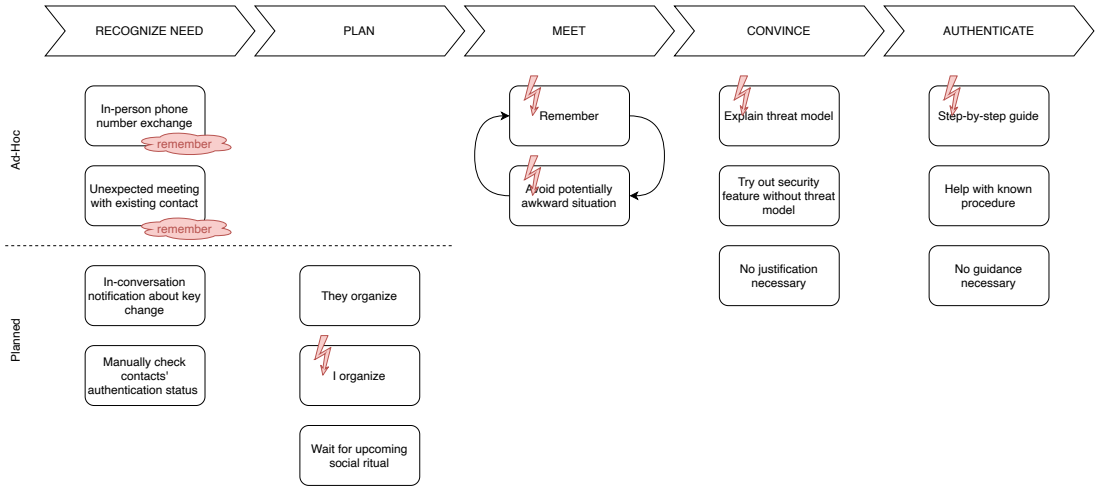


Figure 7.5.: Planning authentication ceremonies: a process overview. Red lightning marks potential barriers.

considering which details of the encounters to describe. While contacts may identify themselves in the descriptions, I avoid disclosing more information than necessary.

7.4. Results

I analyzed the resulting diary entries along three dimensions: (1) What process of planning and conducting authentication ceremonies looks like and what may go wrong; (2) Subjective emotions when planning and conducting authentication ceremonies, i.e., the authentication experience; and (3) The impact of the sociocultural environment on cooperative security mechanisms such as these.

Barriers during the Stages of Authentication Ceremonies

To give an overview of what the authentication process entails and also to embed my autobiographical reports in the correct context, I summarized the different phases of my experiences based on my collected diary entries in Figure 7.5. The following parts describe each of these phases in detail.

Recognizing Need

During the study, I depended on two different approaches to identify who to authenticate. The first and more intuitive approach is an ad-hoc in-the-moment authentication. When meeting and chatting with contacts, I sometimes remembered my wish to authenticate them. However, I could not depend on this method because often, I did not remember that wish in time. More than once, I remembered too late after the person was no longer within reach, e.g., when I sent them a follow-up message or when I wrote a diary entry thinking about missed authentication opportunities. However, as the study progressed, I habituated myself to think about authentication ceremonies when meeting people –

making this approach more practicable. Although it only happened once during the study, I found it easier to authenticate others during the initial exchange of contact details. After we both had each other's phone numbers, I merely had to ask if they used a messenger and if I could authenticate them.

When I recognized the need to authenticate and the contact was not within reach, I had to plan a meeting for the authentication ceremony. For previously authenticated contacts, some secure messengers (e.g., Signal) alert users about encryption key changes before sending a potentially compromised message. In the one instance this happened to me, I wrote my conversation partner that we should authenticate again without making any specific plans. It turns out they had recently gotten a new phone, which is a common reason for changed keys. The alert worked insofar that it reminded me to make specific plans with that contact a week later. Since I did not want to depend on these alerts for this study, I also used a more structured approach. I scoured my list of conversations and checked if I needed to authenticate them (again). Usually, I sorted these lists by the location of the contacts and how easy it would be to meet them in person.

It is difficult to define what constitutes a failure this early in authentication ceremonies. Hence, it is hard to count the number of failed authentication attempts. However, I found it challenging to identify the need for ad-hoc in-the-moment authentication in time. And while I identified the need for planned authentication pretty well, I had a difficult time keeping track of them – at no point was I sure about someone's authentication status without double-checking in the messenger application.

Planning Meetings

Recognizing the need to authenticate while concerned contacts are within reach simplifies the authentication process. When they are not around during this realization, the process becomes more difficult. Then, planning a meeting is necessary.

At first, I planned these meetings around who I needed to authenticate. I tried to remember who to authenticate, checked the authentication status of recent messenger conversations, and sometimes reread the diary. Of course, further factors influenced my decision to set up a meeting, e.g., how much I enjoy spending time with them or the required logistics. Planning meetings was a considerable effort. Often, I documented in the diary that I would like a meeting – without planning one. In other cases, I managed to organize meetings (at least tentatively) and postponed them due to the pandemic. Once, a contact organized a meeting after I vaguely suggested one. This experience encouraged me to try to delegate some of the mental load of authentication ceremonies. So I asked conversation partners to remember to authenticate the next time we meet – with no success.

Later during the study, I stopped planning meetings based on the need to authenticate. Instead, I met who I wanted to meet and waited for suitable authentication opportunities. Hence, I shifted my focus from planning opportunities to identifying them. The Christmas season's social events provided ample authentication opportunities. They relieved me from having to plan separate meetings. Also, they made it possible to authenticate conversation partners when setting up an in-person meeting might have been socially inappropriate.

Meeting

When finally meeting contacts, I had difficulties identifying an appropriate time and place to bring up authentication. Usually, meetings begin with a ritual of greetings and inquiries about well-being. I certainly did not want to disrupt my friends when they recount their struggles with "Yeah, that's very interesting, but can we scan these QR codes now?". That would be disrespectful. It is unlikely that anyone would not want to treat friends like that. For me, this led to a repeating cycle of remembering that I wanted to talk about authentication and then deciding that the current situation was not the most appropriate one. For example, I postponed the topic of authentication during a lunch meal because I thought too many people were sitting at the table, a situation in which it could be impolite and awkward to attract everyone's attention to an authentication ceremony. On a different occasion, I postponed the topic because my conversation partner and I were strolling in a park when I remembered it. It seemed inappropriate to stop walking, stand in a circle, maybe block someone else's path to conduct a ceremony. However, postponing the topic can also backfire. During a night out at the bar with a friend, I completely forgot about the authentication after I decided to postpone it. I forgot so thoroughly that I only remembered a month later, during an analysis session, that authentication was my original intent for that meeting! I had fun regardless.

This cycle of postponing is one of the reasons which made it easier to authenticate people I meet regularly. These more informal meetings include more calm moments in which it is not as rude to take out a smartphone or introduce an entirely new topic. However, since I usually talk in person to contacts I meet regularly, authenticating their secure messaging conversation is less important to my overall security.

Convincing

As a security researcher studying authentication ceremonies, I benefited from my perceived authority on the subject when asking others to authenticate. Had it been someone else, I am unsure if people would have followed along as readily.

If someone had not heard of authentication ceremonies before, I explained my research topic to them, inviting them to follow along with their messenger. This approach of interlocking the explanation of my field of research with their participation felt a bit manipulative. There is practically no way for them to deny my wish without coming across as rude. However, I did not feel too bad – it resulted in education about end-to-end encryption with no perceivable downsides.

For me, the most demanding authentication partner was a fellow security researcher. In pursuit of knowledge, they questioned everything and asked for detailed explanations of the encryption, the threat model, and the feasibility of practical attacks. This scrutiny came from a fellow peer after I described this as my area of research. An uncomfortable situation – it felt like they were questioning the validity of my research, even though they probably were honestly interested in the subject. In the end, they went along with the authentication ceremony, but it was a tough sell.

However, in many cases, requests for authentication ceremonies required no justification. Most of these cases were contacts I had authenticated at least once prior.

Authentication

For the most part, I did not encounter any issues with the authentication ceremony interface and procedure. I have extensive experience conducting authentication ceremonies. When contacts offered me their phone, I declined and instead instructed them in a step-by-step manner. Interestingly, I found that almost everybody (including the people I authenticated previously) had problems initiating the verification procedure. Afterward, they often asked what authentication ceremonies achieve. Usually, I tried to explain the effect in simple terms, e.g., "It makes sure that my phone encrypts the messages so that only your phone can read them. If the codes do not match, someone could be listening in and messing with our messages." At the beginning of this study, I thought my experience prevented me from experiencing usability issues. I was wrong. While I know the ins and outs of my preferred end-to-end encrypted messenger, I encountered a problem re-authenticating a secure chat on Telegram. Usually, I initiated a ceremony by sending a message to the contact, making it easier to find the correct conversation. However, I could not find this secure chat in my list, so I asked my contact to send a message to me – which I did not get. I was confused but did not want to dwell on it, so I gave up. Later I found out that secure chats on Telegram are only available on the devices that created them. Hence, my contact and I would have needed to create a new secure chat in any case.

Summary of My Barriers to Authentication

1. Planing meetings takes effort. Waiting for good authentication opportunities is easier.
2. Forgetfulness. Remembering authentication plans and others' current authentication status is hard.
3. The cycle of postponing. Remembering authentication in unsuitable situations leads to postponing.
4. Convincing others to authenticate can be stressful.
5. Some authentication procedures are still unintuitive.

7.4.1. Emotions during Planning and Conducting Authentication Ceremonies

Bella and Viganò suggested that security mechanisms should provide an enjoyable and beautiful experience, i.e., beautiful security [24]. In my experience during the study, authentication ceremonies did not fulfill the beautiful security criteria. This section describes my subjective emotional experiences while planning and conducting authentication ceremonies.

Overwhelmed by the mental load of planning authentication ceremonies. During most of the study, I felt overwhelmed by the information I needed to keep in mind. Even though I regularly checked the authentication status of my conversations, I felt stressed because I was never really sure about the authentication status of the people I met. Regularly

thinking about authentication and potential opportunities became a significant part of my mental load. Often, I forgot to take advantage of good authentication opportunities. In these cases, I became frustrated and annoyed at myself. In contrast, I was usually even more proud of myself when I identified and remembered an authentication opportunity. I am glad that this study is over. I hope that I can forget about the need to authenticate and other people's authentication status now.

Worried about embarrassment or rejection in meetings. During meetings, I worried about the right time to ask for an authentication ceremony – postponing the request until an appropriate situation came up. Asking in an inappropriate situation could inconvenience my contact, hindering them from more entertaining, interesting, or relevant activities than authentication. In turn, I would be embarrassed for inconveniencing my contact. When the response to my request for authentication was not as enthusiastic as I had hoped, I felt rejected, inadequate, or even threatened. These feelings were at least partially related to the fact that authentication ceremonies are part of my research focus. Hence, skepticism and critique felt like an attack on my identity as a security researcher. In contrast, when my contacts eagerly went along with the authentication ceremony and asked relevant questions that I could answer, I felt respected, confident in my abilities, and proud to teach them something they found interesting and valuable.

Mostly content with the results. It felt good to complete authentication ceremonies and see the resulting "Verified" checkmark. It created a sense of achievement and progress. I was even happier when an entire group received the checkmark, i.e., I had authenticated all of its participants. Collecting is a powerful drive for many people. So, a group checkmark may have a similar effect to badge systems in gamification approaches [307]. However, not all authentication attempts ended on a positive note. Some secure messengers confounded me when they did not behave as expected. For example, one messenger claimed that I had marked someone as unverified while I did not remember doing so – making me doubt reality. In another instance, I wanted to reauthenticate a contact who had gotten a new phone. However, that conversation had remained verified. To this day, I still do not know why.

Carefully treading around conversation partner's emotions. Near the end of the study, I noticed that some contacts felt called out when asked for authentication. They got very defensive and felt the need to justify why they usually did not take certain security precautions. While this did not happen too often, it took some effort to convince them that I did not want to shame their behavior with my request. Usually, I explained that it is appropriate not to take security precautions they do not deem necessary, as long as this is their conscious choice. Thus, secure messenger users may need to consider their conversation partner's emotions about security in their request for authentication.

7.4.2. Sociocultural Aspects of Authentication Ceremonies

Previous result sections focused mainly on describing my experiences while planning and conducting authentication ceremonies. In the spirit of autoethnography, this section

places these experiences into their broader sociocultural context. Thereby examining how sociocultural factors may hinder or facilitate authentication ceremonies.

Authentication ceremonies are embedded in social rituals. Meeting someone exclusively for an authentication ceremony, i.e., meeting without small talk, authenticating, and leaving, seems almost unthinkable – such a meeting sounds ludicrous. In meetings, I generally feel socially obligated to inquire at the very least about the other's well-being. Instead, I embedded my authentication ceremonies without a second thought in social rituals: I went for a walk with my contacts, went cycling with my cycling friends, met contacts for lunch or dinner, or talked with them during the afternoon coffee break at work. After all, I meet friends because I like them and want to catch up with them, not solely for authentication. However, while I do not consider myself shy, I had difficulties arranging meetings with some contacts. In particular, this concerned contacts that I do not know well and with whom I struggled to come up with a good reason for a meeting. Inviting others to a meeting usually implies an interest in a more personal relationship, which does not apply to all my messenger contacts. Hence, authenticating these contacts without arranging a potentially awkward meeting would only be possible if an appropriate situation came up by itself. As a result, some contacts are difficult or even impossible to authenticate if users miss a socially appropriate occasion. Not every authentication ceremony requires a social ritual. For example, I would not feel awkward calling my best friends just for authentication. However, this exception only applies to a few people.

Rules that govern social rituals apply to embedded authentication ceremonies. I authenticated my contacts in physical in-person meetings and remotely over video calls for this study. I decided the mode of meeting according to what felt right to me at the moment. I attempted to stick to familiar settings for authentication ceremonies. When I usually conversed with someone via video call, I arranged a video call. If I usually met someone over dinner, I invited them to dinner. The idea of deviating from this procedure for an authentication ceremony felt off to me. I did not want to appear rude during the meetings and followed familiar conversational conventions. I avoided hijacking conversations centered around entirely different topics. Instead, I waited for an appropriate topic of conversation, where it seemed like phone use, messaging, or security would fit in. In particular, I avoided bringing up authentication ceremonies when others told me about their recent experiences and worries – I wanted to show that I care and would have felt ashamed of interrupting them in these personal moments. I found that not only can a conversation topic remind me of asking for an authentication ceremony, an authentication ceremony usually also influences the conversation topic for several minutes. Conversation partners may have considered it rude not to follow up on a topic that interested me. Usually, this resulted in a short educational episode about security on smartphones, regardless of whether I had intended it like that or not. Lastly, I found that using phones was not always socially appropriate. Seeing my smartphone reminded me of authentication ceremonies, so remembering them became harder in such scenarios. In any case, remembering an authentication ceremony was less challenging when the social situation allowed taking out phones. Striking up a conversation about

messaging and its security was also less challenging when the conversation partner's phone was in view.

Established social practices can make it easier to authenticate others. Asking for an authentication ceremony immediately after exchanging contact details in person did not feel awkward during the study. Hence, this may be a viable approach to authentication when arranging a separate meeting for an authentication ceremony is socially inappropriate. Predictable sociocultural gatherings, such as Christmas, new years celebrations, or birthdays, were a perfect opportunity to conduct authentication ceremonies. Taking part in these gatherings does not require as much planning. Therefore, minimizing planning failures and potential social awkwardness. Also, these gatherings are usually informal so talking about authentication felt less like imposing a topic of conversation on others. Lastly, during some meetings, my planned authentication ceremony influenced others to try authentication ceremonies. This effect is in line with Das et al.'s [52] suggestion that social influence might drive the adoption of a visible security feature. Hence, conducting authentication ceremonies at social gatherings may create social network effects.

7.5. Discussion

Based on the collected diary entries, two main reasons make it hard to authenticate even for motivated and knowledgeable users. First, planning meetings for authentication ceremonies and identifying good authentication opportunities creates a huge cognitive load. The first author had to think about authentication all the time and remember who they wanted to authenticate. They were bound to forget at least some of the time. Caring about authentication after forgetting once or twice requires a considerable amount of security motivation. Second, sociocultural aspects hinder widespread adoption. It is comparatively easy to authenticate people you meet often and know well. For other types of contacts, the experience was different. Sometimes, it might even be socially inappropriate to arrange meetings for authentication.

Simple measures, such as adding consentful and context-sensitive notifications, may alleviate the cognitive load of authentication. In comparison, overcoming sociocultural barriers to authentication ceremonies requires a dedicated design approach.

7.5.1. Designing Cooperative Security

Authentication ceremonies are a prime example of cooperative security mechanisms; two contacts need to cooperate to mitigate MitM attacks against either of them. Hence, well-known design principles for individual human-security interactions cannot solve this design challenge.

Other examples of cooperation from the Usable Security literature come to mind: users wait for other users' reviews of software updates before installing them [269], and users depend on their friends, who have their contact data, to keep their contact data private [224]. Enabling and supporting such cooperative behavior requires designing our security explicitly for it.

In the following, we discuss *cultural transcoding* as a core design issue for cooperative security. According to Manovich [175], cultural transcoding is one of the five principles of new media. They assert that new media consists of a *cultural* and a *computer* layer. Both of these layers influence how the other works. The effort of converting one aspect from one layer to the other is what Manovich describes as transcoding. Cooperative security mechanisms have two similar layers: the computer layer, consisting of cryptographic protocols and security requirements, and the cultural layer, i.e., the social rules that govern our interactions. During the study, the first author had to do the entire transcoding effort unsupported. It was their responsibility to integrate authentication ceremonies into their social life in a socially appropriate manner. It remains to be discussed who is responsible for this cultural transcoding work. At the moment, this task has to be performed by users of secure messaging applications. Integrating transcoding work in the design process for cooperative security mechanisms will move a part of this responsibility from the users to the designers.

For authentication ceremonies, an integration could work as follows. (1) Understanding what kinds of situations are appropriate for authentication ceremonies depending on the context of the relationship and other cultural factors; (2) Support users in identifying these situations; and (3) Support users in initiating these ceremonies in a socially appropriate way, e.g., by integrating them into a well-known ritual.

7.5.2. Self-Inquiry and Autoethnography in Usable Privacy and Security

Qualitative studies in Usable Privacy and Security use diverse methods to study security tools and their users, ranging from field studies to lab studies, using surveys and interviews. All of them are valuable to have in our method catalog. However, self-inquiry studies among security researchers are rare. When we, as researchers, deliberately get involved in studies, we do so in expert roles, e.g., for contextual walkthroughs, designing security features, or analyzing qualitative data. By comparison, we seem to disregard our own user experience with security features. We are users too and hence, our self-reflections and experiences provide indispensable insights into the usability of security and thus form a valuable baseline for user research. If we have issues using security mechanisms, we should treat this as an early warning sign. This does not mean everything is fine if security researchers have no issues. Neither does it mean we should design security features just for ourselves.

Self-Inquiry is a systematic method to investigate our own experience with security features and tools. Over an extended period, it allows deep introspection of tool use. Whenever security and privacy mechanisms intertwine with social or cultural aspects, autoethnography is a viable research approach. In the case of this study, we chose this approach to systematically collect evidence based on experience instead of relying on anecdotal evidence alone. Since authentication ceremonies are highly contextual and depend on social and cultural factors, autoethnography is especially suitable to investigate them.

From our experience in this work, we found two aspects that future researchers who want to use this method may find useful. First, especially for reoccurring behavior, routinization during the course of the investigation may become an issue. Hence, the

diary study itself may influence the behavior under investigation. We coped with this issue using reflection, transparency, and open discussion about its effects. Second, while the experiences and their interpretation are always the first author's, it is helpful to get feedback on how readers with a different background may interpret the situations as reported in the results. Based on that feedback, the authors can add more context to situations when it is necessary to sufficiently understand them.

Personal experiences influence and shape our research ideas. Since we cannot avoid it, we must communicate this influence transparently. We can even embrace this effect by seeking security-relevant experiences to investigate potential research ideas. Documenting these experiences in a diary helps trace the resulting research questions back to them. We plan to explore more kinds of new security technology in this way, documenting experiences, and finding potential research questions based on them.

7.5.3. Implications for Authentication Ceremony Research

The findings of this work suggest a change of direction for future research on secure messaging authentication ceremonies.

First, the results encourage the use of field studies for future work on the adoption of authentication ceremonies. While evaluating interface usability and user comprehension works well in lab settings, researching sociocultural factors is harder. Field studies of people's regular messenger use enable investigation of users' real-world behavior and the sociocultural factors that influence them. Field studies with proposed authentication ceremonies will likely require cooperation with messenger providers to arrive at results with a high ecologically validity.

Second, future research should include sociocultural factors in data collection and design. As this work demonstrates, current authentication ceremonies do not consider sociocultural factors in their design. However, designing ceremonies for these factors requires understanding their contexts of use. Hence, we need to collect data on how users want to authenticate their secure conversations. Given the security context of authentication ceremonies, focusing on the sociocultural context of users who are targeted by surveillance (e.g., members of protest movements) is likely a good first step.

Third, focus research and design on motivated and knowledgeable users with specific threat models. Effective security education is a challenge. Therefore, instead of investing research effort in everyone's security education, it might be wiser to identify specific groups of motivated users and work on removing their barriers to authentication ceremonies. The social influence could then lead to broader adoption of authentication ceremonies [52].

Lastly, our proposed phase model informs future research on authentication ceremonies. While prior work focused on the usability and user comprehension of authentication ceremonies, our results imply that other aspects, such as recognizing the need for authentication, planning, and remembering are crucial and should be better supported by technology. With the different challenges that we identified throughout the process, we argue that individual challenges, such as timing, need to be studied thoroughly and with more depth than in previous studies. For example, future design work may help users to identify conversations that are especially important to authenticate — corresponding

to the "recognize need" phase. Other types of design work may make planning for authentication ceremonies easier or help identify convenient situations for conducting these ceremonies — corresponding to the "plan" or "meet" phase, respectively.

7.5.4. Alternative Approaches to Studying Authentication Ceremonies in the Field

While we preferred an autoethnographic approach for our research questions, other approaches – with different benefits and drawbacks – also work well for studying authentication ceremonies in the field. To support future research endeavors, we briefly discuss alternative approaches.

Diary study. Similar to this work, researchers may use a diary study approach to document participants' experience with existing authentication ceremonies. However, a diary study by itself will likely not get many naturally occurring authentication attempts; limiting the available data. One way to mitigate this issue is prompting participants to try to authenticate or ask participants why they did not conduct authentication ceremonies. The responses might tell us if participants felt it was important to authenticate the messaging conversation in question or how awkward or straightforward they found it to bring up the topic. These prompts likely impact users' natural authentication behavior, thereby reducing ecological validity.

Intervention study. An intervention study could be used to understand the factors that influence participants' authentication plans and behaviors. The interventions could explain threat models, demonstrate the necessary user interaction, or use role-playing to get participants used to authentication ceremonies. Afterward, surveys or interviews may help understand which of the interventions have a promising effect.

Prototype field test with pairwise recruiting. Testing a prototype of an authentication ceremony is difficult in the field since not only the participant but also their conversation partner need to use the prototype. Participants cannot use their regular messenger for this type of field test because prototype versions of authentication ceremonies are usually not interoperable with existing messengers. Since participants can only authenticate a limited set of conversations (with other study participants), this study approach lacks natural interaction. Recruiting participants pairwise ensures that each participant can at least authenticate one other person. However, this approach also makes recruiting more difficult.

Prototype field test in cooperation with a messaging company. Cooperation with an operator of a widespread secure messenger enables study designs with more ecological validity. Prototypes can be rolled out in regular A/B tests. Study participants can authenticate any conversation they like because their conversation partners' user interface adopts accordingly. The benefits of this approach are easy recruiting and natural user interactions. However, some form of a prompt might still be necessary to make users aware of the authentication ceremony and tackling qualitative research questions might

be more difficult when cooperating with a messaging company. Also, establishing these kinds of cooperation with industry partners may prove difficult.

7.6. Conclusion

In this work, we used autoethnography to investigate why even motivated and knowledgeable users may have difficulties using authentication ceremonies.

Based on the collected data, we found that planning and conducting authentication ceremonies results in a huge cognitive load. The first author needed to keep authentication status in mind, plan meetings, and identify opportunities in time. Often, they forgot about the ceremonies, which resulted in a frustrating experience. Additionally, they had to constantly navigate social rituals to integrate authentication ceremonies in socially acceptable ways. Primarily, this navigation was necessary for formal relationships with acquaintances from work or the members of the extended circle of friends. In contrast, authenticating close friends was less complicated.

Consentful and contextual reminders may alleviate the cognitive load in many cases. However, addressing the social aspects of cooperative security mechanisms, such as authentication ceremonies, is more challenging. Integrating cultural transcoding into the design of cooperative security may improve the situation. Using this approach, designers would consider how culture influences security technology and how the security technology may affect cultural practice. Methods of self-inquiry, such as this work, are applicable in Usable Security and Privacy Research. They indicate design flaws early – if we have trouble using a security mechanism, others may have as well. Ultimately, we need to keep in mind that security researchers are users too. For future research on authentication ceremonies, we recommend using field studies to understand real-world use and the sociocultural factors that influence it.

Part III.

Matching Secure Experiences with Actual Security

The contents of the following chapter were published as part of the publication *"Exploring User-Centered Security Design for Usable Authentication Ceremonies"* (CHI 2021) [P6]. This paper was produced in cooperation with my co-authors Lea Gröber and Katharina Krombholz. The following table describes the contributions of all authors to this paper:

Author	**Contribution**
Matthias Fassl	I had the initial idea to redesign authentication ceremonies for E2EE messengers. I planned and conducted the collaborative design workshops. Together with Katharina, I analyzed the qualitative results. I conducted an expert review of the design suggestions. I developed the initial storyboard prototypes. I planned and conducted most of the iterative storyboard prototyping sessions. I planned and conducted the feedback session with the UX expert. I planned, conducted, and analyzed the online evaluation survey of the different prototypes. I wrote the initial draft of the paper and all the following revisions.
Lea Gröber	Lea helped me design background material for one of the prototypes, helped recruit participants and collect data for the iterative storyboard prototyping approach, and contributed to the qualitative analysis of the open-ended questions in the final evaluation survey.
Katharina Krombholz	As my academic advisor, Katharina was involved in the major decisions during this research project. She suggested using collaborative design workshops and contributed to the qualitative analysis of the workshop sessions. She guided me through methodological and ethical questions, reviewed my interpretation of (intermediate) results, and helped me find an appropriate venue for this research project. She also edited draft versions and reviewed the final submissions.

Reference

Fassl, M., Gröber, L. T., & Krombholz, K. (2021). Exploring User-Centered Security Design for Usable Authentication Ceremonies. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, 1–15. `https://doi.org/10.1145/3411764.34451649`

8

Exploring User-Centred Security Design

Security technology often follows a systems design approach that focuses on components instead of users. As a result, the users' needs and values are not sufficiently addressed, which has implications on security usability. In this paper, we report our lessons learned from applying a user-centered security design process to a well-understood security usability challenge, namely key authentication in secure instant messaging. Users rarely perform these key authentication ceremonies, which makes their end-to-end encrypted communication vulnerable. Our approach includes collaborative design workshops, an expert evaluation, iterative storyboard prototyping, and an online evaluation.

While we could not demonstrate that our design approach resulted in improved usability or user experience, we found that user-centered prototypes can increase the users' comprehension of security implications. Hence, prototypes based on users' intuitions, needs, and values are useful starting points for approaching long-standing security challenges. Applying complementary design approaches may improve usability and user experience further.

8.1. Introduction

There is a broad consensus that users should not be required to make complicated security decisions that they cannot make in an informed way. However, users are still required to proactively perform complex security tasks that they hardly understand. An example of such a security task is the authentication of keys when using secure messaging apps. We argue that it is worthwhile to consider users from the beginning of the design process for these kinds of security tasks. To foster discourse on design methods in usable security, we applied user-centered design to a well-understood challenge in usable security, namely *authentication ceremonies* to authenticate or verify keys in instant messaging apps and report our findings and lessons learned in this paper.

In November 2018, the Dutch police decrypted 258,000 messages sent with the E2EE messenger IronChat [223]. Investigations by a Dutch news network [251, 108] revealed that police officers most likely changed all encryption keys, thereby deploying a large-scale MitM attack against IronChat's users. Affected users could have detected and

mitigated this attack if they had used an authentication ceremony to authenticate the used encryption keys. However, few users are aware of these ceremonies [282], and those who are aware have problems conducting them correctly [254, 284, 138].

Most of these authentication ceremonies are based on much older methods to securely pair multiple devices owned by the same user [20, 102, 140, 157]. Following a traditional *Systems Design* approach, these methods have been repurposed for public-key authentication in mobile messaging while overlooking that the two involved devices are owned by different users who are typically not co-located.

Involving multiple people introduces social aspects to the authentication ceremony that were not considered before: e.g., users may feel a potential discomfort when asking others to authenticate or may be embarrassed not to know how to use security features [279]. After discovering that existing ceremonies do not work well, prior work focused primarily on incremental improvements [285, 298] or proposed to remove users from the loop [283, 192] (which potentially leads to a lack of trust [245, 179]). Most incremental improvements use an *Activity-Centered Design* approach to remove possible usability issues, which continues to be important and necessary yet leaves other aspects untouched. Even after all these important scientific contributions using different design approaches, authentication ceremonies still struggle with users' comprehension of their security benefits, long completion times, and consequently low adoption rates.

In this paper, we discuss if and how authentication ceremonies could benefit from an alternative *User-Centered Design* approach. Using this approach, we design authentication experiences from the ground up, involving prospective users, security experts, and a UX expert in the design and evaluation procedures. This integration allows (a) exploring the users' design space of authentication, and (b) incorporating security and UX requirements to enhance user-generated prototypes.

We applied a four-stage *User-Centered Design* process: (1) five collaborative design workshops with ten potential users to gather ideas and drawings that reflect users' perceptions of authentication ceremonies and trust establishment, (2) a security evaluation that narrowed the design space, (3) an iterative storyboard prototyping approach with 18 participants to improve usability and collect participants' preliminary security perceptions, and (4) an online evaluation of the prototypes on Amazon MTurk with $N = 131$ participants.

The evaluation indicated that exposure to the combination lock prototype improved participants' understanding of the security benefits concerning different types of attackers (compared to Signal's current ceremony). This understanding of security benefits likely affects the frequency with which users will conduct authentication ceremonies. Our design approach did not seem to improve other factors, such as usability and user experience. Since these factors are equally important for building authentication ceremonies that people will use, we suggest designers apply complementary design approaches to improve them. We report the methodological lessons that we learned from exploring user-centered and participatory techniques for security design. We are confident that our exploratory contribution sparks an interdisciplinary discourse on when and how to consider prospective users throughout the design process of security technology.

8.2. Related Work and Background

History of pairing mechanisms. In the early 2000s, the secure pairing of devices emerged as a heavily studied research topic [140, 20, 102, 187]. The proposed pairing methods were designed to pair two or more devices owned by a single user. The comparative usability study by Kobsa et al. [157] showed that in cases where both devices had screens, a comparison of PINs, sentences, or images receives the highest usability scores. Starting in 2010 with TextSecure (later renamed to Signal), secure end-to-end-encrypted messaging has become a de facto standard in mobile messaging. Many of these messengers used the aforementioned device pairing methods to implement their authentication ceremonies. However, while device pairing only involves one user, authentication ceremonies in secure messaging involve two users who are potentially not even co-located. Pairing methods involving multiple users are also referred to as social pairing.

Lack of adoption of currently deployed authentication ceremonies. Using device pairing methods for messengers' authentication ceremonies without further considering their context has proven to be insufficient. As an increasing amount of research about the failure rates of these authentication ceremonies demonstrates: Herzberg et al. [138] studied the usability of WhatsApp, Viber, Telegram, and Signal. They found that (1) participants were not aware of the need to authenticate, and (2) 56.5% of the participants failed to authenticate in all messengers after being instructed to do so. Schröder et al. [254] found that the majority of participating CS students failed to detect and mitigate MitM attacks using Signal's authentication ceremony. Vaziripour et al. [284] compared authentication ceremonies of Viber, WhatsApp, and Facebook Messenger. They found that only 14% of their study participants successfully verified the key material without further explanation. When Vaziripour et al. [282] surveyed Iranian Telegram users, they found that only 29.6% had ever used the authentication ceremony in text conversations.

Proposed improvements of authentication ceremonies. Researchers have tried to improve the authentication situation mostly in two different aspects: (1) streamlining the users' authentication activity, or (2) removing the users from the loop. In the first category, Tan et al. [267] focused on improving success rates by identifying a suitable key representation and mode of comparison. Vaziripour et al. [285] streamlined Signal's ceremony by providing easy access and additional guidance. Wu et al. [298] produced new visual indicators, new notification dialogs, and a new simplified notification flow. In the second category, Vaziripour et al. [283] partially removed users from the loop by using Keybase for an authentication method based on social media. Melara et al. [192] proposed a key transparency log that would remove users entirely from the authentication. All these approaches have resulted in valuable improvements regarding specific aspects of authentication ceremonies. Few of these approaches have considered the social aspects of authentication ceremonies or the users' need for a secure experience. Hence, they are important first steps towards solving the challenge. Those aspects need to be addressed from the ground up during the design of security features.

Our work aims to promote the adoption of user-centered security design by presenting a four-stage design process (see Section 8.3) that goes beyond the initial ideation of dialogs and is applicable to entire security tasks. The remaining sections exemplify this design process on the use case of authentication ceremonies in secure instant messaging.

8.3. Design Method

User-Centered Design involves users at every step of the design process and focuses on their goals and requirements. Usually, this process is iterative in nature and driven by evaluation. We chose our process with two main goals in mind: (1) exploring the design space from the users' perspective – including a security evaluation of the specific suggestions and extracting general themes about authentication from the qualitative data; (2) comparing the most promising of the candidates with an existing authentication ceremony – necessarily these candidates need to be fleshed out and developed further for a meaningful comparison. Users are the main focus of our research, but they are neither security nor UX experts – we do not expect them to design usable and secure ceremonies on their own. Security experts are necessary to contextualize participants' conceptual ideas and match them to existing security mechanisms. UX experts are indispensable to designing usable interactions with great security experience. Above considerations and our comprehensive literature survey resulted in a four-stage design process as presented in Figure 8.1:

1. *Collaborative Design Workshops.* We start by collecting users' ideas about how authentication should work including their goals and requirements. We apply a participatory design approach by conducting collaborative design workshops. Participatory design considers the users' cooperation with designers as a possibility to bridge the gap between the users' tacit knowledge, i.e., knowledge that is hard to communicate, and the designers' abstract and analytic knowledge [262]. We structured these workshops similar to the participatory design studies by Weber et al. [291] and Gorski et al. [115]: (1) participants reported about their experience with secure messengers, (2) we created a shared language about MitM attacks and authentication, and (3) participants designed and discussed conceptual authentication ceremonies. Section 8.4 provides a detailed description of the collaborative design workshops.

2. *Narrowing Down the Design Space.* The design workshops resulted in a design space with many creative conceptual ideas. We conducted a security evaluation to narrow this design space to feasible ideas. We selected concepts that users mentioned commonly and perceived as secure. If possible, we matched those concepts with security mechanisms that actualized the concepts' perceived security. We removed concepts that participants mentioned rarely and concepts without matching security mechanisms. Section 8.5 provides a detailed description.

3. *Iterative Storyboard Prototyping.* After selecting viable concepts, we design prototypes based on them. We refine and evaluate these prototypes with users in an iterative process. We apply prototyping techniques because they can

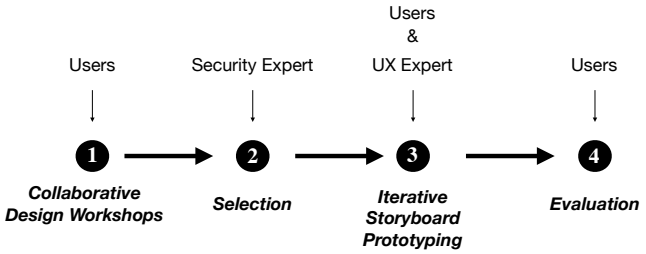


Figure 8.1.: Overview of the design-process and the involved parties. We describe each of the steps in a separate section in the remaining paper.

also be used to analyze the work process of prospective users and simulate possible future work [30]. For each concept, we sketched all possible states of the user interface, which resulted in a storyboard. We presented these storyboard prototypes to potential users and asked them to describe the concepts behind them and estimate the prototypes' security benefits. Then, we encouraged them to redesign the storyboards using pen and paper. After two iterations, we presented the storyboards and the feedback we received to a UX expert with several years of experience and improved them accordingly. Section 8.6 provides a detailed description of the iterative storyboard prototyping.

4. *Evaluation.* After concluding the iterative development of prototypes, we evaluate the prototypes' secure experience and usability with prospective users. In a between-subjects online study, we evaluate the three resulting prototypes against a storyboard prototype of Signal's authentication ceremony. We recruit participants on Amazon MTurk and randomly assign conditions. Afterward, they respond to a questionnaire about the prototypes' usability, user experience, perceived security, and protection against different threats. Section 8.7 provides a detailed description of the evaluation.

Security requirements. The purpose of authentication ceremonies is to mitigate MitM attacks. To do this securely, the conversation partners have to compare the key material they use. The conversation is secure if all conversation partners agree on the correct key material. Comparing key material is possible in three ways: (1) out-of-band, (2) in-band, and (3) using zero-knowledge proofs.

For the first kind of comparison, the conversation partners have to meet, which requires planning for potential future security requirements. The second and third types of comparison allow an in-the-moment approach to security. In our study, we did not prescribe any method of application in order to not restrict the intuitions of our participants.

Before conversation partners decide to use an authentication ceremony they have to negotiate the need for one. The person who identifies the need for additional security has to explain the purpose and necessity of authentication ceremonies. This explanation requires at least a high-level view of potential attackers' capabilities. Additionally, the

user experience of the ceremony needs to provide convincing evidence of its protective power – even if users do not exactly understand its technical background.

Ethical considerations. For all three parts of this work, we collected basic demographic data but no personally identifiable information. For the workshop, we collected the participants' email addresses to organize the workshops and communicate the results. These were stored separately from the study data. All participants were informed about the purpose and procedure of the study. Before the workshop, we asked all participants to sign a consent form; for the iterative storyboard prototyping, we asked for verbal consent. The online study had a consent form on the first page, including all necessary information on data collection and processing. We compensated all participants for their time. Our university's ethical review board (ERB) approved the study.

8.4. Collaborative Design Workshops

We conducted workshops to collect the prospective users' ideas of authentication ceremonies, e.g., how they imagined the process and what motivation they need to complete it. Similar to Weber et al. [291] we divided each workshop into three phases: (1) discussion of experiences with secure messaging, (2) creating a shared language for basic security concepts, and (3) prototyping conceptual ceremonies. In the first phase we asked all participants about their common messenger usage. Additionally, we encouraged participants to report negative and positive experiences with these tools. The goal of this phase was to acquaint the participants with each other and to identify general challenges with secure messaging.

During the second phase, we presented a slide show (see Figure E.1 in the Appendix) explaining end-to-end encryption, threat models, and MitM attacks to create a common base of knowledge and to establish a common language among the participants. In the last phase, we asked the participants to provide prototypes of conceptual authentication ceremonies and to explain their ideas verbally and through drawings. In the beginning of this prototyping phase, we asked the participants to provide suggestions on how to ensure that they are communicating with the intended person if they could meet in person only once or not at all. Later on, we asked them how they establish trust in the offline world, and if and how those strategies could be translated to electronic communication. After about an hour we debriefed the participants and discussed remaining open questions.

Analysis. We collected different types of data: a set of drawings of conceptual prototypes, audio recordings, and our written notes. Two independent coders traversed the notes and corresponding drawings to systematically assign codes, a process known as *open coding*. We used the audio recordings to clarify misleading notes or drawings, which were difficult to understand. The resulting inter-coder agreement (Krippendorff's $\alpha = 0.69$ for the experience reports and $\alpha = 0.67$ for the prototypes) allows us to draw tentative conclusions from the data. We grouped the resulting codes into categories to identify the most common concepts for authentication ceremonies among our participants. The full protocol and codebook are presented in Section E.1.1 and E.1.2 of the Appendix.

Since the workshops were conducted in a different language the researchers translated the codebook, quotes, and the shown drawings.

Pilot study. We conducted a pilot study consisting of three sessions with one participant each. Contrary to the rest of the study, the participants had basic knowledge about cryptography. Since those participants came up with rather unusual prototypes for authentication ceremonies, we assumed that the procedure would also work for other participants. Based on the pilot study we concluded that the notion of trust needs to be well-defined when asking participants how they would establish trust. Therefore, we decided to add two offline trust scenarios to our study design: (1) meeting a previously unknown bank advisor, and (2) handing over a package that was accepted for an unknown neighbour.

Recruitment and participants. We conducted five sessions with two participants each. We invited interested users who used secure messaging applications, who described themselves as having a lay person's understanding of cryptography, and who did not have concrete threat models in mind. We deliberately excluded participants with either a background in cryptography, or with concrete threat models. Both types of participants will already know about authentication ceremonies and have preconceived notions on how they should work – which would narrow their design space. We created a dedicated website to inform about the study and advertised it via email, Facebook, and Twitter. Additionally, we used snowball recruitment to quickly find interested and qualified participants. We compensated them for their time with food and non-alcoholic drinks during the sessions, and by offering future security advice (which one participant accepted).

All ten participants either graduated from a university or were currently attending one. Four participants received some kind of training related to computer science, but had no further knowledge in security or cryptography. Seven participants were female and three were male. The average age was 28.2 (min=22, max=35, sd=4.76). The participants self-reported their knowledge on cryptography and IP networks on a scale from one ("very little") to six ("very much"). The participants rated their knowledge about cryptography (m=2, sd=1.3) as well as IP networks (m=2, sd=1.6) as low. This reflects that our target population should only have a lay person's understanding of cryptography.

8.4.1. Results

In the following, we present the experience reports, the most common conceptual prototypes that came up during the workshops, and the priorities and expectations that participants explicitly named.

Experience reports. At the beginning of each session, participants reported on their experience with secure instant messengers. Convenience was the participants prime reason to praise messengers, e.g. *Telegram* and *WhatsApp* have large user base and work on all platforms. Peer pressure is an important reason for choosing messengers: *"If I*

have a close friend who insists on only using WhatsApp and I really want to communicate with him, then I am forced to use WhatsApp, even if I don't want to – otherwise I have no way of communicating with that friend". Features were equally important for users, e.g., participants liked *Telegram* for its sticker packages. All participants expressed annoyance about the diversity of apps and lacking inter-operability. Some participants criticized messengers that require a phone number. *Signal* but also to a lesser degree *Telegram* were criticized for their lacking quality of service. Many participants did not trust *WhatsApp* and *Facebook Messenger* because of Facebook's bad privacy reputation. Interestingly, usability did not seem to be a major concern for our participants. Participants were mostly unaware of authentication ceremonies' existence. The few aware participants perceived them as confusing rather than helpful: *"I don't really understand how it [WhatsApp encryption] works, because it says encryption is used, but when you access the contact data, you can encrypt it again with some kind of code, so I don't get that. [...] and you have to be in the same place to do that, that's very bothersome."*

Trust establishment. The participants proposed numerous ways of authentication in electronic communication and provided 20 conceptual prototypes. We categorized the concepts into six methods of establishing trust: (1) *shared knowledge*: comparing knowledge that is only known to the conversation partners, (2) *picture-based*: showing pictures or videos of conversation partners, (3) *social*: asking friends or trusted contacts if they have authenticated the conversation partner, (4) *institutional*: trusting institutions to correctly authenticate people, (5) *habituation*: building up trust in the identity of the conversation partner over long periods of time, and (6) *measurement-based*: using technological measurements to test if the conversation could currently be under attack. Fifteen of the suggested prototypes were from the first three categories, which suggests that these come more intuitively to mind than others. The other three categories of trust establishment were not as popular and only had one or two suggestions each.

Shared Knowledge: Nine out of ten participants proposed an identification method based on shared knowledge immediately after we confronted them with the possibility of communicating with an intruder. The three most common concepts were: (1) exchanging a password used for accessing conversation, (2) agreeing on code words and communicating using Spy Speak (a common TV trope), and (3) asking personal questions that only the other could answer. They reported high confidence in these methods, since they assume that only their conversation partner knows the agreed code words or can answer the personal questions.

Picture-based: Six participants suggested a picture-based authentication of conversation partners and three of them provided a conceptual prototype for this method. Participants suggesting this were quite confident that they were talking to the right person afterwards, since they usually knew the face of their communication partners. However, most of them noticed that an attacker could spoof pictures. Therefore, the resulting trust would increase if senders could prove that the pictures are recent or if real-time communication, i.e. a video-chat, is used.

Social: All participants reported everyday life situations in which they receive information about identities and trust from their social contacts, but most were uncertain how

this process of establishing social trust could be translated to electronic communication. The three participants who provided a conceptual prototype for social authentication wanted the messaging client to automatically establish which of their contacts is trusted by one or more friends. The participants reported that the resulting trust from social authentication would be medium to low, suggesting that social authentication can only be part of a more extensive authentication concept and that trust transitivity highly depends on the friend who verified the contact.

Institutional: During the discussion about establishing trust in the offline world, five participants mentioned that they would ask for some form of institutional identification card. Two other participants said that in business scenarios they would check the name tag of their conversation partner to establish the person's name and their affiliation. This form of trust is based on the issuing organization: if a bank or a government vouches for someones identity, the trust in the organization is transferred to the person in question. However, none of the participants had a suggestion on how to translate this form of trust establishment to electronic communication, which means that we did not receive a conceptual prototype for this form of trust.

Habituation: Offline relationships with neighbors, colleagues, and even bank employees indicate that some kind of trust can be built up over time. Almost all participants said that this is not a fool-proof way of establishing trust, but that they nonetheless depended on this method in some ways. Participants usually agreed that this method could be useful in electronic communication as well. They said that time builds valid identities because information collected over time can be matched with information from other sources. Measuring this trust involves either counting the number of messages between conversation partners or measuring the time since the last key change.

Measurement-based: Even participants who are reluctant to conduct authentication ceremonies with every contact might still want to verify the communication security in more sensitive circumstances. Testing based trust establishment reflects this belief and offers different ways to test the communication channel for eavesdroppers. Approaches to testing-based trust were mostly technology based, one included meeting up and comparing the received messages in order to reveal if any manipulations took place, and another checked the quality of transmission.

Drawings. In the following, we present three conceptual prototypes based on most frequent coding categories. We translated the participants' pen and paper drawings to English.

Figure 8.2 shows a prototype based on *shared knowledge*. Two communicating partners meet in person and exchange a passphrase, which one of them uses to upload a document afterwards. Only the person who knows the passphrase word can download the document at a later point in time. As the code word is negotiated offline, attackers do not know the passphrase, assuming security properties are resistant to guessing. Figure 8.3 shows the corresponding UI, a conversation that is inaccessible until the password has been entered without exceeding the limit of guesses.

Picture-based authentication ceremonies assume that seeing the actual person invokes trust in the person's identity. A major problem with this approach is spoofing. As a mitigation strategy one participant proposed to request images showing the communica-

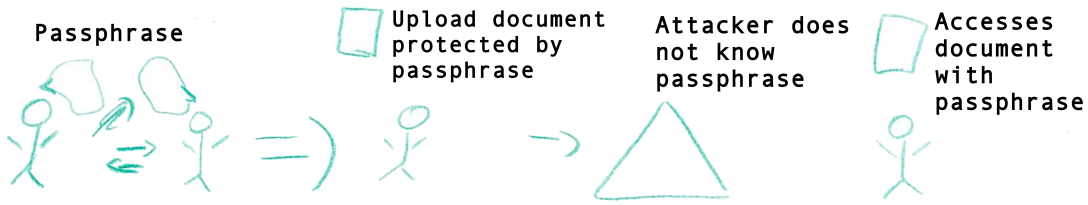


Figure 8.2.: In-person meeting for exchanging a passphrase that protects against attackers.

Figure 8.3.: Conversation partners need a previously exchanged password to access the conversation.

tion partner performing a specified task. Figure 8.4 on this page shows an example of such a task used as an encryption code. However, the participant who designed this method was not fully convinced about the resistance to image manipulations.

All workshop participants preferred automatic to manual *social* authentication, but had difficulties drawing an automatic process. Most of them focused on the visualization of trust levels in the UI. Figure 8.5 in the Appendix shows the use of color codes corresponding to the trust status associated with a particular contact. Green was used for trusted friends, yellow for contacts that have been authenticated by trusted friends, and red was used for all other contacts. The designer emphasized that a trust network should be shallow, i.e. trust information should always come from a trusted friend. Multiple layers of trust inheritance are confusing for users and reduce the confidence in the result. The proposed trust level for "vouched for by a trusted friend" was suggested as medium to high, which seems promising for a method without required user interaction.

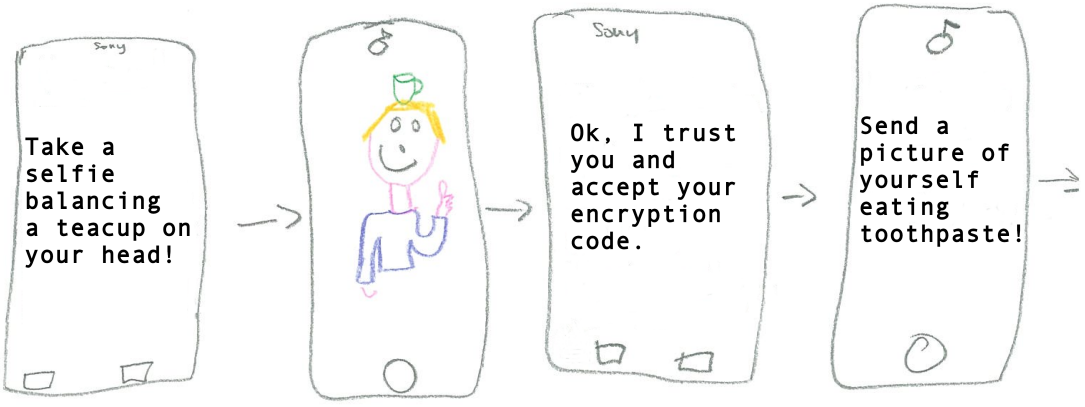


Figure 8.4.: Conversation partners authenticate each other by taking pictures of themselves executing a task defined by the other partner.

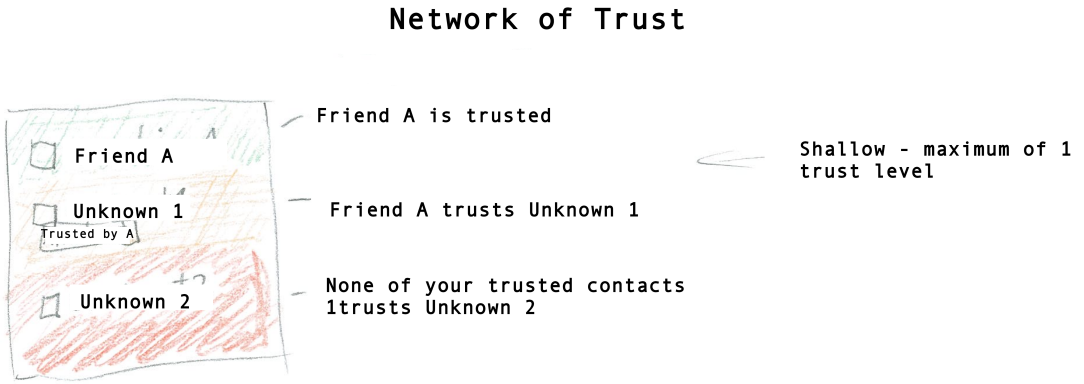


Figure 8.5.: A contact list showing the origin of trust information and color-coded entries based on the trust status.

8.5. Narrowing Down the Design Space

The collaborative design workshop resulted in 20 conceptual designs and qualitative data about different authentication schemes in the offline and online world. One of the authors, a security and privacy professional with experience in the industry (2 years) and academia (2 years), narrowed down this design space based on the following criteria: (1) How common is the suggestion? and (2) Is it possible to actualize the perceived security of the concept design? The first criterion assumes that suggestions are common because many people understand them intuitively, making them valuable as an authentication ceremony for the general population. The second criterion combines the users' tacit knowledge about authentication with the researcher's experience with security and privacy technology. We excluded conceptual designs which the security professional could not match to appropriate security mechanisms. In the following, we present the three chosen conceptual designs with corresponding threat models and security mechanisms.

Shared knowledge → Combination lock. The most common concept design is based on shared knowledge. In those designs, participants suggested sharing a code word or a passphrase which is then used to control access to the conversation. There are two kinds of threats against this kind of authentication: (1) attackers guess a weak password, or (2) attackers intercept the password while users exchange it. As a countermeasure, this concept suggests passwords and not allow users to exchange the password in-band. The Socialist-Millionaire-Protocol (SMP) which is currently used by the Off-the-Record protocol [11] can be used to implement this concept. It is an online protocol that provides a zero-knowledge proof that both parties possess the same secret without actually disclosing any information about the secret. As Alexander et al. [11] mentioned, even secrets with a very low entropy are secure against offline dictionary or brute-force attack. Boudot et al. [32] provides a full security analysis. Based on this scheme we propose a prototype utilising the concept of *combination locks*, where all conversation participants have to set their combination lock to the shared secret before they can join the conversation. We continue to develop this proposal using iterative storyboard prototyping in the next section.

Picture-based → Selfies. Several participants mentioned that sending each other pictures of themselves establishes trust in the conversation partner's identity. Since the pictures are transmitted in-band, a targeted attacker could manipulate the pictures in real-time or use past pictures for authentication. However, the concept would provide security against simple large-scale attacks such as the crackdown on IronChat by the Dutch Police in 2018 [108]. This method encodes information about key fingerprints into gestures. Therefore, recipients of those pictures verify not only the identity of the person they are talking to, but also that a person uses the same key material as them. Adoption of this authentication method could be negatively impacted if users need to compare more than five pictures. Assuming a gesture alphabet of size 32, it is possible to compare 15 to 30 bit of the key material using this approach. Based on this scheme we propose a prototype utilising the concept of *selfies*, where conversation participants have to provide a series of selfies to others to authenticate themselves. We continue to develop this proposal using iterative storyboard prototyping in the next section.

Institutional → ID cards. Most participants were familiar with ID cards as a way to authenticate other people. The process of showing each other an ID card is well-established in the offline world and we consider this mode of authentication well-aligned with common approaches to key authentication. Messengers that implement this need to provide a user interface that mimics an ID card that users can show each other. Since users have to meet in person to check their ID cards (simulated in the UI), attackers cannot influence the authentication process as long as the devices are not compromised. The security is based on key verification, which works by comparing the encryption keys of the conversation partners over a secure channel. This is achieved by integrating a QR code of the key fingerprint into the simulated ID card (refer to (c) in Figure 8.6). In this case, they meet in person and the compare keys automatically as suggested by Tan et al. [267]. Based on this scheme we propose a prototype utilising the concept of *ID cards*, where conversation participants have to verify the others' simulated ID card

in the messaging app. We continue to develop this proposal using iterative storyboard prototyping in the next section.

8.6. Iterative Storyboard Prototyping

Narrowing down the design space resulted in three design concepts with corresponding threat models and security mechanisms. We developed detailed storyboard prototypes for each of them using Sketch, a vector graphics editor that supports user interface prototyping. Each storyboard starts in an unauthenticated state and ends in an authenticated state. Each state of the user interface, i.e. changes after each tap, is included as its own still image in these storyboards.

We used an iterative approach with alternating field-work and revision of the prototypes. During the field-work, we explained a scenario to the participant and conducted a walk-through of the storyboard prototype. For each element of the storyboard we asked the participants to describe what they see and what they would do next. We noted hesitation and obvious confusion as an implicit feedback in the field notes and asked the participants explicitly about it afterwards. After the walk-through, we asked questions to explore the participants' understanding of the prototype. We asked them (1) how they would describe the authentication process to a friend, (2) to describe how the process affects the security of the conversation, (3) to rate trust in the additional security on a 10-point Likert scale. After reflecting on the security of the process we encouraged all participants to redesign all storyboard prototypes, we emphasized that they could change the design, the phrasing, the order of the screens, or add additional screens. Participants marked their suggested changes directly on the printed storyboard prototypes. We noted all answers in the corresponding field notes. Since the prototyping sessions were conducted in a different language the researchers translated the field notes and the prototype annotations prior to the analysis. We used an online survey to collect demographics (age, gender, study program, type of occupation, and responses to the affinity for technology interaction (ATI) scale [88]).

We extracted misconceptions, most frequently made suggestions, and improvements without negative side-effects from the resulting feedback. We used those suggestions and adapted the storyboard prototypes accordingly. After we improved all storyboard prototypes we recruited new participants in the field and started a new iteration of the prototyping approach.

Recruitment and participants. Two members of our department's administrative staff participated in a pilot-study. Their results are included in the final result as no changes were made, except for minor adjustments in the field notes template. We recruited participants around our university's main plaza. All participants provided verbal consent. We compensated them for their time with a candy bar. $N = 18$ people provided feedback in the storyboard prototyping process. The participants' age was between 17 and 50 (m=26.545, sd=9.136), 33.3% of them were women and 66.6% of them were men. About a third (27.3%) of the participants were not studying or declined to answer the question about their field. Only one participant studied a STEM program. One third of the participants were students, a third was employed, 26.7% were out of work, and one

participant was self-employed. The average ATI score was 4.128 (sd=0.763), which indicates an affinity for technology interaction slightly above the average population score (3.5 [88]).

8.6.1. Results

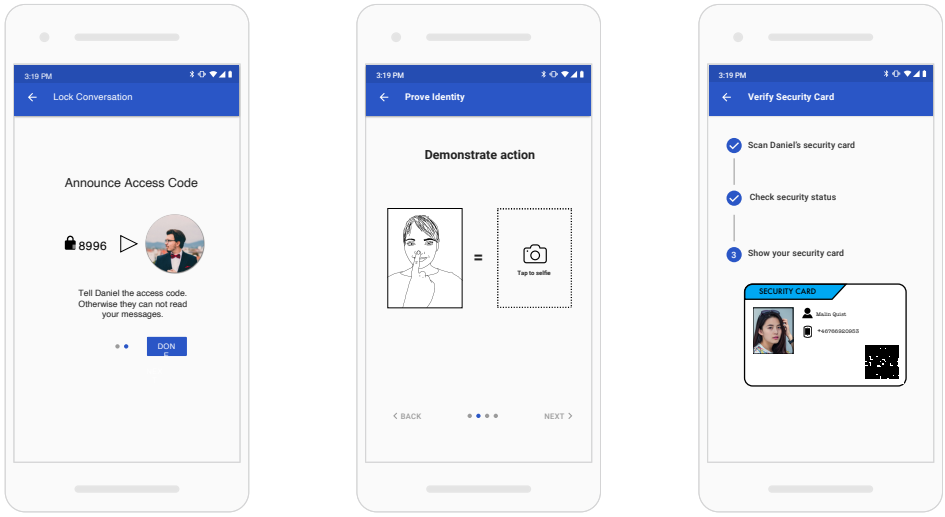
Perceived security and trust. Since authentication ceremonies often need the cooperation of two users it is necessary that users are able to describe ceremonies in simple terms. The majority (13) of participants provided short and functional descriptions, the others either responded with step-by-step explanations (4) or an explanation why they would not use such a ceremony (1). Eleven participants said that the ceremony had a positive effect on security, four did not know, two thought it would impact security negatively, and one would not use such a ceremony. To quantify the perceived security we asked participants to rate the increased security on a scale from 1 to 10. The selfies based prototype received the highest average score of 7.08, the combination lock and the ID card prototype received a lower but similar score (6.08 and 6.2). When we asked about the participants' reasons for their assessment, some described the strengths of an approach "being hard and time-consuming to fake the process" (P10) or remained cautious "one can never be entirely sure".

User-reported usability issues. The participants' feedback contained two categories of problems: (1) they wanted either more information, or (2) they wanted to improve the user interface flow. In the first category, the warning message that notifies them was the most common topic. Participants wanted to change its placement, color, the information contained, or the kind of buttons included in it. Another major concern was the kind of visualisation of a successful authentication: participants suggested different colors, symbols, or obvious messages to do that. In the second category, the user interface flow, participants wanted to simplify and streamline the ceremonies. Their suggestions were concerned with reducing the amount of actions they had to do manually. Additionally, they wanted more information on their progress and wanted to change the visible buttons to make it easier for them to navigate through the ceremonies. Two participants also wanted to change input fields, because they felt more comfortable with pin entry pads than with number wheels.

Social and cultural issues. For the authentication prototype based on personal pictures, we selected an alphabet of 32 gestures. We selected gestures that are easy to do with one hand, that are easy to recognize, and that do not have any negative political or insulting meanings. However, during the storyboard prototyping iterations we found that our curated list still contains gestures with negative meaning depending on cultural context.

8.6.2. UX Expert Feedback

Asking potential users to give design feedback is helpful because it gives them a concrete method to describe what they understand, what aspects irritated them, and how they would resolve those problems. In most cases, we cannot implement the participants'



(a) Combination lock based prototype (Malin).

(b) Selfies based prototype (Malin).

(c) ID card based prototype (Malin).

(d) Combination lock based prototype (Daniel).

(e) Selfies based prototype (Daniel).

(f) ID card based prototype (Daniel).

Figure 8.6.: Impression of the prototypes' authentication interaction between the fictional characters Malin (blue) and Daniel (magenta).

design suggestions without further consideration. After a total of $N = 18$ participants and two iterations of storyboard prototyping, we did not encounter any new types of problems or design suggestions from our participants. We involved a UX expert to help us improve the user interface design further. The UX expert that we recruited from our institution has several years of experience in UX design for different research facilities. In the first meeting, we presented and explained the storyboard prototypes, provided information about the design process, and discussed the feedback we received from the participants. After the meeting, the UX expert annotated all storyboard prototypes in detail and gave suggestions for improvements. In the second meeting, we discussed the annotated storyboard prototypes and resolved misunderstandings.

The UX expert gave feedback on the design of the authentication ceremonies themselves, as well as the motivational cues in the messenger's interface. Regarding the authentication ceremonies, we implemented the following suggestions: a radical reduction of text on each screen, reduction of visual noise in the interface, consistent placement of icons, and communicating only one aspect per screen with consistent use of progress visualization. They also suggested using imagery and animations to communicate that security "happens" in the background, similar to the approach by Distler et al.[63]. However, we did not implement this suggestion since we could not decide on fitting imagery and our paper-based prototyping approach does not work well with animations.

The UX expert also provided feedback on the motivational cues in the messenger's interface that lead up to the authentication ceremony. We implemented their suggestion to visualize the security status before and after security actions in a consistent manner. Additionally, we discussed several issues with the security warning message. The UX expert suggested avoiding the advertisement-banner effect by integrating the warning into the conversation itself. The messenger could force users to pay attention to the warning by requiring interaction, such as a slider or a finger tap, to access information. While we did not implement these suggested changes to the warning message, we consider them an interesting future research direction.

At the end of the second meeting, the UX expert suggested several sources of design inspiration, helpful books, and design tools that could help in the future. Figure 8.6 shows impressions of the three resulting prototypes.

8.7. Evaluation

During the collaborative design workshops, we found that (1) participants who understood the purpose and consequence of the ceremony were willing to invest an additional effort for some of their contacts, and (2) some participants felt reassured about the security if they were able to participate in the security process. Therefore, a good authentication ceremony provides users with an intuition about the security it provides in different situations, and also increases the users' perceived security.

Procedure. We conducted a between-subjects online survey ($N = 131$) on Amazon MTurk and randomly assigned participants to one of four conditions. The four conditions

Table 8.1.: The quantitative responses (SUS, UEQ-S, perceived security, and security ratings against threat models) in all four conditions.

	Combination Lock		Selfies		ID Cards		Signal's	
Participants	35		38		30		28	
	avg.	std.	avg.	std.	avg.	std.	avg.	std.
SUS	55.93	15.12	50.07	12.73	54.25	15.48	52.14	16.39
UEQ-S	1.42	0.83	1.24	0.90	1.12	1.02	1.26	1.05
Perceived Security	5.60	1.13	4.92	1.35	5.26	1.46	5.39	1.23
Threat Models	4.84	0.76	4.24	1.13	4.32	1.07	4.81	0.81

Table 8.2.: Separate one-way univariate analyses (ANOVA) on the individual outcome measures.

Outcome Measure	Sum of Squares	*df*	Mean Square	F	p	partial η^2	
SUS	694.6	3	231.54	1.0186	.77	.0235	
UEQ-S	1.437	3	0.4790	0.5187	.77	.0121	
Perceived Security	8.841	3	2.9471	1.7035	.51	.0387	
Threat Models	9.972	3	3.3241	3.4762	.07	.0759	.

Sign. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

include the three developed prototypes and Signal's current authentication ceremony in the same presentation format and design language as the other prototypes.

At the start of the survey, we presented the same messaging-related scenario to all participants. This scenario introduces a threat model: the potential risk of losing one's job if the messages are intercepted or sent to the wrong person – this was chosen since (a) it could have serious consequences, (b) it is a common enough situation s.t. participants are able to immerse themselves into it. The scenario also introduces the notion of an authentication ceremony and links the participants to a randomly assigned condition. Each of the four conditions shows a storyboard of an authentication ceremony. Participants may click anywhere to receive an indication of the clickable features in each still image. After the participants complete the authentication ceremony, we link them to the evaluation survey. The survey covers the following four quantitative measures: (1) *SUS* (10 items): evaluation the prototypes' usability using the Systems Usability Scale (SUS), (2) *UEQ-S* (8 items): evaluation of the prototypes' user experience using the User Experience Questionnaire (UEQ-S), (3) *Perceived Security* (1 item): participants' rating of their conversation's general security after completing the ceremony [7-point Likert scale], and (4) *Threat Models* (5 items): participants' rating of security that their ceremony provides against five attackers with different capabilities [7-point Likert scale] (detailed items Section E.3.3 of the Appendix). Even though users' perceptions on security or the threat models are not necessarily accurate, they will affect how often and for which purpose they will use authentication ceremonies. We also collected qualitative information on (1) the participants' reasons for their perceived security, and (2) the participants' contacts they would consider authenticating in the future.

Additionally, we asked for information on messengers used, whether they have seen an authentication ceremony before, and if they would conduct an authentication ceremony in the future. To show the validity of our sample, we also measured the affinity for technology interaction (ATI) scale. The full questionnaire is included in Section E.3.3 of the Appendix.

Analysis. We hypothesized that participants who experience one of the three developed prototypes will have an increased perception of security and an improved understanding of the type of threat models they protect against – when compared to Signal's current authentication ceremony.

We use a MANOVA to measure the global effects of the choice of prototypes on the four outcome measures (*SUS*, *UEQ-S*, *Perceived security*, and *Threat Models*). We apply separate univariate analyses to measure the effect of the prototypes on each of the outcome measures. In case of a significant statistical effect on the outcome measures, we use pairwise planned contrasts between the developed prototypes and the control group to understand which prototypes is responsible for this effect. The required sample size for a medium effect size of $f^2(V) = 0.625$ and a *power* $= 0.95$ is 144 participants, which was our lower bound recruitment goal. We used *open coding* to analyse the free text responses to the qualitative questions. One researcher coded all answers, thereby creating the initial codebook consisting of 15 codes. A different researcher used this codebook to code all answers independently. This resulted in an inter-rater agreement of Cohen's $\kappa = 0.69$, which is a *satisfactory* agreement.

Recruitment and participants. We conducted a pilot test with two participants to refine our survey and to determine participant compensation based on completion time (10 resp. 15 minutes). We implemented their minor suggestions for improvements in the final version of the study.

For the final study, we recruited participants on Amazon MTurk. We required a 99% approval rate for past assignments. We paid each participant USD 2.50 which results in a USD 10 per hour wage. We received a total of 217 completed questionnaires. 82 participants were already familiar with one of the study's conditions, because they had seen Signal's authentication ceremony before. Four participants failed the two Likert scale attention check questions (taken from Huang et al. [143]). After removing them, our final dataset consists of $N = 131$. The participants' average age was 30.58 ($sd = 8.10$). About two-thirds (63%) of them were men, about one-third (36%) were women, and one person (1%) was non-binary. The majority (66%) of participants had a college degree, 12% had a vocational degree, 19% completed high-school, and 3% did not complete high-school or preferred not to say. The average ATI score (3.58) is near the expected average of 3.5 [88]. Participants reported using on average 2.48 ($sd = 1.24$) messengers (most commonly WhatsApp, iMessage, and Telegram). 12.97% of them would authenticate most or all of their contacts. The rest would authenticate on average 5.16 ($sd = 2.94$) of their contacts. Table E.1 in the Appendix presents all collected demographic information.

8.7.1. Results

Participants experience one of the prototypes (Combination Lock, Selfies, ID Cards) or the reference prototype modeled after Signal's ceremony. Each of the four groups had 28 to 38 participants. The average affinity for technology interaction (ATI) in each group was similar and ranged from 3.53 to 3.67, whereby 3.5 is the expected average [88].

Quantitative responses. Participants evaluated the prototype's usability (SUS) and user experience (UEQ-S). The prototypes' SUS scores ranged from 50.07 to 55.93. SUS scores above 71 are considered *acceptable* and scores below 51.7 are considered *unacceptable* [21, 246]. The prototypes' UEQ-S scores ranged from 1.12 to 1.42. Values range from -3 (horribly bad) to +3 (extremely good), whereby results above +0.8 indicate a *positive evaluation*.

To understand the participants' perceptions of the prototypes' security benefits we asked them to rate their perceived security on a 7-point Likert scale, and rate (again on a 7-point Likert scale) the respective ceremony's effectiveness against five specific threat models. The prototypes' perceived security ranged from 4.92 to 5.60. The participants' rating of security against the five threat models ranged from 4.24 to 4.81, whereby a rating of 7 indicates confident and correct evaluation of security in all cases. The calculated Cronbach's alpha for these five threat models items is $\rho_T = 0.83$ which indicates *good* ($0.8 < \rho_T < 0.9$) internal consistency. Table 8.1 provides an overview of the resulting measurements.

Statistical tests. Using Roy's largest root, there was a significant global effect of the experienced authentication ceremony on the outcome measures, $\theta = 0.09$, $F(4, 126) = 2.702$, $p = .03$, $\eta p^2 = .0789$[1].

To find out which of the four outcome measures (*SUS*, *UEQ-S*, *Perceived Security*, and *Threat Models*) is affected by the different authentication ceremonies we ran separate one-way ANOVAs on them. After applying Holm-Bonferroni correction, these separate univariate analyses revealed a marginally significant effect on the outcome measure *Threat Models*, $F(3, 127) = 3.48$, $p = .07$, $\eta p^2 = .0759$[1]. Table 8.2 shows the results of separate ANOVAs for all outcome measures.

We used planned contrasts (with a separate linear regression model) between the different ceremonies to find out which of the ceremonies affected the outcome measure *Threat Models*. These planned contrasts with Signal's authentication ceremony revealed that (a) the combination lock ceremony significantly improved the outcome measure *Threat Models*, $t(127) = 1.982$, $p = .049$ (one-tailed), and (2) that the selfies-based ceremony significantly worsened the outcome measure *Threat Models*, $t(127) = -2.194$, $p = .03$. Table 8.3 shows the result of all planned contrasts on the outcome measure *Threat Models*.

Qualitative responses. We asked all participants to provide a reason for their perceived security rating. The responses to the combination lock-based prototype included rather detailed explanations, many of them stating that the security of the conversation is

[1] Calculated partial eta squared s.t. $.06 < \eta p^2 < .14$ are considered *medium* effect sizes.

Table 8.3.: Planned contrasts on the outcome measure *Threat Models* using a separate linear regression model.

Contrast	Estimate	Std. Error	t	p	
(Constant)	4.55231	0.08606	52.895	<.001	***
CL vs. SI	0.28769	0.14515	1.982	.049	*
SE vs. SI	-0.31021	0.14138	-2.194	.03	*
ID vs. SI	-0.23231	0.15279	-1.520	.13	

Sign. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Abbr.: CL = Combination Lock, SE = Selfies, ID = ID Card, SI = Signal's

based on the knowledge of the access code: *"Only the person with the code can access the conversation."* (P75). One participant thought 4-digits might be too short for security and another one thought about the difficulties to distribute the shared knowledge in a secure manner. Participants who saw the selfies-based prototype felt reassured by the pictures of the communication partners, but commonly had a problem to connect this with confidentiality: *"I can understand the confirmation of the person on the other end. I can see how the ceremony confirms the party you are communicating with. I have trouble understanding how the data is further secured in the space in between."* (P85) The ID card based prototype resulted in misconceptions about the security implications of the authentication ceremony. Participants assumed that it could be stolen or copied, contrary to the technical reality.

Necessary additional interactions conveyed a feeling of improved security. Several participants connected the action of scanning QR codes with added security: *"Due to the double QR code verification (the two participating phones mutually scanning each other)"* (P25). However, since one participant mentioned that scanning a QR code does nothing for security, this reasoning is potentially shaped by prior experience or knowledge. Many participants also explained why they did not fully trust the authentication ceremony: *"However, I do know that there will be that small section of individuals that would still be able to hack this system if they really wanted to."* (P81) Some of those participants described threat models (such as a conversation partner forwarding information) and others just overestimated the capabilities of attackers.

8.8. Limitations

Improving authentication ceremonies is a long-standing challenge and several approaches have already been applied to it with limited success. We explore how a different design approach, namely *User-Centered Design*, is applicable in this case despite its drawbacks. We hope that the lessons we learned during this study start a discourse on the benefits and pitfalls of applying this design approach to security.

Necessarily, *User-Centered Design* studies focus on the goals and requirements of users – who then severely influence outcomes. We recruited participants for our collaborative design workshops and iterative storyboard prototyping sessions in stable, economically

rich countries of the global north. This population faces few threats in their daily life – which influences the resulting prototypes. These participants were also not security or design experts, and we did not expect them to come up with technical secure concepts. Instead, we explicitly included security and design experts in our design method. However, users are experts when it comes to their perceptions, values, intentions, and mental models – and User-Centered Design can help to incorporate this expertise into designs that work for users, not against them.

Many authentication ceremonies require in-person meetings even though users' need for additional security arises in the moment. Ensuring secure conversations requires planning these in-person meetings ahead of time, which might decrease the usefulness of the authentication ceremony to regular users. We did not prescribe one type of ceremony, since we did not want to restrict the participants' intuitions. Qualitative results (in Section 8.4 and 8.7) suggest that in-person meetings build trust in the security mechanism.

Collaborative design workshop participants need a grasp of the security issues in order to suggest solutions. We used a slide show (see Figure E.1 in the Appendix) to explain these issues in high-level terms to them. The resulting prototypes did not include these explanations, since (a) we cannot expect users to read them (outside of a lab environment), and (b) the user experience should communicate the prototypes' security implications.

Our evaluation used a scripted online experience instead of a lab study. Previous work [12, 37] found that remote asynchronous usability testing discovers fewer usability issues than lab testing, which is offset by easier participant recruitment. However, our main concern is not usability testing but rather the users' comprehension of the ceremonies' security implications. Encouragingly, Wu et al. [298] used a similar approach to evaluate their authentication ceremony designs. We assume that the effects found in a scripted online experience should be even more pronounced in a real-world scenario.

8.9. Discussion

We begin this section by discussing lessons learned from applying a user-centered design process to a well-researched security problem in secure instant messaging, and then continue to discuss unexpected findings and how these results fit into the existing related work.

8.9.1. Methodological Lessons We Learned about User-Centered Security Design

We explored how user-centered and participatory design techniques can be applied to a well-studied security problem. From this exploration, we learned how some aspects worked out better than we expected and which aspects we would have approached differently in hindsight.

Framing of the design problem affects the entire design process. The framing of the initial design problem impacts the entire design process and its outcomes. It specifies which strategies and which kinds of solutions are suitable for the problem at hand – and who

should contribute in which manner they are allowed to contribute. Hence, this framing should be chosen carefully and explicitly.

We based our design problem entirely on previous research on authentication ceremonies in secure instant messaging. This was possible since this niche-problem has already been studied extensively and the continuing issues are well-documented. This is a valid and common approach in research, however, it also means that assumptions from previous research also influenced our work. A different approach of framing a design problem in secure instant messaging could have involved asking participants with an increased reliance on security (e.g. activists, members of oppressed minorities, health-care workers, sex-workers, ...) about their day-to-day uncertainties and fears about secure communication.

Explicit choice of participants is necessary. Working with a universal definition of an unmarked user has been a problem in the first wave of HCI [47]. Similarly, it is a commonly observed issue in User-Centered Design that designers tend to imagine users that are similar to themselves [48, 207]. This erases the challenges of groups that are unlike the involved designers, and reinforce the societal power hierarchy.

To avoid unmarked users in design studies for security, we suggest keeping the following groups in mind: affected users, idealistic users, and non-users. Ermoshina et al. [75] differentiated between users with specific and concrete threat models, who were consequently invested in learning and using security tools, and users with very abstract threat models who had an interest in security tools but used them more for emotional and idealistic reasons than fearing concrete negative effects. Since the security and usefulness of some security tools rely on the number of total users, it is equally important to focus on the attitudes and requirements of non-users [250]. All three groups need to be involved to build widely-deployed security mechanisms that provide meaningful security against various kinds of users' threat models.

In our design process, we did not focus on users with specific threat models in mind, instead, our recruitment efforts yielded mostly idealistic users with rather abstract threat models. In hindsight, this allowed us to focus more on user comprehension and motivational aspects – assuming that affected users are usually motivated and more concerned with issues of usability. Our choice of participants also enabled us to find social and cultural aspects to the design of authentication ceremonies: (1) that requesting an authentication might seem like a sign of distrust in the conversation partner, (2) that they might feel pressured to provide a reason for their authentication request, or (3) that they might be expected to explain how the ceremony works when they actually do not know. One participant of the iterative storyboard prototyping sessions did not like the concept of the selfies-based prototype. They thought that sending selfies of themselves comes across as narcissistic or that some of the gestures are inappropriate depending on the cultural context. Even though our selection of participants worked out well in our case, we would approach the question of suitable participants with more care in future design studies.

Clear expectations from participants' involvement in the security design process. In the beginning, we did not have a clear expectation of our participants' conceptual designs –

misleading us about the work that would still remain in subsequent design stages. This early learning experience led us to include an explicit security evaluation and feedback from a UX expert into the design process.

Designers need to have clear expectations of which expertise participants can bring to the design process. They account from their lived experience to inform the design process about problems with existing approaches, current workarounds, and the users' threat models. Participants can also provide intuitions about security procedures they would expect to see or that they would find especially convincing. Our collaborative design workshops provide evidence for that, seeing that the resulting prototypes are novel and engaging. Additionally, participants can provide insights on secure experiences after they experience them. This works either by using traditional interview techniques or redesigning the low-fidelity prototypes themselves. The latter approach is especially useful when participants have difficulties expressing their desired changes verbally.

However, we need to stress again that participants cannot provide expertise on security, usability, or user experience design. The results from our collaborative design workshops included several designs that were not secure and very hard to implement securely at all. Consequently, expertise in these areas has to come from other involved parties.

Focus on qualitative evaluation in the prototyping phase. As Greenberg et al. [118] noted that HCI papers tend to quantitatively evaluate early designs even when a qualitative approach would be more appropriate. Throughout our design process, we observed that the qualitative, rather than the quantitative, evaluation of our prototypes provided more thorough and actionable information about their underlying issues and benefits. A particularly relevant example is the participants' mental association of the user experience with the achieved levels of security. While we could tell from quantitative measures that participants believed in a prototype's security, we required qualitative data to understand the reasons for these beliefs. These reasons were sometimes unintended, unexpected, and consequently, insightful. In the future, we would focus more on qualitative evaluations of our early designs instead of comparing prototypes quantitatively early on.

8.9.2. Outcomes from our Endeavour to Design Appropriate Authentication Ceremonies

Unexpected findings regarding our resulting prototypes. In hindsight, two unexpected findings add necessary context to the resulting prototypes and their evaluation results: (1) Qualitative results suggest that participants strongly associate the act of scanning QR codes with security – which means that QR codes potentially evoke a perception of security even without understanding the security mechanism itself. This association would have influenced the evaluation of the ID card prototype and Signal's current ceremony. (2) Reviewing the documentation from the collaborative design workshops, we note that participants either understood MitM attacks as an impersonation attack or an interception attack – both interpretations are incomplete but correct. Consequently, we received some suggestions that protect against impersonationand others that protect against interception. Qualitative results indicate that participants exposed to the *selfies* or the *ID card* prototypes were confident about their contact's identity but unsure

how the procedure protects against interception – even though it technically would. In contrast, participants that used the *combination lock* prototype understood how it provides security even though some thought 4-digit combinations were insufficient.

The iterative storyboard prototyping uncovered that adapting the messengers' UI flow (details are in Section E.2.2 of the Appendix) can provide security guarantees. The combination lock prototype locks users out of a conversation until they have entered the correct access code. This design choice was consistent with the workshop participants' conceptual ideas. Since users in these scenarios are required to authenticate, the transmission of unauthenticated messages indicates an ongoing MitM attack. Such a guarantee based on messengers' UI flow is a clear improvement over the state-of-the-art.

Increasing adoption rates. The comparative evaluation of our prototypes did not find improved usability or user experience. However, both are prerequisites for increased adoption rates. We found three different explanations for this, each leading to another remedy. Our prototypes could have improved the usability and user experience in minor ways not detectable with our study's number of participants. Such potential minor improvements could be found with more participants. It is also possible that our online user experience does not compare well to the physical counterpart. Repeating the evaluation with high-fidelity Android-based prototypes could provide more meaningful results for these two measures since the experience would resemble real-world circumstances. We also consider the possibility that our comparatively short user-centered design process might not be suited for gaining usability improvements. Other complementary design approaches, such as activity-centered design, could be applicable and improve usability and user experience.

A broad social acceptance of authentication ceremonies is necessary to increase adoption rates. Social acceptance might be an issue for our prototype based on sending selfies that show specific gestures. Sending pictures containing silly gestures or even selfies to contacts might be inappropriate depending on the cultural and social contexts. Hence, business-related authentication will require different kinds of pictures than authentication amongst relatives and friends.

Aspects such as discoverability, motivation, and nudges were not our primary research goals. Nevertheless, they are crucial aspects of increasing adoption rates. Users could either plan their security ahead, regardless of their immediate requirements or they (unexpectedly) require increased security in the moment of use. Users that employ the first approach have a security motivation but might need reminders at convenient times (motivation and context-sensitive nudges). The second approach requires an understanding of the available security mechanisms and that users find and use these mechanisms in an appropriate time-frame (understanding, discoverability, and usability). Most authentication ceremonies rely on a planned security approach because they often rely on in-person meetings. However, ceremonies need to support both approaches to be useful in more situations and for more types of users.

Our results in the context of related work. When Vaziripour et al. [284] let their study participants authenticate their communication partners without telling them about authentication ceremonies, they used several techniques that also came up during our

collaborative design workshops. Namely, *send pictures*, *recognize video*, *recognize voice*, and *shared knowledge*. This overlap of techniques suggests that these social approaches to authentication might generalize to a larger population – a finding that could inform future authentication ceremonies.

Necessary additional interactions in the form of authentication ceremonies boosted some of the participants' perceived security. Fully automatic authentication (e.g., with CONIKS [192]) would make these ceremonies superfluous, thereby reducing the perceived security for this population. However, most participants stated that they would only authenticate friends, partners, and family – making automatic approaches useful for other contacts.

8.10. Conclusion

Authentication ceremonies in secure instant messaging are a well-researched security problem [138, 254, 284, 285, 283, 282, 267, 298]. A few different approaches (applying either systems design or activity-centered design) have been explored to improve their usability and adoption rate. Our user-centered design approach is based on the assumption that fundamental improvements of these ceremonies require rethinking the entire design from the user's perspective. We used a four-stage design process including collaborative design workshops, selecting viable candidates, iterative storyboard prototyping, and a mixed-methods online evaluation. Even though the quantitative comparison of our prototypes did not reveal usability or user experience improvements, we found that one of our prototypes increases the users' comprehension of the ceremonies' security benefits.

We also learned several important lessons from applying user-centered design to security problems: (1) Participants have important participatory roles in the security design processes. Mainly framing the design problem regarding threat models and social aspects, informing designers and security experts about their intuitions on convincing secure experiences, and improving prototypes with their iterative feedback; (2) The choice of participants needs to be explicit and consistent. The constructed notion of the "universal user" could be combated by,e.g., differentiating users according to the details of their threat models (concrete – abstract), or by their current (non-)use of security features; and (3) Focusing on qualitative evaluations is necessary to understand if participants correctly associate the user experience with the achieved security levels.

Part IV.

Discussion and Conclusion

9 Discussion

I started this thesis by recognizing the possible impact of secure experience on users' understanding of security implications and their informed choice about which tools provide appropriate security. However, with each of my case studies, I uncovered more of the underlying complexity of the secure experience, making it challenging to design. I also identified deceptive secure experiences, i.e., instances of security and privacy theater, and want to discuss their origins.

Based on what I discovered through my case studies, I propose a method to embed secure experience into a design process to facilitate future awareness of the concept and, subsequently, adoption.

Since this is a methods-driven thesis, I will also reflect on my choice of case studies, the generalizability of my results, and the problems and opportunities of applying different methodological approaches to security research problems.

9.1. The Complexity of Secure Experience

In the case studies I found several underlying factors that impact the resulting secure experience, the user interface, users' mental models and threat models, and the social norms around security practices.

User Interaction

The most obvious factor of secure experiences – and the one I expected – was how the interface was designed. The literature review of our paper *"Stop the Consent Theater"* (Chapter 3) showed that websites and also larger consent management platforms used manipulative user interface design in their consent notices. The consent notices used several psychological tricks in their user interfaces: overwhelming users with choice, which leads to disengagement [172]; suggestive button placement and design to manipulate visitors into accepting more cookies than they actually would like; and categorization of data collection purposes as "essential" or "legitimate" that do not match visitors' expectations.

In our paper on user perceptions of anti-stalkerware (Chapter 4) we found different user interface design elements that convey a reassuring security experience: One of the apps had a UI element that was constantly visible, which gave users the impression of a

guard watching over them; the app's regular maintenance actions were made visible to communicate that the app was active, even if it did not identify potential threats; and immediate UI reactions to user actions communicated vigilance.

In our paper on the Tor over VPN practice (Chapter 5), we studied how combining different security tools affected the secure experience. Since users had a secure experience with both tools independently, they used additive reasoning that increased their overall secure experience – because neither tool's user interface suggested anything otherwise.

Mental Models of Technology

Users' security perceptions depend on their mental model of the underlying technology. To understand how a security practice reduces risks, they also require a functional understanding of what software does, i.e., a mental model. I encountered this issue in our work on the Tor over VPN practice (Chapter 5) since it is unintuitive how a VPN and the Tor browser provide anonymity to its users. Because of its branding as the Tor *browser*, some users thought it was only a browser with privacy-enhancing features. Such a mental model of the Tor browser leads to a rational conclusion that a VPN is necessary to protect the communication on the network, leading to a secure experience of combining Tor and VPN. In our work on cookie consent notices (Chapter 3), we also identified an issue with associated mental models. Explaining the current system of tracking online behavior for personalized advertisements is difficult. Consequently, creating functional mental models is unlikely for web users. Making matters worse, online platforms benefit from users not understanding the details of online ad-tracking since they get revenue from the advertisements. Hence, they have no interest in explaining it in more detail than they legally need to. For users, this means annoyance rather than a secure experience.

Threat Models

Threat models influence the secure experience by focusing users' attention on protection from a specific threat. In our work on anti-stalkerware (Chapter 4), this becomes clear because the studied reviews often contained references to technology-mediated surveillance by others or positive and negative experiences over time with the anti-stalkerware. Both types of experiences understandably focused attention on specific kinds of threats and, consequently, software functionality. These surveillance-based threats are usually closely intertwined with patriarchal power in intimate partner relationships. In our work on the Tor over VPN practice (Chapter 5), many practitioners did not have a specific threat model. The lack of threat models allowed them to attach their secure experience to other factors, such as the prevalent social norms in a community.

In general, I did not find many instances in our research where we identified concrete and realistic threat models, i.e., potential threats people wanted to protect themselves from. While this may not correspond to researchers' common ideal of informed users who make conscious decisions, it reflects users' day-to-day reality: having other priorities than keeping up with an evolving threat landscape while still expecting safety and security by default.

Social Norms around Security Practices

If users are unsure about what security practices are appropriate, it is understandable (and only natural in a cooperative society) to seek advice from others. Either by asking for recommendations, experiences with security practices, or observing what others do in similar situations, i.e., a *social proof* [51].

We observed this effect in our paper on the Tor over VPN practice (Chapter 5). The security and anonymity effects of combining two network-level privacy tools, such as VPN and Tor, are hard to judge, even for technical experts. Our analysis of newspaper articles, marketing material, and social media discussions showed that users align their usage with the observed practices of others they perceived as knowledgeable. Similarly, in our paper on anti-stalkerware (Chapter 4), people relied on recommendations from family members and other trusted sources they perceived as knowledgeable – supporting a view that security practices are adopted at least partially for social reasons.

Sometimes, security behavior is adopted to become an accepted member of a security-minded user group – the threats that necessitate the practice are less critical. We found references to such a mindset in our research on the Tor over VPN practice (Chapter 5) that emphasized belonging to a group of security enthusiasts over specific threat models. In my autoethnography on authentication ceremonies in end-to-end-encrypted messengers (Chapter 7), I also described practicing authenticating encryption keys primarily as a social cue of belonging to a group. This social aspect of security practices is also documented outside IT security. Lingel et al. [168] investigated the information secrecy practices of a punk rock subculture in New Brunswick, New Jersey. They described how the localized social context and group dynamics drive technological use independent of relevant threat models.

Hedonic and Pragmatic Qualities of Secure Experience

User experience divided into *hedonic* and *pragmatic* qualities [130]. The *hedonic* quality refers to the product design dimensions without obvious relation to task-related goals, such as originality and innovativeness. The *pragmatic* quality refers to design dimensions that are more directly related to the task-related goals, such as usability, effectiveness, and efficiency. See Section 2.3 in the Background and Related Work Chapter for a more detailed explanation of the experience of technology.

Of the factors that I described above, the user interface factor is mostly *hedonic* in nature because it has the ability to reassure users about the security, unrelated to the actual task goals. The other factors, mental models, threat models, and social norms, mostly affect the *pragmatic* quality of secure experience, i.e., they inform the task-related goals of which security practice is important and why. This difference became clear in our paper on the user-centered security design of authentication ceremonies in end-to-end-encrypted messengers (Chapter 8), where the designs focused on the *hedonic* qualities of interaction. In our qualitative evaluation of the resulting prototypes, we found that while the user interaction received good feedback, some participants reported not perceiving it as secure because they did not understand how it would add security to their communication. It did not align with their understanding of the underlying technology, which confused them.

Implications of this Complexity

The described complexity is what makes designing secure experiences in a holistic way intractable. Since it depends on a host of other background factors that can differ widely between groups of users and the experience can hardly be "designed" in the traditional sense, how they experience interaction is unique to each person.

Hence, achieving secure experiences is a matter of an iterative design process that incorporates feedback from people rather than a definitive list of design elements that can be included. The entire user interaction needs to align with (simplified but functional) mental models, threat models, and prevailing social norms. The overarching theme for designing user interfaces is one of *reassurance*, usually by giving the appearance of vigilance and quickly responding to user actions.

9.2. The Current State of Secure Experience

Currently, the concept of secure experience is not widely used. Which does not mean that no one has a secure experience. In fact, I argue that some of the stalkerware detection products we investigated in Chapter 4 *intentionally* designed the user interaction to provide a secure experience. This product category faces a marketing problem: In the best (and most common) case, there is nothing malicious to detect. However, if there is nothing to detect, users might start wondering if the product is working correctly or even working at all. To solve this conundrum, they are designing reassuring and secure experiences for monetary profit.

However, even without *intentional* design, users can still experience user interactions as secure – an *unintentional* secure experience. I argue that this applies to most types of widely used security mechanisms in some way. Our research on the Tor over VPN practice (Chapter 5) showed that a combination of security tools may result in an *unintentional* secure experience that was not planned. Related work [70, 245, 3] indicates that the focus on usability, which hides the complexity of security, decreases or removes the secure experience. Hence, users cannot tell anymore which tools provide security and which do not based on the user interaction design that *unintentionally* removed secure experience.

Whether we try to design them or not, users will have a secure experience. Because this concept has not been integrated into the security design process, many security and privacy tools communicate and endorse unintended security practices as a consequence. As a research community, we should try to be aware of the potential of secure experiences, investigate how they might affect the choice of security practices, and design them in a way that matches recommended security practices and protection from relevant threats.

9.3. Integrating Secure Experience into Design

Integrating secure experience into the design of security mechanisms is a challenge. Designers' influence over all of the described complex and interrelated factors is limited, making it intractable to suggest drop-in design elements to make software "feel" secure. Hence, I focus on a process-driven design method to integrate the concept of a secure

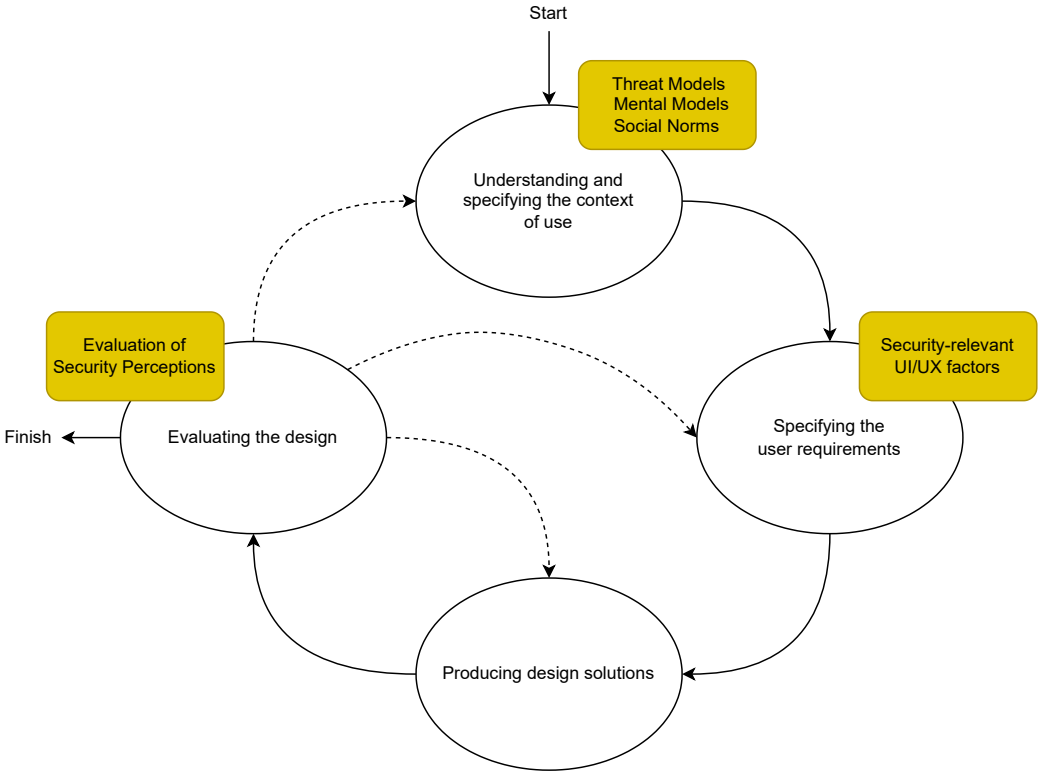


Figure 9.1.: A User-Centered Design Process (adapted from EN ISO 9241-210:2019 [74]) that integrates the identified factors of Secure Experience

experience. The resulting feedback loop in the design process gauges the design's effect on the resulting secure experience. This increases the intentionality of secure experience, i.e., making it less likely to deal with unintentional good or bad secure experiences, with the added benefit of increasing the awareness of the concept.

The well-established user-centered design process [74] is already an iterative process with feedback loops. It is a good candidate design process to incorporate secure experience. Figure 9.1 shows an overview of the proposed user-centered design process, adapted to the identified factors of secure experience. The design process has four iterative stages: (1) understanding and specifying the context of use, (2) specifying the user requirements, (3) producing design solutions, and (4) evaluating the design.

1. **Understanding and specifying the context of use:** The context of use depends on the threat models, the prevalent social norms around a security and privacy practice, and the potential users' mental models of the technology. Understanding these might be possible through a focused literature research of the related work or a fresh investigation in the form of a pre-study. Triangulation is one way of getting a good overview of the context of use: In our research on the Tor over VPN practice (Chapter 5), we combined behavior measurements, a user survey, and a qualitative analysis of the background information material to

that effect. Depending on the type of security and privacy mechanisms, it might be possible to use an autoethnography approach (Chapter 7) to understand the relationship between social norms and technology use in detail. Other approaches that I tested in this thesis include contextual inquiry at scale (Chapter 6) or the collaborative design workshop that we applied in our research on user-centered security design (Section 8.4) – it has the potential to reveal the context of use by eliciting tacit user knowledge. In any case, the iterative design process reveals inconsistencies and an unexpected context of use. However, understanding and specifying it as early as possible helps reduce the necessary time to viable design solutions.

2. **Specifying the user requirements:** As soon as the context of use is clear, this context needs to be translated to specific user requirements. Since security and privacy are secondary user interaction goals, these user requirements will mostly be focused on the primary interaction goal, e.g., communicating efficiently with friends. However, knowing the context of use, the requirements should also specify how to communicate a specific threat model and guide users in a way that makes sense with a simplified mental model of the underlying technology while respecting the prevailing social norms. The resulting user interface and interaction need to provide reassurance that the security goal is achieved and that the design solutions provide ongoing protection from the identified risks.

3. **Producing design solutions:** Producing design solutions works the same way as for other, non-security-relevant designs. In our research on the design of authentication ceremonies in end-to-end-encrypted messengers, we used iterative storyboard prototyping (Section 8.6). In the course of this process, we also learned to appreciate the benefits of working together with an experienced UX designer. I would suggest that design teams that mainly consist of members with a strong security background should try to collaborate with UX designers to complement their own experience.

4. **Evaluating the design:** Evaluating the designs for the resulting secure experience can use several different approaches, both quantitative and qualitative. In our research on user-centered security design (Section 8.7), we used a single-item scale that asked participants to rate the security. Distler et al. used a similar approach to evaluate the perceived security of encryption [61]. In one of their other works [63], they used a scale for psychological needs fulfillment that contained the psychological need for security. In the long term, future work could complement the established UX evaluation scale AttrakDiff [132] with additional items related to perceived security, privacy, and safety. In addition to a quantitative evaluation, a qualitative evaluation is crucial for an iterative design process. In our user-centered security design paper, we asked participants an open-ended question about how the prototype provided security. This explanation was useful to point out the origin of misconceptions and inform the next iteration of the design process.

The evaluation step in the design process will likely unearth new insights about the

context of use, incomplete user requirements, or misleading design solutions. Hence, the process, as shown in Figure 9.1, includes feedback loops to each of the previous steps. The design process can be stopped if the resulting design solutions provide an adequate secure experience or, at the very least, not a misleading secure experience.

9.4. The Ethics of Designing Secure Experiences

Careful readers will have undoubtedly identified an ethical issue with purposefully designing secure experiences: they have the potential to be deceptive. This mirrors ongoing issues with manipulative design (sometimes called dark patterns), for example, in cookie consent notices.

However, it is not like this thesis introduces deceptive secure experience. Capitalist market forces have already found out that they can use design to increase the perceived security of anti-stalkerware (Chapter 4) or use influencer-based marketing campaigns to change prevailing threat models, mental models, and social norms around the use of VPNs [10]. So, misuse of this concept is nothing new. Also, the current approach of ignoring the secure experience during the design process can also lead to deceptive secure experiences that misinform users.

I argue that a conscious effort is necessary to systematically detect deceptive secure experiences, criticize them in established products, and avoid designing them into new security mechanisms. In the ideal case, a secure experience could have a positive effect by aligning with actual security properties, which could reassure and inform users about effective security practices.

9.5. Methodological Reflections

One of the goals of the thesis was to investigate how existing methods from HCI could be adapted to security and how new approaches could help in designing secure experiences. I approached this using several case studies to understand the concept of secure experiences in detail. In this Section, I reflect on my experiences in this process, the validity of the approach, and the difficulties I encountered.

The suitability and generalizability of my choice of case studies

The point of choosing specific case studies in this thesis was twofold: First, to understand the phenomenon of secure experience in detail based on specific case studies that I could investigate. Second, to explore how existing methods could be adapted to security and privacy issues while understanding the potential barriers and the benefits of doing so.

Since this thesis primarily applies a qualitative research approach, I chose case studies where I expected a lot of rich insights into the concept of secure experience. I chose the thoroughly researched topic of cookie consent notices (Chapter 3), anti-stalkerware apps where I knew that intentional secure experience design was used (Chapter 4), and a combination of security practices (Chapter 5) with unclear security benefit. These edge cases of secure experience helped me understand different aspects of the concept, whereas more similar case studies would probably have resulted in fewer insights.

With every case study, the question of generalizability comes up at some point – which of these insights can be applied in other, maybe more general cases? As Braun and Clarke [35] explain, strict generalizability is usually not a meaningful goal in qualitative research. Instead, they advocate for a more flexible 'transferability' approach that leaves it to the readers to identify how the results apply to similar kinds of situations and people. To improve the transferability of the work in this, we provided a detailed description of the context of our case study and the circumstances and context of each interaction. While at least some aspects of each case study will be transferable to other research topics, like the possible misconceptions of combining different security and privacy tools, the generalizability of the case studies is limited.

In this Discussion section, I gave an overview of the parts of the results that are transferable to other use cases and offered ways to integrate the concept of secure experience in the research and design of security practices.

Difficulty of adapting methods from HCI to security

It is not necessarily hard to adapt existing methods from HCI to security. However, I usually encountered one of the three issues:

Problem 1: Security is not the primary interaction goal people are interested in. The conventional methods in HCI use research and design to target a specific primary interaction goal. However, security and privacy are not primary interaction goals; they are important but secondary goals. Users would like to communicate *securely* with their chat partners, and they want to maintain their privacy by keeping control over who gets access to their personal information while still being able to share personal pictures or distribute contact information without worrying about potential misuse. Hence, researching (and proposing design solutions) for security and privacy issues is only possible when bundled with a specific primary interaction goal. As a consequence, the security and privacy results are always tangled together with the primary interaction goal. Untangling and transferring these results to other applications and primary interaction goals is challenging.

However, since it is challenging to untangle different interaction goals, the research community might have to contend with researching security, privacy, and safety issues specific to each primary interaction goal – with limited transferability to other primary interaction goals.

Problem 2: Security interactions are usually rare. Since security interactions are not the primary thing that users are interested in, consistent design efforts have reduced them to the absolute minimum so as not to unnecessarily annoy users. This also means these interactions have become rare in everyday use. Hence, it is much easier to study how people interact with them in tightly controlled lab environments.

In contrast, modern third-wave HCI research and design focuses on everyday lived experience – a change that was heavily influenced by the prevalence of ubiquitous computing devices. In a world of ubiquitous computing devices, where many social interactions become technology-mediated, security and privacy decisions and practices

also become embedded into our everyday experience. Consequently, security and privacy research will have to shift the focus, similar to third-wave HCI, to everyday experience in the field as well.

One way to achieve this shift to field research is to emphasize research collaborations with industry partners. For example, messenger providers such as WhatsApp, which had around 2 billion monthly users in 2023, are perfectly situated to help research the effect of different user interaction mechanisms or user education.

Problem 3: Lack of validated psychometric measurement scales for different areas of security and privacy. One issue that I encountered when adapting established methods to security was the lack of psychometric scales specific to security and privacy. These are especially useful when evaluating the effect of different types of educational or technical interventions. Often, established scales, for example, the affinity for technology scale [88], are useful even when they are not directly related to security and privacy. However, for this thesis, a reliable scale to evaluate the secure experience would have made the evaluation and detection of security theater easier. Future work might investigate how an established UX evaluation scale, like the AttrakDiff 2 [132], might be adapted to include security and safety aspects. Systematic development of psychometric scales is still a rather new topic in the Usable Privacy and Security field. However, recent publications, like Hasan et al.'s [129] psychometric scale to measure how people value other peoples' privacy (VOPP), give an outlook on how the field could benefit from additional work in developing psychometric scales.

Change of methodological approaches opens up new avenues for research

One thing that I noticed during my work on this thesis was that the methods I used influenced the types of insights and, consequently, the following choice of further research topics. This can lead to a better understanding of already well-established research problems.

For example, before I used an autoethnography to study the use of authentication ceremonies in end-to-end-encrypted messengers (Chapter 7), it was already well-known that users did not know about this security feature; they did not understand its purpose, and that they had difficulties conducting these ceremonies correctly. Using an autoethnographic approach, I found that embedding these ceremonies into social interactions can create friction that makes the widespread adoption of this security feature unlikely in its current form. In our paper on the Tor over VPN practice (Chapter 5), we used a triangulation method – including a behavior measurement, user survey, and a media analysis – to understand a specific privacy practice in detail. From that work, we learned that social norms and perceived expectations can contribute to the spread of security and privacy practices – even when the exact effect or threat model is unclear.

Both of these examples highlight how using different methods can contribute new insights, in this case, the role of social aspects of using security and privacy technology. This aligns well with the emerging sub-field of social cybersecurity [299]. That applying different methods to study well-established research problems may lead to new insights

is also one of the core arguments of the philosopher of science Paul Feyerabend in his book *"Against Methods"* [81].

9.6. Social Cybersecurity Research and Design as a Way Forward

During the research work for this thesis, I became fascinated with the social aspects behind security and privacy practices. These became evident during my work on the Tor over VPN practice (Chapter 5) and the authentication ceremonies in end-to-end-encrypted messengers (Chapters 7 and 8).

Of course, the fact that the social context is important to understanding security and privacy practices is not a new insight in the Usable Privacy and Security field. Park et al. [211] investigated how account sharing, a practice that is not recommended, is used to create trust and closeness in romantic relationships. Wei et al. [292] discussed how social privacy norms affect the social acceptability of surveillance and control in different social contexts, i.e., between intimate partners and in parent-child relationships. Hassoun et al. [133] how information processing is an inherently social practice for Gen Zers, impacting how misinformation spreads. Das et al. [52] analyzed how social aspects influence the adoption of security features. Wu et al. [299] provided an overview of security and privacy research focused on social aspects. However, in general, the social aspects of security behaviors are still underinvestigated – likely because of the methodological difficulties described in the previous Section.

Interestingly, information security practices have been an established topic of research in sociology since at least the 1980s. When Adler and Adler studied secrecy amongst drug dealers [8], Merten et al. studied the enculturation of secrecy amongst junior-high school girls [195], and Lingel et al. studied the information secrecy practices in the punk rock subculture [168]. Ideally, we, as a research community, could transfer this knowledge and research practices into modern-day security, privacy, and safety research.

Up until now, the main approaches to researching and designing security and privacy mechanisms have been highly individualized, similar to how computer use is expected in the global north. However, humans are social beings who are embedded in inherently social and cultural environments. In the physical world, we often tackle safety risks as a community, with mutual support and social control. We should explore how to transfer these community-based social approaches to security, privacy, and safety to computer-mediated online interactions. Hence, future work could explore how to design effective cooperative security and privacy mechanisms and interventions that focus on providing mutual aid – similar to interactions in the offline world. The visibility of user actions to others is crucial for this to work. The social norms regarding mutual support can only come into play when others' behavior is communicated to members of the same social group. This change of design space also requires a shift in perspective. Instead of designing a user-machine interaction, it is necessary to focus on the possible user-to-user interaction and the social and cultural norms that come with them. This line of design research would bridge the current gap between individualized and community-based approaches based on mutual support that would improve everyone's well-being in our increasingly computer-mediated interactions.

10 Conclusion

This thesis started out with the recognition that users' secure experience – be it good or bad – affects their opinion on the software's security and, consequently, their choice of security practices. Avoiding a mismatch of secure experience with the provided protection from threats, a *security theater*, requires a conscious design of these secure experiences. To tackle the question how the design of secure experiences could avert security theater, I structured the thesis into three parts: (1) Recognizing security theater in case studies, (2) adapting methods from HCI to security, and (3) designing security.

However, the case studies in this thesis (Chapters 3, 4, and 5) revealed the different factors that can influence users' secure experience with software: threat models, social norms, mental models of technology, and user interaction. This complexity makes it challenging to design a secure experience since the designer only has control over one of these aspects. Even then, users might experience the same type of user interaction differently, making a fixed set of design guidelines to create secure experiences unhelpful.

Usually, the secure experiences this thesis uncovered in the case studies were unintentional. First and foremost, this underlines the potential of improving security perceptions by incorporating the secure experience concept into already existing design processes. Integrating this concept into the design process would also, in general, increase the awareness of the idea, making designers think about the *hedonic* qualities in addition to the *pragmatic* qualities of security-relevant user interactions. In Section 9.3 of the Discussion, this thesis proposes a method to integrate the factors of secure experience into a user-centered design process. In the process description, I explain how each chapter translated to the design process steps: Understanding and specifying the context of use, specifying the user requirements, producing design solutions, and evaluating the design.

In this thesis, I also adapted existing methods from HCI to security and tried new methodological approaches to broaden the scope of security research. In three case studies (Chapters 3, 4, and 5), I studied examples of security and privacy theater, hoping to understand the underlying factors that lead to this mismatch of experience with the protection abilities. Understanding these underlying factors can help the research community systematically detect security theater instances. Many of these underlying factors depend on the security mechanisms' context of use. I applied and adapted several methods from HCI to understand the context of use in detail, including systematic literature review (Chapter 3), media analysis (Chapter 5), analysis of app

reviews (Chapter 4), contextual inquiry at scale (Chapter 6), behavior measurements (Chapter 5), and autoethnography (Chapter 7). To further the research community's understanding of secure experience, future work could develop a psychometric scale that encapsulates the underlying factors this thesis has identified. This would be especially useful for systematically detecting security theater and evaluating the design in a user-centered design process. It might make sense to either extend a standard user experience evaluation scale, such as AttrakDiff 2 [132] or develop a new validates psychometric scale, such as Hasan et al.'s [129] psychometric scale of the value of other people's privacy (VOPP).

Further, future work might intensify research on the social aspects and the context of using security, privacy, and safety mechanisms. As Chapter 3 and 5 of this thesis suggest, security-focused companies have found ways to influence the secure experience. They do so not only with user interaction design but also with a marketing strategy [10] to influence threat models, mental models, and surrounding social norms. This showcases another reason why it is hard to design secure experiences: the context of use is malleable and may evolve over time, impacting the corresponding secure experience. Hence, social cybersecurity research and socially-aware design are increasingly necessary to understand the use context. The Usable Privacy and Security research field has already started to explore Social Cybersecurity [299]. Thereby aligning itself with a long tradition of ethnographic work that focused on the social aspects of information secrecy practices, e.g., amongst teenage girls [195], in the punk rock subculture [168], or in the drug world [8]. Focusing on the social aspects of cybersecurity allows the research community to highlight problems with existing societal power dynamics [177] – a common denominator behind threats to privacy and safety. As such, social cybersecurity research could be the bridge to shift the focus from security to the grand challenge of ensuring safety for everyone [265].

Part V.

Appendix

A The Curious Case of Tor over VPN

A.1. Survey Questionnaire

1. Have you accessed the Tor network previously (e.g., by using the Tor Browser)?
   - ◯ Yes
   - ◯ No
   - ◯ I don't know what Tor is
2. Do you currently have a VPN client (e.g., NordVPN, ExpressVPN, PIA, ...) installed on one of your devices?
   - ◯ Yes
   - ◯ No
   - ◯ I don't know what a VPN is

Skip to question 6 if participant has not answered yes to the first and second question

3. How would you prefer to access the Tor network?
   - ◯ I would prefer **to connect** to a VPN for accessing the Tor network.
   - ◯ I would prefer **not to connect** to a VPN for accessing the Tor network.
   - ◯ I have no preference.
4. Please explain the reasons for your preference on accessing the Tor network (which you chose in the previous question):
   _____
   _____
   _____
5. Can you run 2 miles in 2 minutes? *Note: This is one of Huang et al.'s [143] attention check questions. Participants that answered 'Disagree' or 'Strongly Disagree' were included in the final dataset.*

- ◯ Strongly Agree
- ◯ Agree
- ◯ Neutral
- ◯ Disagree
- ◯ Strongly Disagree

6. What describes your use of VPNs better (on devices that have a VPN client installed):
   - ◯ I **connect** to a VPN in specific situations, usually, I am **not connected** to a VPN.
   - ◯ I **disconnect** from VPN in specific situations, usually, I am **connected** to a VPN.

   If participant answered to connect to a VPN in specific situations only:

a) What are these specific situations?

_____
_____
_____

If participant answered to always be connected to a VPN:

b) Why are you always connected to a VPN?

_____
_____
_____

7. How old are you?

_____

8. What is your gender?
   - ◯ Woman
   - ◯ Man
   - ◯ Non-binary
   - ◯ I prefer not to say
   - ◯ I prefer to self-describe:

     _____

A.2. Codebooks for the Survey on Tor over VPN beliefs

Table A.1 and A.2 contain the codebooks for the qualitative analysis of the open-ended questions in the online survey about Tor over VPN beliefs.

A.3. Codebook for Tor over VPN information sources

- Communication
  - Ambiguous phrasing may suggest Tor/VPN combination
  - List of beneficial security advice implies combination is useful
  - Present Tor/DarkNet in negative light to make VPNs more attractive
  - Reads like a VPN advertisment
  - Conceptions of Technology
- Conceptions of Technology
  - Layers of security
  - Supports misconception that Tor is just Browser and VPN is necessary to protect traffic
  - Supports misconception that Tor is just a (special) VPN
- Control Belief
  - Onion over VPN feature easy to use
  - simple and effective
- Effects of Tor/VPN combination (Behavioral Belief)
  - Benefit of Tor/VPN depends on threat model
  - Combination of Tor and VPN reduces security
  - Combination of VPNs obfuscates traffic patterns
  - depends personal threat model
  - No harm in trying
  - Peculiar Tor/VPN combination does not add security / anonymity
  - Tor over VPN does not add security / anonymity
  - Tor over VPN hides Tor Usage
  - VPN does not hide Tor usage (well)
  - VPN over Tor protects from malicious exit nodes
  - VPN protects from Tor entry node
- Instructions (Control Belief)
  - Anonymous VPN necessary for Tor/VPN

  - Instruction to access the darknet with Tor/VPN
  - Onion over VPN feature is unnecessary
  - Recommending Tor/VPN combination w/out specific effects
  - Tor/VPN optional but beneficial
  - VPN is required to access Tor
- Normative Belief
  - "Tor by itself goes against best practice"
  - Appeal to authority
  - Assumed general practice
  - complex mix of VPNs
  - Credibility through officially supported VPN feature
  - Layering Proxies as a inside-joke Meme
  - Necessary practice to become part of the "security-focused" users
  - Tor/VPN combination employed by expert user
  - Trustworthy source of security advice
- Threat Model (Behavioral Belief)
  - Filter bubble affecting available information
  - Government Censorship
  - Government Repression
  - Harvard Bomb Threat
  - Oppressive Social Community
  - State agencies run Tor nodes
  - Tor is potentially unsafe because of US Government funding
  - Tor is untrustworthy
  - Tor vs. Onion-over-VPN depends on Trust Consideration
  - Trust in no-log policy crucial for Tor/VPN combination
  - Trust in some party (ISP,VPN, or Tor) is necessary
  - VPNs are untrustworthy

Table A.1.: Codebook for the open-ended question about participants' VPN usage preferences

Code	Description
No answer	Participant's response does not answer the question (makes no sense, obvious copy&paste, etc.)
General improvement of security and privacy	Using a VPN improves participant's privacy and security (without specifying)
Secure feeling	Participant expresses an improved feeling of safety and security
Untrusted network	Participant does not trust their ISP / WiFi / otherwise shared network
Hide user's location from service	Hiding location is only necessary in specific cases
Hide personally identifiable information (PII) from observers	Participant uses a VPN to protect their transmission of sensitive personal data
Hide the user's specific activity from observers	Participant wants to hide their activity from observers
Secure access to high-value target	Accessing high-value targets such as systems at home or cloud services
Protection from potentially insecure sites	Participant doesn't trust a specific service and wants to have additional protection in that case
Legal reasons	Using a VPN to protect from legal troubles
Filesharing	Using a VPN to share files
Circumvent geo-blocking	Accessing content that would otherwise not be available in the participant's region
Circumvent terms of service (ToS)	Avoid blocks from services for ToS violations (e.g., multiple account use)
Accessing Tor	Explicitily using a VPN to access the Tor network
Solving technical issues	Improving routing or checking if services are unavailable from other locations as well
Performance reasons	Participant only uses VPN when absolutely necessary because of limited speed or bandwith
Workplace requirements	The participant's working environment requires a VPN connection

Table A.2.: Codebook for the open-ended question about participants' reasons to use Tor over VPN

Code	Description
No answer	Participant's response does not answer the question (makes no sense, obvious copy&paste, etc.)
Distrust in Tor	Mistrust in the safety and security of the Tor network or browser
General security benefits	VPNs increase security, privacy, or anonymity in general (Expectation of linear additive protection)
Legal reasons	Legality of downloading and or using the Tor browser is unclear
No added benefits	No additional security / anonymity benefit of using a VPN to access Tor
Protection from ISP / guard node	Hide Tor activity from ISP or from guard node
Recommendations	Friends, experts, or other trusted sources recommend (Tor+)VPN use
Secure feeling	VPNs increase feeling of security
Too slow	(Tor+)VPN is too slow
Tracking protection	VPN protects from tracking
Always uses a VPN	Participant is always connected to a VPN and connects to Tor when necessary

B

Comparing Users' Perceptions of Anti-Stalkerware with the Technical Reality

B.1. Codebook for the Thematic Analysis

Table B.1 shows the initial codebook. Table B.2 shows the codebook we used to focus on the users' perception of the case-study apps' safety and security.

Table B.1.: Initial codebook that included users' general perceptions about the apps.

Codes	Description	AntiSpy	Lookout	Total
Effect +	Review reports an event that demonstrated the app's efficacy	119 110	41 39	160 149
Experience +	Review focuses on the app's great user experience	155 161	19 12	174 173
Performance +	Review highlights the technical performance of the app (e.g., quick scans or low battery drain)	40 37	12 14	52 51
Usability +	Review reports that the app is easy to understand and/or use	20 20	11 10	31 30
Payment +	Positive experience with payment for the app itself or the subscription	18 19	3 3	21 22
Response +	Positive experience with responsive app developers or support team	8 7	10 12	18 19
Privacy +	Reviewer praises the app for its privacy-preserving approach	5 4	4 5	9 9
Effect -	Review reports an event that demonstrated the app's inadequacy	28 27	27 23	55 50
Experience -	Review focuses on the app's bad user experience	1 2	1 0	2 2
Performance -	Review highlights the bad technical performance of the app (e.g., battery drain, slow scans, or bugs)	94 101	18 18	112 119
Usability -	Review reports that the app is hard to understand and/or use	47 42	15 18	62 60
Payment -	Negative experience with payment for the app itself or the subscription	48 44	20 21	68 65
Response -	Negative experience with unresponsive app developers or support team	22 21	2 2	24 23
Privacy -	Reviewer perceives the app as privacy-infringing	5 4	4 3	9 7

Table B.2.: The codebook for the second coding iteration that focused on the users' perception of the app's effectiveness, i.e., the *effect* code in the previous codebook.

Codes	Description	AS	LO	Sum
Real Life Safe	Experience report of an event where app protected reviewer from harm	52 50	27 26	79 76
Test passed	Reviewer tested the app's detection capabilities and was satisfied by the results	5 6	3 3	8 13
Secure Feeling	Experience of using the app gave reviewer a feeling of security	67 70	9 9	76 79
Notifications	Prompt notifications about security incidents gave reviewers a secure feeling	17 19	2 2	19 21
Real Life Fail	Experience report of an event where app failed to protect reviewer from harm	13 12	8 8	21 20
Test Fail	Reviewer tested the app's detection capabilities and was not satisfied by the results	13 15	8 9	21 24
Insecure Feeling	Experience of using the app did not reassure reviewer about its security	8 9	10 10	18 19
Likes Feature	Reviewer praise a specific feature of the app	10 10	3 3	13 13
Misses Feature	Reviewer complains about a feature they had before or would like to have	27 24	5 5	32 29
Update	Review concerned changes to the app by a software update	13 14	2 2	15 16
Time of Experience	Reviewers reference their long usage experience with the app to communicate their trust in the app's capabilities	19 19	3 4	22 23

Abbreviations: AS = AntiSpy, LO =Lookout

C

Understanding the Low Adoption of Authentication Ceremonies with Autoethnography

C.1. Guideline for Research Diary Entries

Based on external triggers:

- Is it a memory?
  - *Try to remember as much details as possible. Use smartphone images and chat conversation history for memory clues.*
- Is it a plan to authenticate?
  - *Is it an abstract or a very concrete plan to authenticate?*
  - *How did I get the idea for this plan?*
- Is it an ad-hoc authentication?
  - *How did I get the idea to authenticate in that moment?*
- Details about the authentication ceremony itself
  - *What was the social context*
  - *Describe conversation topics before authentication ceremony*
  - *How did I remember the authentication ceremony*
  - *How did I ask to authenticate*
  - *Were instructions necessary*
- Is it a memory of a missed authentication opportunity?
  - *Try to think about the possible reasons for missing the opportunity.*

Based on periodic reminders:

- With whom did I meet?
- Did I have an (abstract or concrete) plan to meet these people?
- Did I miss an ad-hoc opportunity to authenticate?
- Do I know if these people I met have a secure messenger?
- Do I chat with them using a secure messenger?

Reflection entries (every two weeks):

- Unstructured thoughts about the social aspects of different meetings, e.g., the personal relationships and the context of these situations.

C.2. Codebook

- Planing and Conducting Authentication Ceremonies
  - AC with in-person exchange of contact details
  - Ad-hoc authentication
  - Ad-hoc check of authentication status
  - Almost forgot plan to authenticate
  - Asking for AC can be awkward in large groups
  - Bad timing of keyreset notification
  - Check contact list for authentication status
  - Cognitive load of planning authentication
  - Concrete plan to authenticate
  - Confused about messenger's UI response
  - Did not want to seem intrusive
  - Explanation of messenger's UI necessary
  - Fear of missing authentication opportunities
  - First time authentication
  - Forgot planned authentication
  - Negotiation about purpose and effects of AC
  - Negotiation: introduce AC as a personal research topic
  - No negotiation necessary (for re-authentication)
  - No planning necessary for people who you meet often
  - Not in the same city (makes planning meetings difficult)
  - Notification about required authentication

  - Other person organized meeting
  - Pre-warned conversation partner about authentication
  - Re-Authentication
  - Recognize need for authentication
  - Recognize need for authentication shortly after meeting
  - Recognize need for AC while planning meeting with a messenger
  - Remembered after meeting during messenger use
  - Remembered because of Phone use during meeting
  - Remembered because of rare or difficult meeting
  - Talk about my research reminded me of AC
  - Tread carefully as not to disturb other kinds of social rituals
  - Unclear how ohter person's authentication status came to be
  - Unexpected change of authentication status
  - Unspecific plan to authenticate in the future
  - Used different messenger for planning and follow-up conversation
  - Wait for upcoming social ritual
- Emotional Experience
  - Angry - Frustrated - Annoyed
  - Anger about my own forgetfulness
  - Bad - Stressed - Overwhelmed
  - Demonstration of competence
  - Desire to explore
  - Disgusted - Disappointed
  - Disgusted - Disapproving - Embarrassing
  - Fearful - Anxious - Worried
  - Fearful - Insecure - Inadequate
  - Fearful - Rejected
  - Fearful - Threatened - Exposed
  - Happy - Accepted - Respected
  - Happy - Content - Free
  - Happy - Interested - Curious
  - Happy - Proud - Confident
  - Happy - Proud - Successful
  - Peaceful - Thoughtful - Pensive

  - Sad - Guilty - Ashamed
  - Sad - Vulnerable - Fragile
  - Satisfaction of success / Amending a mistake
  - Surprised - Amazed - Astonished
  - Surprised - Confused - Perplexed
  - Surprised - Excited - Eager
  - Surprised - Excited - Energetic
  - Surprised - Startled - Shocked
  - Unsure about current authentication status
  - Unsure about future possibilities to meet
- Sociocultural Aspects
  - ACs and Security becomes topic of conversation
  - AC awkward in specific situation
  - AC harmonizes well with established social practice of contact detail exchange
  - AC not (yet) embedded in other social ritual
  - AC without social ritual
  - Authentication in front of others may lead to replication
  - Cooperative planing of Security
  - Demonstration and Explanation of ACs
  - Do not want to burden others with a meeting or an AC
  - Educate others about Security
  - Habituating myself to treat any meeting as a potential AC
  - Meetings not only about authentication, requires socially acceptable framework
  - Meetings with far-away friends is a social ritual
  - Meetings with Friends that I meet often are less ritualistic
  - Postpone AC until contact is geographically closer
  - Smartphone is a visual reminder (with questionable social acceptance)
  - Social Anxiety / Fear of Judgement impacts demand for AC
  - Socially inappropriate to plan a meeting
  - Social aspects more important than threat model
  - Social ritual
  - Topic of conversation triggers AC
  - Unsure if meeting will take place
  - Videocall with contact because it is the usual way of meeting

D

Transferring Update Behavior from Smartphones to Smart Consumer Devices

D.1. Instructions on finding Update Settings

Android

Operating System Updates: (1) Check if developer options are activated - until version 8.x they are found at the bottom of the main settings menu. From version 9 they are found in the system settings menu; (2) If developer options are activated and the corresponding menu exists: check if "Automatic System updates" are activated (default) or not.
Application Updates: (1) Open your Google Play Store Application; (2) Tap the hamburger-menu in the upper-left corner to open the Play Store menu; (3) Scroll down to the settings option; (4) Tap on the option "automatic updates".

iOS

Operating System Update: (1) Open the iOS settings; (2) Scroll down to the option "General" and tap it; (3) In this menu the entry "Software update" should be in the second place; (4) Wait for the listing to load, the option for automatic updates should be at the bottom of display.
Application Updates: (1) Open the iOS settings; (2) Scroll down and choose the option "iTunes & App Store"; (3) Below the heading "Automatic Downloads" there is an option for applications; (4) A green button shows that automatic downloads are enabled, and a grey button shows that they are not.

D.2. Formative Field Study

D.2.1. Questionnaire

- **Update Settings:**
  (1) Which operating system and which version is currently installed on your phone? (2) What is your current setting for automatic operating system updates? (3) Why did you choose this setting? (4) What is your current setting for automatic application updates? (5) Why did you choose this setting?

- **Demographic data:**
  (6) What is your gender (female, male, diverse, I prefer not to answer)? (7) How old are you? (8) What is your major (for students) or what is your occupation (for non-students)? (9) Please tell us how well the following statements apply to you (1 = Not at all ...7 = Very much): (a) It is difficult for me to convince computers to do what I want. (b) Concerning computers, I don't think I am very competent. (c) I think I am a skilled computer user. (d) I can help others with their computer problems. (e) I find it difficult to learn new computer software.(f) I am able to learn a programming language. (10) Provide three things that you regularly do in order to keep your smartphone or your personal data secure.

D.2.2. Demographics, Codebooks, and Update Settings

Table D.1.: Participants' demographics in the formative field study

	#	m	sd
N	52		
Age		23.62	3.71
Self-Efficacy (all)		5.21	1.26
Self-Efficacy (w/out students)		4.88	1.27
Gender			
Women	13		
Men	37		
Preferred not to say	2		
Students	45		
Computer Science	13		
Business	8		
Teaching	7		
Law	6		
Psychology	4		
Other	7		
Non-students	7		

- **Reasons for OS update settings:**
  Do not remember (18); Maintain control over installed software (2); General desire

to be up to date (2); Installed OS does not provide automatic update option (2); Practicality (1); Compatibility problems (1); Data cap on their mobile contract (1); Security (1)

- **Reasons for application update settings:**
  Did not change it (15); Maintain control over installed software (8); Data cap on their mobile contract (7); Practicality (4); Annoyance (4); Do not remember (4); General desire to be up to date (2); Installed OS does not provide automatic update option (2); Not enough storage space (1); Security (1)

- **Security-relevant day-to-day behavior:**
  Authentication (23); Self-Denial of potentially useful products or features (18); Check data protection specific settings (16); Password management (10); Secure network access (10); Backup (7); Use common sense (7); Protection software (6); Encryption (4); Physical access control (4); Updates (4); Others (4)

Table D.2.: Distribution of OS and application updates for Android and Apple users.

	OS Updates		App Updates	
Apple	On:	**17**	On:	**15**
	Off:	**6**	Off:	**10**
Android	DO on/ UP on:	**3**	Always:	**1**
	DO off/ UP off:	**1**	WiFi only:	**21**
	DO off/ UP on:	**16**	Never:	**3**
Total		**43**		**50**

Annotations. Number of participants that chose the possible option. DO = Developer options. UP = Updates.

Table D.3.: Distribution of OS and application updates regarding self-efficacy.

	Self-efficacy ≥ 4				**Self-efficacy < 4**			
OS	OS updates		App updates		OS updates		App updates	
Apple	On:	**4**	On:	**14**	On:	**1**	On:	**1**
	Off:	**16**	Off:	**8**	Off:	**2**	Off:	**2**
Android	DO on/ UP on:	**3**	Always:	**1**	DO on/ UP on:	**0**	Always:	**0**
	DO off/ UP off:	**1**	WiFi only:	**13**	DO on/ UP off:	**0**	WiFi only:	**8**
	DO off/ UP on:	**11**	Never:	**3**	DO off/ UP on:	**5**	Never:	**0**

Annotations. Number of participants that chose the possible option. DO = Developer options. UP = Updates.

D.3. Online Survey

D.3.1. Questionnaire, Demographics, Reasons for (De)activation, and Update Avoidance Behavior

1. **Update settings on your smartphone:**
   (1) Which OS do you use on your smartphone? [Android, iOS, other]; (2) Which exact version of the chosen OS do you use?; (3) Take a screenshot of your OS update settings and upload it; (4) Have you changed those settings in the past?; (5) Why did you choose this setting? (6) Take a screenshot of your app update settings and upload it; (7) Have you changed those settings in the past?; (8) Why did you choose this setting?

2. **Personal expectations about updates:**
   For all update notifications shown in Section D.3.2: (1) You are just about to use (insert device here) and the following update notification pops up; (2) How important do you think is this update? [5-point Likert scale]; (3) State your reasons for the last answer; (4) What kind of changes would you expect from such an update?; (5) When would you update? [Now, Later, Never]; (6) State your reasons for the last answer; (7) How would you change the update notification?
   Afterwards: (1) How large do you think is the share of app updates that are relevant for security? (2) How large do you think is the share of OS updates that are relevant for security? (3) How should an update be presented so that you perceive it as security-relevant?

3. **Reasons for (de)activation of automatic updates:**
   (1) Other people gave the following reasons for their activation of automatic updates. Please state how much you agree with them [5-point Likert scale]: I want to keep up with the current version, It is convenient to have them done automatically, Installing updates is good for security, I am annoyed by notifications in case of manual update installation, other; (2) Other people gave the following reasons for their deactivation of automatic updates. Please state how much you agree with them [5-point Likert scale]: I want to control which software and which version is installed on my phone, I fear compatibility problems with other software, my phone contract includes a low amount of data, I am annoyed by automatic updates, My phone does not have enough free storage for updates, others.

4. **Update avoidance behavior:**
   (1) At what time of day do you charge your phone battery?; (2) At which location do you usually charge your phone battery?; (3) At which locations is your phone usually connected to a WiFi network?; (4) At which times of the day is your phone connected to the WiFi, so that automatic updates could be installed?

5. **Personality:**
   (1) Psychological Reactance Scale [141]; (2) Affinity to Technology scale; (3) Big Five Inventory scale [Agreeableness and Conscientiousness]

6. **Demographic data:**
   (1) Gender; (2) Age; (3) How would you rate your knowledge of German?; (4) Type of occupation; (5) Field of occupation; (6) Highest completed educational level; (7) Available household-income per month

7. **Comments:**
   (1) Did you experience technical problems during this questionnaire?; (2) Please describe your problems; (3) General comments

Table D.4.: Participants' demographics in the online survey

	#	m	sd
N	91		
Age		29.13	8.39
Affinity for Technology Interaction Scale		4.56	0.95
Reactance to Autononmy Scale		2.85	0.59
Big Five Inventory Scale			
Extraversion		2.90	0.36
Agreeableness		3.52[1]	0.57
Conscientiousness		3.51[1]	0.49
Neuroticism		3.15	0.39
Openness to Experience		2.69	0.55
Gender			
Women	16		
Men	73		
Preferred not to say	2		

[1] The above average scores for conscientiousness and agreeableness are noteworthy, since they correlate with increased security awareness [116].

D.3.2. Update Notifications

We showed participants five update notifications: (1) Figure D.1 shows a system update, (2) Figure D.2 shows several available application updates, (3) Figure D.3 shows an open dishwasher that displays a notification of an ongoing update, (4) Figure D.4 shoes an available update for self-lacing basketball shoes, and (5) Figure D.5 shows an available update in a car.

D.3.3. Codebook

- **Curiosity (11)**
- **Update Preparation** after relevance check (11), Update as soon as electricity and/or internet available (84), use own WiFi (6), Prevention of data loss [after Backup (15), Threat of data loss intimidating (9)]

Table D.5.: Ranked reasons for (de)activing automatic updates

Reasons for deactivation	m	sd
I want to control which software (version) will be installed	4.45	1.82
My phone contract has a limited data cap	4.09	2.11
I am concerned about potential compatibility problems	3.47	1.74
I am annoyed by automatic updates	3.46	2.01
My phone has insufficient storage space for updates	2.97	1.9
Reasons for activation		
security reasons	5.49	1.41
"stay up to date"	5.19	1.54
convenience	5.14	1.68
annoying update notification	4.62	1.79

Table D.6.: Participants who avoid charging their battery and connecting to WiFi at the same time might demonstrate update avoidance behavior

Time of day	Charge battery	WiFi[1]
morning	5	40
before noon	4	36
noon	0	27
afternoon	3	0
dinnertime	8	55
night	68	75
whenever necessary	4	not applicable

[1] Multiple choice response

- **Scheduling** Time for Update [Immediately (61), At next opportunity (116), Point in time of no importance (11), No time for Updates (22)], Update prevents use (150), not while out and about (25), Not leave pending/remove notification (42), no counter-argument apparent (9), No disruption because in background/finished quickly (40), App Updates do not disrupt use (15)

- **Scepticism** IoT incomprehension (99), Incomprehension (40), no demand/unimportant (95), New devices error-prone (3),

- **Principles of importance** Update size implies importance [Small changes are unimportant (7), Update important if finished quickly (2), Bigger Update → later, smaller → sooner (3), Important because of long installation duration (1), System Updates take longer (7)], Rare Updates important (16), Updates are important (113), System updates are important (79), Update only important for used apps (63), Apps differ in importance (2), important → sooner (20), Updates unimportant for IoT devices (133)

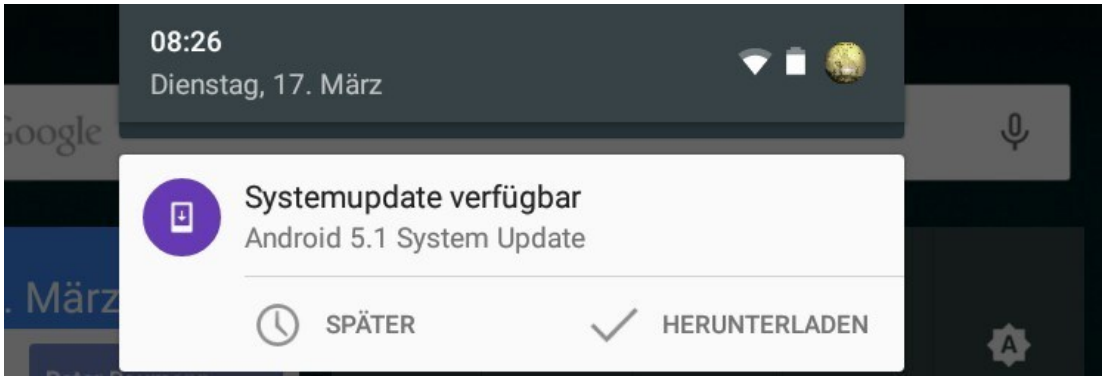


Figure D.1.: Android System

- **Expected changes** User Interface [UI Changes (75), Device in mint condition through update (3), Updates important for UX change (2), Improved Usability (27)], No noticeable changes (244), Maintenance (185), Bug-fixing (290), New features (252), Improvements (114), Performance (198), Changes anticipated by users (4), Safety (59), Privacy (8), Security (338), (only) devices attached to network need be up-to-date (3)
- **Negative Experiences** Never change a running System (51), Wait for field reports (11), Updates can cause errors (17), No update because of space lacking (2), Negative experience long duration (2), Negative experience as reason without explanation (7)
- **Update Deployment** Right to a say [Choice to delay makes update unimportant (4), Right to a say desired (49), Choice when to install update (25)], Information through notification [Notification no boost to confidence (10), Improved update notification (58), visual information within notification (32), More information within notification (349), Notification emphasizes importance (2), Notification as source of information (16), Less information within notification (49), Notify through phone (4)], Automatic updates [Automatic updates preferred (48), Unimportant updates should happen unsupervised (6), Critical updates automatically (22)], Timing of notification disruptive (14), Timing of notification convenient (1), Download vs. installation of update (4), No suggestion (529)

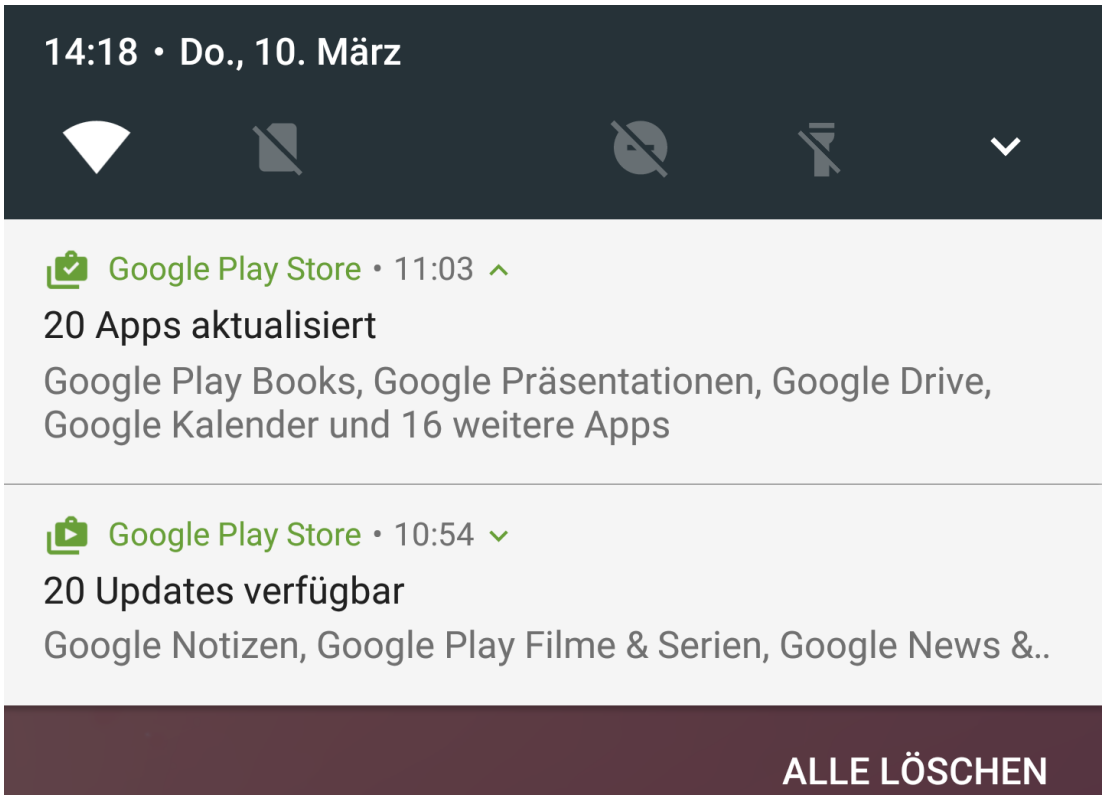


Figure D.2.: Android App

Figure D.3.: Dishwasher

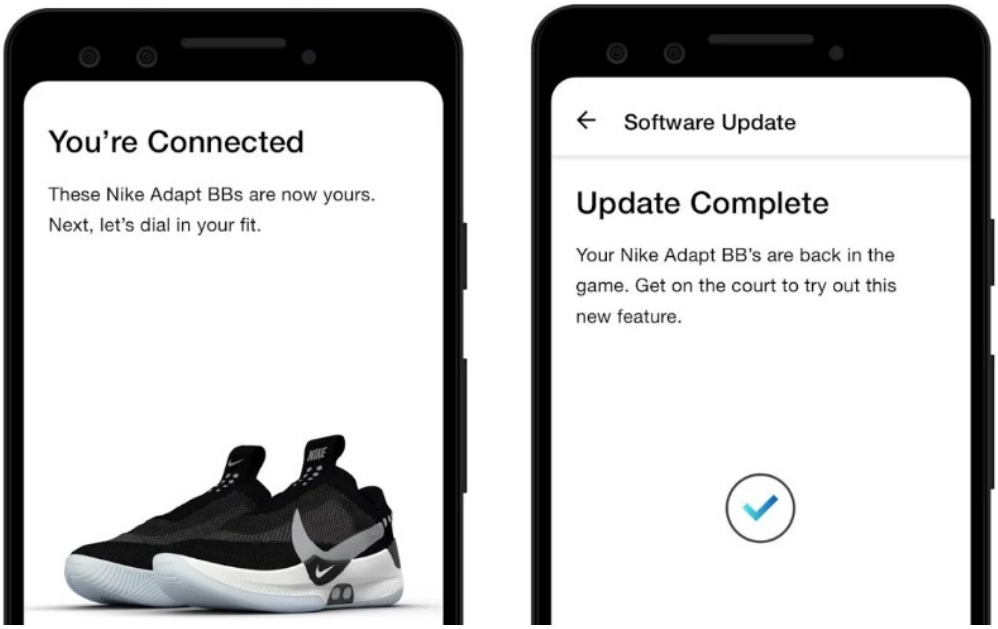


Figure D.4.: Basketball Shoes

Figure D.5.: Car System

E

Exploring User-Centered Security Design for Usable Authentication Ceremonies

E.1. Collaborative Design Workshop

E.1.1. Procedure

We split the workshops into three phases as common in Participatory Design studies according to Spinuzzi [262] and similar to the study by Weber et al. [291] about SSL warning messages. Of those three phases the design phase was the longest with a length of about 40 to 45 minutes and we tried to limit each workshop to about an hour. In the following sections we describe the scenarios and questions we used to elicit responses from the participants.

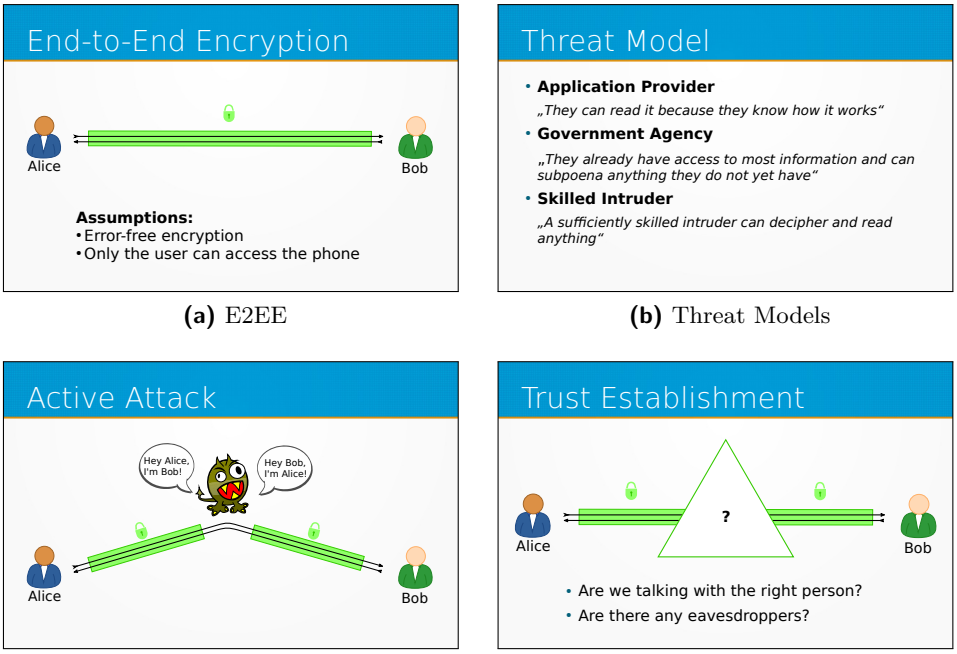
Discussion of Experiences.

(1) With which secure messengers have you had experience?; (2) What kind of positive or negative experiences have you had with secure messengers?; (3) Which of those messengers do you not use anymore, and why?

Creating a Shared Language.

We presented a short slide show (see Figure E.1) explaining a few basic concepts needed to discuss authentication in instant messaging. The following questions covered in that slide show: (1) What is end-to-end encryption and under which circumstances is it secure?; (2) What are common threat models of users and for which of them does end-to-end encryption help? *(as mentioned by Vaziripour et al. [284] and Renaud et al. [242])*; (3) What is a Monster-in-the-Middle attack?; (4) What are the unknown properties in unauthenticated conversations?

We encouraged the participants to ask question during and after the presentation to ensure that everyone understood the basics.



(a) E2EE **(b)** Threat Models

(c) MitM Attack **(d)** Trust Establishment

Figure E.1.: Slideshow presented during the collaborative design workshops

Designing Concept Ceremonies.

At the beginning of the design phase we presented the following scenario to the participants: *"You and a colleague from another branch office want to discuss a surprise party for a mutual friend. Your mutual friend with technical skills wants to find out exactly what you are conspiring but you definitely do not want to ruin the surprise and installed a secure messenger. You do not really know that colleague too well and you may or may not be able to meet him in person. What would you do in order to convince yourself that you are in fact talking to the right colleague and that nobody listens in?"*.

In the discussion that ensued afterwards we encouraged the participants to draw their ideas in order to explain them. Additionally, we frequently asked which conversation partner had to act and in which order they had to act according to their concept ideas.

In order talk about intuitive authentication approaches commonly used in the offline world, we presented a picture of an entrance hall of a bank and a picture of a delivered package in front of a door. With these artifacts in mind we asked the following questions: (1) You accepted a package of unknown value for an unknown neighbour and a few days later someone comes by to pick it up. What convinces you that you are handing the package over to the right person?; (2) You have been assigned a new bank advisor and you are waiting in the entrance hall to talk about a loan. What convinces you that you talking to your bank advisor and are not victim of a fraud?

Afterwards, we discussed if any of those approaches could be translated to electronic communication and if so, how that would work.

E.1.2. Codebook for Experience Reports

- Prevalence WhatsApp Prevalence; Tedious Diversity; Large User Base; Peer Pressure as a reason to use messengers; Client Availability on many platforms
- Privacy Personal Data unimportant; Concerns about Mass-Surveillance; Health Data considered important; Secret Chats available; Ephemeral Messaging; Images not deletable; Too many required App authorizations; Uploading Contact Information; Wiretapping Device; Minimal or no data storage as a sign of quality
- Trust Loss of Trust due to sponsoring; Facebook as a sign of insecurity; Open Source as a sign of quality; No trust in Telegram
- Aesthetics Pleasant Design; Puristic Design; Ugly
- Obstacles Message-Loss after re-install; Tedious Diversity; Complicated Group-Chats; Missing Notifications; Intermittent message-loss; Incomprehensible Authentication Ceremony; Phone-Number necessary; Forced to use app because of technical reasons; Technical Limitations of Device; No pictures next to names; Difficult to use
- Features Telegram Stickers; More convenient than SMS; Sticker selection and trading; Only short Videos sendable; Easy to use voice messages; Usable without phone-number; Messages available on different devices; File-Transfer impossible

E.1.3. Codebook for Prototyping Sessions

- Shared Knowledge Personal Questions; Exchange Secret Keys; Exchange Code-Words; Exchange symmetric encryption keys; Knowledge of personal information; QR-Code confirmation; Continuously confirm identity with known information
- Pictures Picture-Check; Recognition of person; Profile pictures not important; Recognition from photo
- Social Inconspicuous identity-check; Check reactions of person in question; Avoid seeming suspicious; Lookism; Check reactions of social environment; Trust Network; Trust Third Party; Suspicious Behaviour; Fits the environment; Friendliness
- Institution ID-card; Name-Tag; Authentication by possession; Verification Institution; Verification by Experts; Fingerprint; Signature; Place of Conversation; Crypto-ID-card
- Habituation Time and Reoccurring Messages; Visualize Habituation; Writing-Style as expected; Speech-Style
- Testing Compare Chat-history; Uncover Message Manipulation; Continuously confirm identity with known information; Manual Check by the users; Check if identity is plausible; Measure Transmission Latency
- Wishes and Requirements Base-Level of Trust; Manual Check by the users; Avoid seeming suspicious; Automatic detection of attack; Small choice when selecting secrets; Inconspicuous identity-check; Warning after security-errors; Careful after incidents; Minimal personal information required; Effort only worth it for few

- REACTIONS Change communication medium

E.2. Iterative Storyboard Prototyping

E.2.1. Field notes

Implicit feedback during walkthrough. During the prototype walkthrough with participants we noted hesitation, apparent confusion, and remarks by the users. We referred back to those notes during the explicit design feedback.

Comprehension questions. (1) How would you describe this process to a friend? [Free text answer]; (2) How do you think does this process affect the security of the conversation? [Free text answer]; (3) Why do you think the conversation will be less or more secure after this process? [Free text answer]; (4) How strong is your trust in this additional security (1 = no trust, 10 = strong trust)? [10-point scale]; (5) Explain why you chose this level of trust: [Free text answer]

Explicit design feedback and suggested changes. We asked participants what they would change about the prototype. We told them that they could change all descriptions, colors, dialogues, and the control flow. We gave them a pen and discussed every page of the paper-based prototype.

E.2.2. Adapted UI Flow for Combination Lock Prototype

During the iterative storyboard prototyping phase we construct a UI flow for the combination lock prototype. This prototype required locking users out from the conversation until enter the correct combination. This added an extra state to the UI flow that allows the ceremony to add an additional security guarantee. Figure E.2 shows the adapted UI flow. An authentication ceremony can fail in two ways: (1) either the SMP answer is dropped / infinitely delayed, or (2) the answer is invalid. In the first case a user that has not provided the correct combination *yet* is not able to send a message. Hence, a message arriving before the answer is an indication for an ongoing attack. In the second case users that failed to enter the correct code are not able to send a message. Hence, a message arriving after a failed authentication attempt indicates an ongoing attack.

E.3. Evaluation

E.3.1. Scenario

Figure E.3 shows the two parts of the introductory scenario that all participants experience before they continue to their selected authentication ceremony. It asks participants to imagine that they are messaging a friend about their boss, in that moment they think about possible negative consequences if the boss new about this.

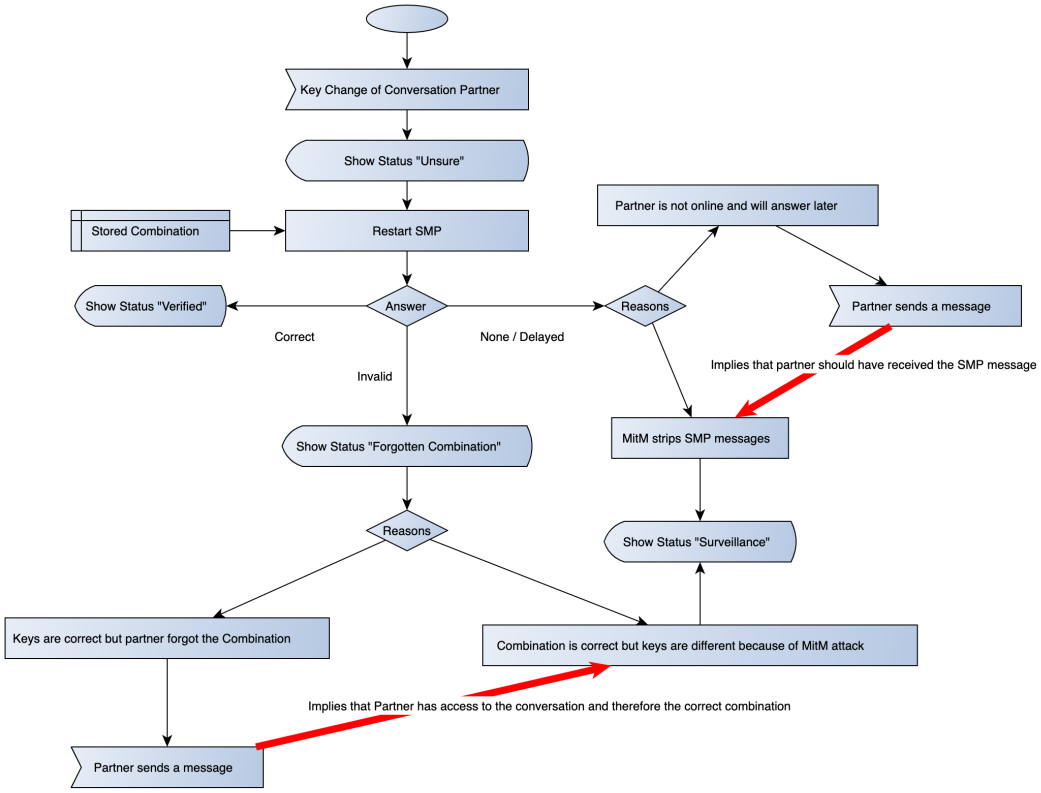


Figure E.2.: Adapted UI flow that makes it possible to differentiate between MitM attacks and regular use.

E.3.2. Prototypes

Figure E.4 and E.5 show the storyboards that participants had to click through when they were assigned to the **ID card based prototype**: Tap to scan card, then (a) Scan card of conversation partner, (b) Successful scan, (c) Shown own ID card, and (d) Conversation partner scans ID card.

Figure E.6 shows the storyboards that participants had to click through when they were assigned to the **Combination lock based prototype**: (a) Set a combination, (b) Conversation successfully locked, (c)Instructions to announce code, (d) Sharing code with conversation partner.

Figure E.7 shows the storyboards that participants had to click through when they created selfies for to the **Selfie based prototype**: (a)Instructions for three selfies, (b) Instruction to demonstrate action, (c) Action demonstrated, and (d) Continue to conversation partner's perspective.

Figure E.8 shows the storyboards that participants had to click through when they verified the selfies from the **Selfie based prototype**: (a) Instructions to verify the conversation partner's three selfies, (b) Verify required action on selfie (3x), and (c) Conversation secured.

E.3.3. Questionnaire

Perceived security. (1) How secure do you think is the conversation? [7-point Likert]; (2) Explain why you think so: [Free text answer]

Understanding of threats. (1) I think the conversation is secure from attackers that are able to steal messages while they are in transit [7-point Likert]; (2) I think the conversation is secure from attackers that have access to messages on the messenger's servers. [7-point Likert]; (3) I think the conversation is secure from attackers that change how the messenger's servers work. [7-point Likert]; (4) I think the conversation is secure from attackers that can intercept messages in transit and send fake messages. [7-point Likert]; (5) I think the conversation is secure from attackers that have access to my phone. [7-point Likert] (reversed)

Usability. (1) Systems Usability Scale (SUS); (2) User Experience Questionnaire (UEQ-S); (3) Describe one thing you liked about the presented authentication ceremony: [Free text answer]; (4) Describe one thing you did not like about the presented authentication ceremony: [Free text answer]

Messenger usage. (1) I use the following mobile messengers: [Multiple Choice: Signal, WhatsApp, Wire, Wickr, Telegram, Viber, Threema, Line, Tox, Briar, iMessage]; (2) Have you seen one or more of the following screens?[Multiple Choice: Screenshots of Signal, WhatsApp, Wickr, Wire, Threema]; (3) For which people in your contact list would you invest extra time and effort to increase your conversation's security: [Free text answer]; (4) For how many people in your contact list would you invest the additional effort? [List: 0,...,10,most,everyone]

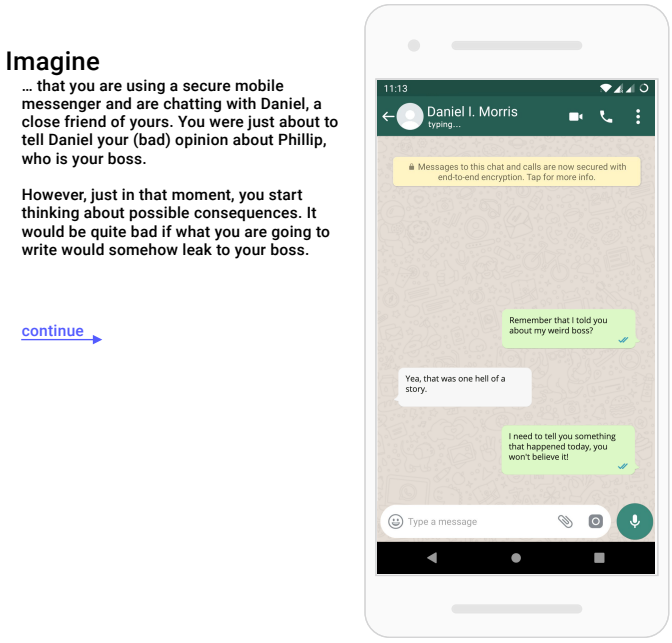
Demographic data. (1) What is your age?; (2) What is your gender? [woman, man, non-binary, prefer not to disclose, prefer to self-describe]; (3) What is your highest qualification?; (4) Affinity to Technology (ATI) scale

E.3.4. Demographics

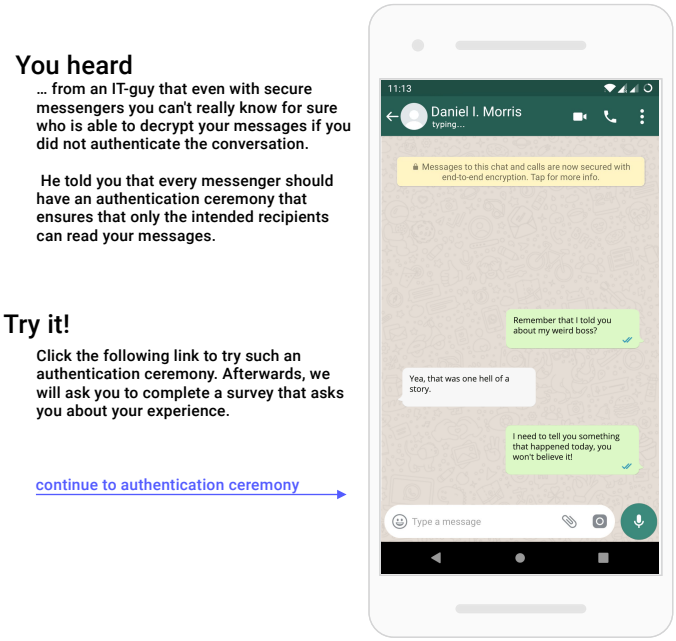
Table E.1 shows the demographic information of the evaluation participants.

Table E.1.: Demographic information from the evaluation

Demographic Variable	N	Percent	Mean	SD
Age	131		30.58	8.10
Gender				
Woman	47	35.77		
Man	83	63.36		
Non-Binary	1	0.76		
Education				
9th - 11th grade	2	1.53		
High school graduate or equivalent	10	7.63		
Some college, no degree	15	11.45		
Bachelor' degree	54	41.22		
Master's degree	31	23.66		
Doctoral degree	1	0.76		
Associate's degree: academic program	9	6.87		
Associate's degree: occupational, technical or vocational program	4	3.05		
Professional degree	4	3.05		
I would rather not say	1	0.76		
Affinity for Technology Interaction (ATI)			3.58	0.63
No. of used messengers			2.48	1.24
WhatsApp	115	87.79		
Telegram	55	41.98		
Viber	31	23.66		
Signal	8	6.11		
Threema	10	7.63		
Line	19	14.50		
iMessage	71	54.20		
Wire	10	7.63		
Tox	3	2.29		
Briar	2	1.53		
Wickr	1	0.76		
No. of seen authentication screes			1.20	0.49
Threema	26	19.85		
WhatsApp	90	68.70		
Wickr	29	22.14		
Wire	12	9.16		
No. contacts participants would authenticate			5.16	2.94
Most	13	9.92		
Everybody	4	3.05		



(a) Introduction to scenario



(b) Explanation of a potential solution

Figure E.3.: Scenario that participants experience before their selected condition.

Click through the user interface of this authentication ceremony.

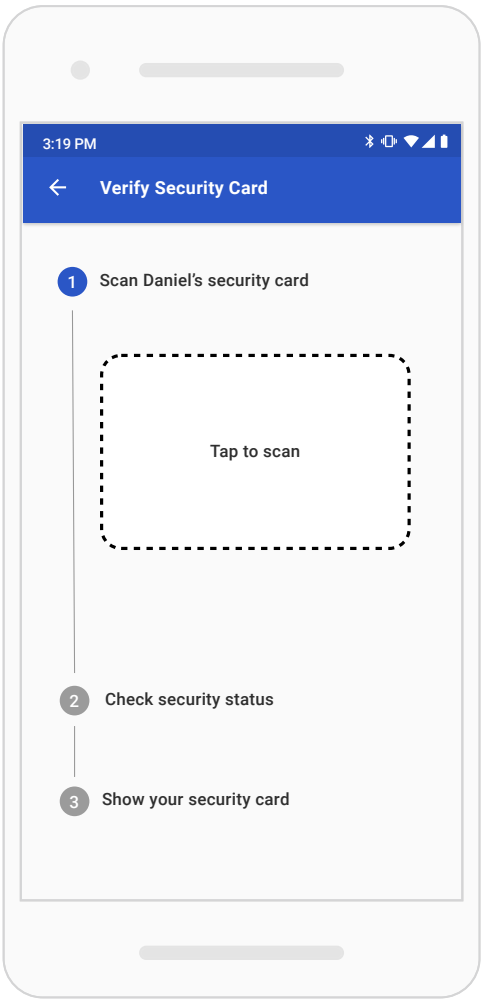
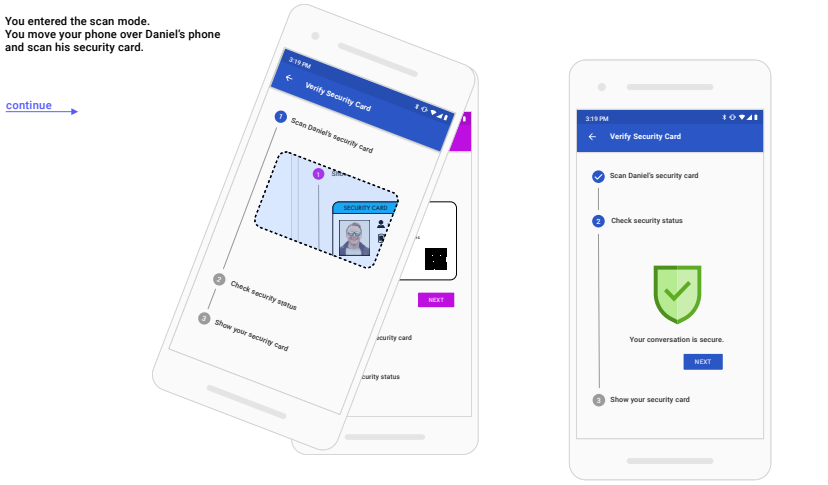
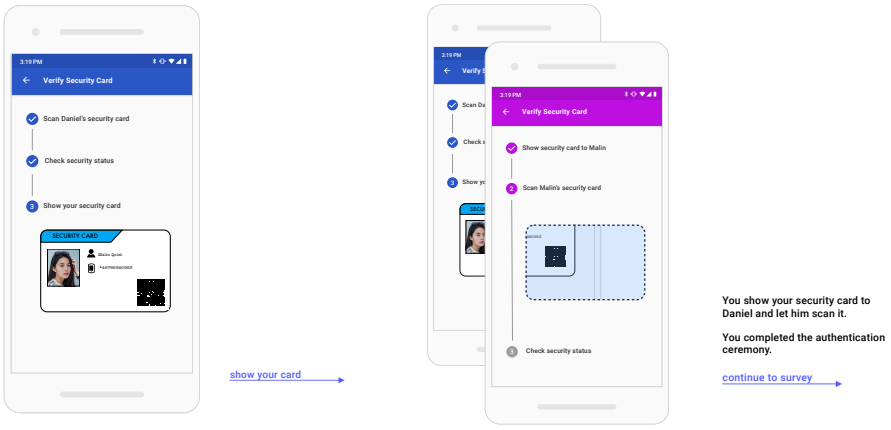


Figure E.4.: Start of the ID cards based authentication ceremony: Tap to scan card

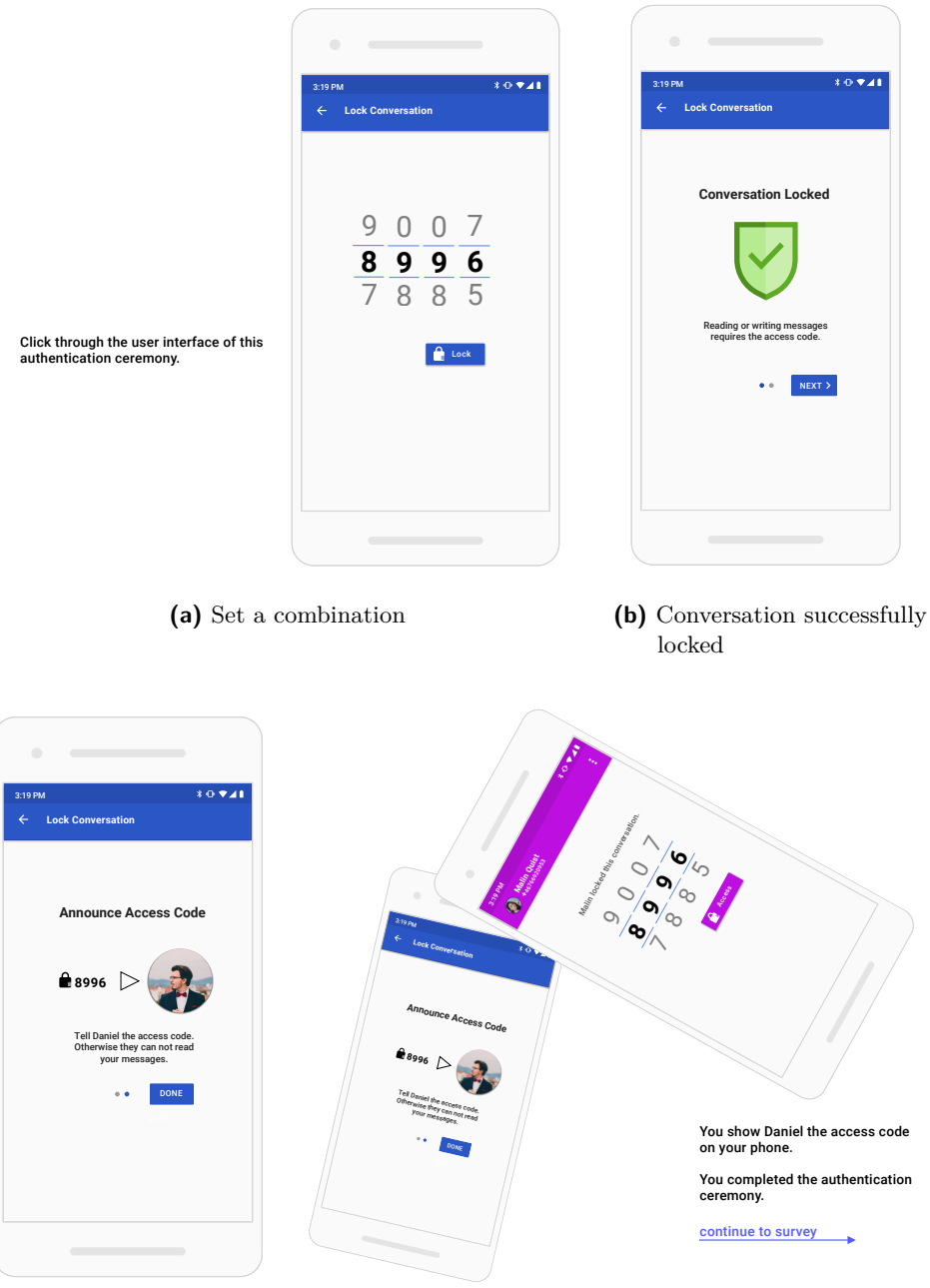


(a) Scan card of conversation partner **(b)** Successful scan



(c) Show own ID card **(d)** Conversation partner scans ID card

Figure E.5.: ID card based prototype



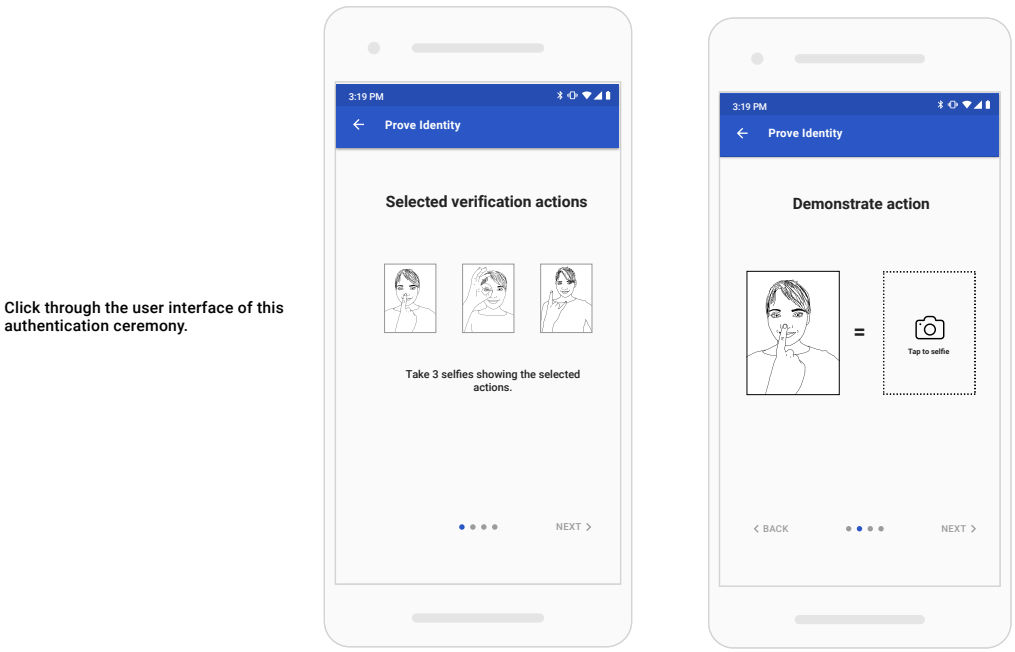
(a) Set a combination

(b) Conversation successfully locked

(c) Instructions to announce code

(d) Sharing code with conversation partner

Figure E.6.: Combination lock based prototype



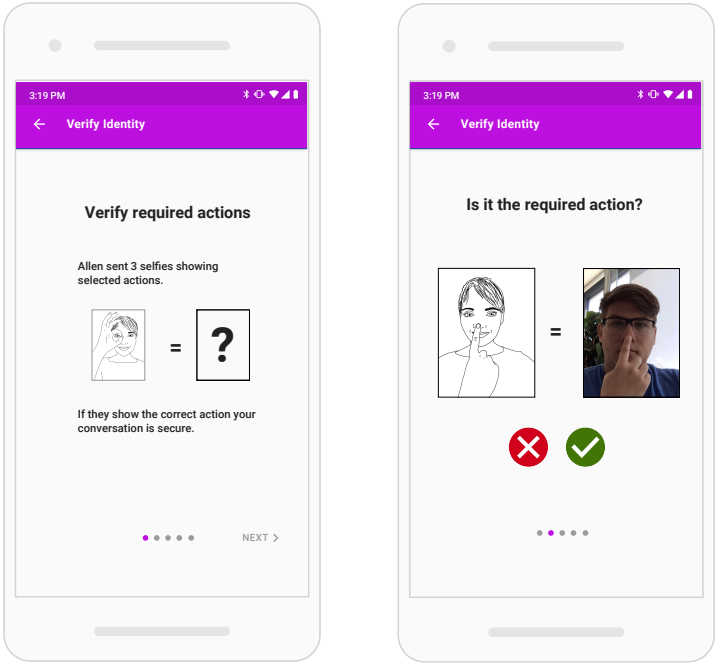
(a) Instructions for three selfies

(b) Instruction to demonstrate action (3 times)

(c) Action demonstrated (3x)

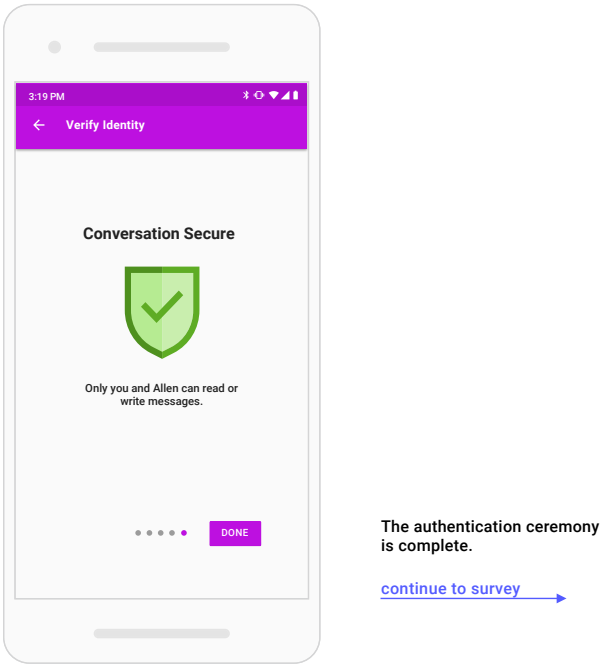
(d) Continue to conversation partner's perspective

Figure E.7.: Creating selfies for the selfie based prototype



(a) Instructions to verify the conversation partner's three selfies

(b) Verify required action on selfie (3 times)



(c) Conversation secured

Figure E.8.: Verifying selfies for the selfie based prototype

Bibliography

Author's Papers for this Thesis

[P1] **Fassl, M.**, Gröber, L. T., and Krombholz, K. Stop the Consent Theater. In: *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, May 2021, 1–7.

[P2] **Fassl, M.**, Anell, S., Houy, S., Lindorfer, M., and Krombholz, K. Comparing User Perceptions of Anti-Stalkerware Apps with the Technical Reality. In: *Proceedings of the Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*. USENIX Association, Boston, MA, USA, 2022, 135–154.

[P3] **Fassl, M.**, Ponticello, A., Dabrowski, A., and Krombholz, K. Investigating Security Folklore: A Case Study on the Tor over VPN Phenomenon. *Proc. ACM Hum.-Comput. Interact.* CSCW2 (Nov. 2023).

[P4] **Fassl, M.**, Neumayr, M., Schedler, O., and Krombholz, K. Transferring Update Behavior from Smartphones to Smart Consumer Devices. In: *Computer Security. ESORICS 2021 International Workshops*. Ed. by Katsikas, S., Lambrinoudakis, C., Cuppens, N., Mylopoulos, J., Kalloniatis, C., Meng, W., Furnell, S., Pallas, F., Pohle, J., Sasse, M. A., Abie, H., Ranise, S., Verderame, L., Cambiaso, E., Maestre Vidal, J., and Sotelo Monge, M. A. Vol. 13106. Springer International Publishing, Cham, 2022, 357–383.

[P5] **Fassl, M.** and Krombholz, K. Why I Can't Authenticate — Understanding the Low Adoption of Authentication Ceremonies with Autoethnography. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI '23. ACM, Hamburg, Germany, 2023, 15.

[P6] **Fassl, M.**, Gröber, L. T., and Krombholz, K. Exploring User-Centered Security Design for Usable Authentication Ceremonies. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI '21. ACM, Yokohama, Japan, May 2021, 1–15.

Other Papers of the Author

[S1] Distler, V., **Fassl, M.**, Habib, H., Krombholz, K., Lenzini, G., Lallemand, C., Cranor, L. F., and Koenig, V. A Systematic Literature Review of Empirical Methods and Risk Representation in Usable Privacy and Security Research. *ACM Transactions on Computer-Human Interaction* 28, 6 (Dec. 2021), 1–50.

[S2] Distler, V., **Fassl, M.**, Habib, H., Krombholz, K., Lenzini, G., Lallemand, C., Koenig, V., and Cranor, L. F. Empirical research methods in usable privacy and security. In: *Human Factors in Privacy Research*. Ed. by Gerber, N., Stöver, A., and Marky, K. Springer International Publishing, Cham, 2023, 29–53.

[S3] Gröber, L., **Fassl, M.**, Gupta, A., and Krombholz, K. Investigating Car Drivers' Information Demand after Safety and Security Critical Incidents. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI. ACM, Yokohama, Japan, May 2021, 1–17.

[S4] Ortloff, A.-M., **Fassl, M.**, Ponticello, A., Martius, F., Mertens, A., Krombholz, K., and Smith, M. Different Researchers, Different Results? Analyzing the Influence of Researcher Experience and Data Type During Qualitative Analysis of an Interview and Survey Study on Security Advice. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI '23. ACM, Hamburg, Germany, 2023, 21.

[S5] Ponticello, A., **Fassl, M.**, and Krombholz, K. Exploring Authentication for Security-Sensitive Tasks on Smart Home Voice Assistants. In: *Proceedings of the Seventeenth Symposium on Usable Privacy and Security*. SOUPS. USENIX Association, Virtual Conference, 2021, 18.

Other references

[1] @opera. *Twitter Message: "You Always Have a Choice with Us When It Comes to Blocking Cookie Dialogs. It's Also Available on Opera Touch for iOS Devices."* Nov. 2020. URL: `https://twitter.com/opera/status/1325870165033152513` (visited on 02/11/2021).

[2] Abu-Salma, R. and Livshits, B. Evaluating the End-User Experience of Private Browsing Mode. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, Apr. 2020, 1–12.

[3] Abu-Salma, R., Sasse, M. A., Bonneau, J., Danilova, A., Naiakshina, A., and Smith, M. Obstacles to the Adoption of Secure Communication Tools. In: *2017 IEEE Symposium on Security and Privacy (S&P)*. IEEE, San Jose, CA, USA, 2017, 137–153.

[4] Ackerman, M. S. The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility. *Human–Computer Interaction* 15, 2-3 (Sept. 2000), 179–203.

[5] Ackerman, M. S., Cranor, L. F., and Reagle, J. Privacy in e-commerce: examining user scenarios and privacy preferences. In: *Proceedings of the 1st ACM Conference on Electronic Commerce (EC '99)*. ACM, Denver, CO, USA, 1999, 1–8.

[6] Acquisti, A., Adjerid, I., Balebako, R., Brandimarte, L., Cranor, L. F., Komanduri, S., Leon, P. G., Sadeh, N., Schaub, F., Sleeper, M., Wang, Y., and Wilson, S. Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online. *ACM Computing Surveys* 50, 3 (Oct. 2017), 1–41.

[7] Adams, A. and Sasse, M. A. Users are not the enemy. *Communications of the ACM* 42, 12 (1999), 40–46.

[8] Adler, P. A. and Adler, P. The Irony of Secrecy in the Drug World. *Urban Life* 8, 4 (Jan. 1980), 447–465.

[9] Agrawal, P. and Trivedi, B. A Survey on Android Malware and their Detection Techniques. In: *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*. IEEE, Coimbatore, India, 2019, 1–6.

[10] Akgul, O., Roberts, R., Namara, M., Levin, D., and Mazurek, M. L. Investigating Influencer VPN Ads on YouTube. In: *2022 IEEE Symposium on Security and Privacy (SP)*. IEEE, San Francisco, CA, USA, 2022, 876–892.

[11] Alexander, C. and Goldberg, I. Improved user authentication in off-the-record messaging. In: *Proceedings of the 2007 ACM Workshop on Privacy in Electronic Society (WPES 2007)*. ACM, 2007, 41–47.

[12] Andreasen, M. S., Nielsen, H. V., Schrøder, S. O., and Stage, J. What happened to remote usability testing?: an empirical study of three methods. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, 2007, 1405–1414.

[13] AppBrain. *Anti Spy Mobile PRO: Changelog*. 2022. URL: `https://www.appbrain.com/app/anti-spy-mobile-pro/com.antispycell` (visited on 06/08/2022).

[14] Apple. *App Security Overview*. 2021. URL: `https://support.apple.com/guide/security/app-security-overview-sec35dd877d0/web` (visited on 06/08/2022).

[15] Apple. *App Store Review Guidelines*. 2021. URL: `https://developer.apple.com/app-store/review/guidelines/` (visited on 06/08/2022).

[16] Arp, D., Spreitzenbarth, M., Hübner, M., Gascon, H., and Rieck, K. Drebin: Effective and Explainable Detection of Android Malware in Your Pocket. In: *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, Internet Society, San Diego, CA, USA, 2014.

[17] Asad, M. Prefigurative Design as a Method for Research Justice. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 1–18.

[18] Bacci, A., Bartoli, A., Martinelli, F., Medvet, E., Mercaldo, F., and Visaggio, C. A. Impact of Code Obfuscation on Android Malware Detection based on Static and Dynamic Analysis. In: *Proceedings of the 4th International Conference on Information Systems Security and Privacy - ICISSP*. SciTePress, Funchal, Madeira, Portugal, 2018, 379–385.

[19] Bakiu, E. and Guzman, E. Which Feature is Unusable? Detecting Usability and User Experience Issues from User Reviews. In: *2017 IEEE 25th International Requirements Engineering Conference Workshops (REW)*. IEEE, Lisbon, Portugal, 2017, 182–187.

[20] Balfanz, D., Smetters, D., Stewart, P., and Wong, H. C. Talking To Strangers: Authentication in Ad-Hoc Wireless Networks. In: *Proceedings of Network and Distributed System Security Symposium 2002 (NDSS '02)*. Internet Society, 2002.

[21] Bangor, A., Kortum, P. T., and Miller, J. T. An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction* 24, 6 (2008), 574–594.

[22] Bannon, L. J. From Human Factors to Human Actors: The Role of Psychology and Human-Computer Interaction Studies in Systems Design. In: *Design at Work.: Cooperative Design of Computer Systems.* 1991.

[23] Beautement, A., Sasse, M. A., and Wonham, M. The compliance budget: managing security behaviour in organisations. In: *Proceedings of the 2008 New Security Paradigms Workshop*. NSPW '08. Association for Computing Machinery, New York, NY, USA, 2008, 47–58.

[24] Bella, G. and Viganò, L. Security is Beautiful. In: *Security Protocols XXIII*. Ed. by Christianson, B., Švenda, P., Matyáš, V., Malcolm, J., Stajano, F., and Anderson, J. Vol. 9379. Springer International Publishing, Cham, 2015, 247–250.

[25] Bellini, R., Tseng, E., McDonald, N., Greenstadt, R., McCoy, D., Ristenpart, T., and Dell, N. "So-Called Privacy Breeds Evil": Narrative Justifications for Intimate Partner Surveillance in Online Forums. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (Dec. 2020).

[26] Bellissimo, A., Burgess, J., and Fu, K. Secure Software Updates: Disappointments and New Challenges. In: *1st USENIX Workshop on Hot Topics in Security*. 2006, 37–43.

[27] Binkhorst, V., Fiebig, T., Krombholz, K., Pieters, W., and Labunets, K. Security at the End of the Tunnel: The Anatomy of VPN Mental Models Among Experts and Non-Experts in a Corporate Context. In: *31st USENIX Security Symposium (USENIX Security 22)*. USENIX Association, Boston, MA, USA, 2022, 3433–3450.

[28] Biryukov, A. and Pustogarov, I. Bitcoin over Tor isn't a Good Idea. In: *2015 IEEE Symposium on Security and Privacy*. SP. IEEE, San Jose, CA, May 2015, 122–134.

[29] Bødker, S. When second wave HCI meets third wave challenges. In: *Proceedings of the 4th Nordic Conference on Human-computer Interaction: Changing Roles*. ACM, Oslo Norway, Oct. 2006, 1–8.

[30] Bødker, S. and Grønbæk, K. Cooperative Prototyping: Users and Designers in Mutual Activity. *International Journal of Man-Machine Studies* 34 (1991), 453–478.

[31] Böhme, R. and Köpsell, S. Trained to accept?: a field experiment on consent dialogs. en. In: *Proceedings of the 28th International Conference on Human Factors in Computing Systems (CHI '10)*. CHI. ACM, Atlanta, GA, USA, 2010, 2403–2406.

[32] Boudot, F., Schoenmakers, B., and Traoré, J. A fair and efficient solution to the socialist millionaires' problem. *Discrete Applied Mathematics* 111, 1-2 (2001), 23–36.

[33] Boyd, M. J., Sullivan Jr., J. L., Chetty, M., and Ur, B. Understanding the Security and Privacy Advice Given to Black Lives Matter Protesters. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama, Japan, May 2021, 1–18.

[34] Braun, V. and Clarke, V. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101.

[35] Braun, V. and Clarke, V. *Successful Qualitative Research: A Practical Guide for Beginners*. SAGE, Los Angeles, 2013.

[36] Brunvand, J. H. *The Study of American Folklore: An Introduction*. Second. Norton, New York, 1978.

[37] Bruun, A., Gull, P., Hofmeister, L., and Stage, J. Let your users do the testing: a comparison of three remote asynchronous usability testing methods. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, 2009, 1619–1628.

[38] Chamberlain, A., Bødker, M., and Papangelis, K. Mapping Media and Meaning: Autoethnography as an Approach to Designing Personal Heritage Soundscapes. In: *Proceedings of the 12th International Audio Mostly Conference on Augmented and Participatory Sound and Music Experiences*. AM '17. ACM, London, UK, Aug. 2017, 1–4.

[39] Chanchary, F. and Chiasson, S. User Perceptions of Sharing, Advertising, and Tracking. In: *Proceedings of the Eleventh USENIX Conference on Usable Privacy and Security*. SOUPS. USENIX Association, Ottawa, Canada, 2015, 53–67.

[40] Chang, H. *Autoethnography as Method*. Developing Qualitative Inquiry v. 1. Left Coast Press, Walnut Creek, CA, USA, 2008.

[41] Chatterjee, R., Doerfler, P., Orgad, H., Havron, S., Palmer, J., Freed, D., Levy, K., Dell, N., McCoy, D., and Ristenpart, T. The Spyware Used in Intimate Partner Violence. In: *2018 IEEE Symposium on Security and Privacy (SP)*. IEEE, San Francisco, CA, USA, May 2018, 441–458.

[42] Chen, N., Lin, J., Hoi, S. C. H., Xiao, X., and Zhang, B. AR-Miner: Mining Informative Reviews for Developers from Mobile App Marketplace. In: *Proceedings of the 36th International Conference on Software Engineering*. ICSE 2014. ACM, Hyderabad, India, 2014, 767–778.

[43] Cheng, J., Wong, S. H., Yang, H., and Lu, S. SmartSiren: Virus Detection and Alert for Smartphones. In: *Proceedings of the 5th International Conference on Mobile Systems, Applications and Services*. MobiSys '07. ACM, San Juan, Puerto Rico, 2007, 258–271.

[44] Cialdini, R. B. *Influence: Science and Practice*. Pearson Education, 2009.

[45] Coalition Against Stalkerware. *The State of Stalkerware in 2019*. Apr. 2020. URL: `https://media.kasperskycontenthub.com/wp-content/uploads/sites/100/2020/03/18084439/Kaspersky%5C_The-State-of-Stalkerware-in-2019%5C_Updated.pdf` (visited on 06/08/2022).

[46] AV-Comparatives. *Android Test 2019 - 250 Apps*. Jan. 2019. URL: `https://www.av-comparatives.org/tests/android-test-2019-250-apps/` (visited on 06/08/2022).

[47] Cooper, G. and Bowers, J. Representing the user: Notes on the disciplinary rhetoric of human-computer interaction. In: *The Social and Interactional Dimensions of Human-Computer Interfaces*. Ed. by Thomas, P. J. Cambridge Series on Human-Computer Interaction. Cambridge University Press, 1995, 48–66.

[48] Costanza-Chock, S. *Design Justice: Community-Led Practices to Build the Worlds We Need*. Information Policy. MIT Press, 2020.

[49] Cranor, L. F. A Framework for Reasoning About the Human in the Loop. In: *UPSEC'08 Proceedings of the 1st Conference on Usability, Psychology, and Security*. 2008.

[50] Das, S., Kim, T. H.-J., Dabbish, L. A., and Hong, J. I. The Effect of Social Influence on Security Sensitivity. In: *10th Symposium On Usable Privacy and Security*. SOUPS. USENIX Association, 2014, 143–157.

[51] Das, S., Kramer, A. D., Dabbish, L. A., and Hong, J. I. Increasing Security Sensitivity With Social Proof: A Large-Scale Experimental Confirmation. In: *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. CCS. ACM, Scottsdale, AZ, USA, Nov. 2014, 739–749.

[52] Das, S., Kramer, A. D., Dabbish, L. A., and Hong, J. I. The Role of Social Influence in Security Feature Adoption. In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. CSCW '15. ACM, Vancouver, BC, Canada, Feb. 2015, 1416–1426.

[53] De Luca, A., Das, S., Ortlieb, M., Ion, I., and Laurie, B. Expert and Non-Expert Attitudes towards (Secure) Instant Messaging. In: *Proceedings of the Twelfth USENIX Conference on Usable Privacy and Security*. SOUPS '16. USENIX Association, Denver, CO, USA, 2016, 147–157.

[54] Dechand, S., Naiakshina, A., Danilova, A., and Smith, M. In Encryption We Don't Trust: The Effect of End-to-End Encryption to the Masses on User Perception. In: *2019 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE, Stockholm, Sweden, June 2019, 401–415.

[55] Dechand, S., Naiakshina, A., Danilova, A., and Smith, M. In Encryption We Don't Trust: The Effect of End-to-End Encryption to the Masses on User Perception. In: *2019 IEEE European Symposium on Security and Privacy (EuroS&P)*. 2019.

[56] Degeling, M., Utz, C., Lentzsch, C., Hosseini, H., Schaub, F., and Holz, T. We Value Your Privacy ... Now Take Some Cookies: Measuring the GDPR's Impact on Web Privacy. en. In: *Proceedings 2019 Network and Distributed System Security Symposium*. NDSS. Internet Society, San Diego, CA, USA., 2019, 1–15.

[57] Demjaha, A., Spring, J., Becker, I., Parkin, S., and Sasse, A. Metaphors considered harmful? An exploratory study of the effectiveness of functional metaphors for end-to-end encryption. In: *Proceedings of the Workshop on Usable Security*. USEC. Internet Society, San Diego, CA, USA, 2018.

[58] Dingledine, R. *[Tor-Talk] Tor plus VPN (Was Re: Hi All!)* Jan. 2012. URL: `https://lists.torproject.org/pipermail/tor-talk/2012-January/022917.html` (visited on 11/23/2022).

[59] Dingledine, R., Hopper, N., Kadianakis, G., and Mathewson, N. One Fast Guard for Life (or 9 months). In: *Proceedings of the 14th Privacy Enhancing Technologies*. PETS. Amsterdam, Netherlands, 2014.

[60] Dingledine, R. and Mathewson, N. *Tor Path Specification*. Dec. 2021. URL: `https://github.com/torproject/torspec/blob/master/path-spec.txt` (visited on 11/22/2022).

[61] Distler, V., Lallemand, C., and Koenig, V. Making Encryption Feel Secure: Investigating how Descriptions of Encryption Impact Perceived Security. In: *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, Genoa, Italy, Sept. 2020, 220–229.

[62] Distler, V., Lenzini, G., Lallemand, C., and Koenig, V. The Framework of Security-Enhancing Friction: How UX Can Help Users Behave More Securely. In: *New Security Paradigms Workshop 2020*. NSPW '20. ACM, Online, USA, Oct. 2020, 45–58.

[63] Distler, V., Zollinger, M.-L., Lallemand, C., Roenne, P. B., Ryan, P. Y. A., and Koenig, V. Security - Visible, Yet Unseen? In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2019)*. CHI '19. ACM, Glasgow, Scotland, UK, 2019, 1–13.

[64] Dodier-Lazaro, S., Abu-Salma, R., Becker, I., and Sasse, M. A. From Paternalistic to User-Centred Security: Putting Users First with Value-Sensitive Design. In: *Workshop on Values in Computing*. ACM, 2017.

[65] Ducklin, P. *The Google Play "Super Antivirus" That's Not so Super at All...* Jan. 2018. URL: https://nakedsecurity.sophos.com/2018/01/19/the-google-play-super-antivirus-thats-not-so-super-at-all-report/ (visited on 06/08/2022).

[66] Duebendorfer, T. and Frei, S. *Why Silent Updates Boost Security.* TIK 302. ETH Zürich, 2009.

[67] Dutkowska-Zuk, A., Hounsel, A., Morrill, A., Xiong, A., Chetty, M., and Feamster, N. How and Why People Use Virtual Private Networks. In: *Proceedings of the 31st USENIX Security Symposium.* USENIX Association, Boston, MA, USA, 2022, 3451–3465.

[68] Dykstra, J. and Spafford, E. H. The case for disappearing cyber security. *Communications of the ACM* 61, 7 (June 2018), 40–42.

[69] Edward Snowden [@Snowden]. *Use Tor. Use Signal. https://t.co/VLvBsbVHKs.* Sept. 2016. URL: https://twitter.com/Snowden/status/778592275144314884 (visited on 11/22/2022).

[70] Edwards, W. K., Poole, E. S., and Stoll, J. Security automation considered harmful? In: *Proceedings of the 2007 Workshop on New Security Paradigms - NSPW '07.* ACM, New Hampshire, USA, 2008, 33.

[71] egress. *Cybersecurity Hype: How to Manage Expectations vs Reality.* 2022. URL: https://www.egress.com/media/14bhfms5/egress_cybersecurity_hype_report.pdf (visited on 01/26/2024).

[72] Ellis, C. and Bochner, A. Autoethnography, Personal Narrative, Reflexivity: Researcher as Subject. In: *Handbook of Qualitative Research.* Ed. by Denzin, N. K. and Lincoln, Y. S. Second. Sage Publications, Thousand Oaks, CA, USA, 2000, 733–768.

[73] Emms, M., Arief, B., and van Moorsel, A. Electronic Footprints in the Sand: Technologies for Assisting Domestic Violence Survivors. In: *Privacy Technologies and Policy.* Ed. by Preneel, B. and Ikonomou, D. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014, 203–214.

[74] *Ergonomics of Human-System Interaction Part 210: Human-centred Design for Interactive Systems.* Tech. rep. EN ISO 9241-210:2019. European Standards, 2019.

[75] Ermoshina, K., Halpin, H., and Musiani, F. Can Johnny build a protocol? Coordinating developer and user intentions for privacy-enhanced secure messaging protocols. In: *The 2nd European Workshop on Usable Security (EuroUSEC).* IEEE, 2017.

[76] European Court of Justice. *Declaration of Consent by Means of a Pre-Ticked Checkbox.* Jan. 2019. URL: http://curia.europa.eu/juris/document/document.jsf?docid=218462&doclang=EN (visited on 02/11/2021).

[77] Fagan, M., Khan, M. M. H., and Buck, R. A study of users' experiences and beliefs about software update messages. *Computers in Human Behavior* 51 (Oct. 2015), 504–519.

[78] Feal, Á., Calciati, P., Vallina-Rodriguez, N., Troncoso, C., and Gorla, A. Angel or Devil? A Privacy Study of Mobile Parental Control Apps. In: *Proceedings on Privacy Enhancing Technologies*. Vol. 2020. Apr. 2020, 314–335.

[79] Fenske, E., Mani, A., Johnson, A., and Sherr, M. Distributed Measurement with Private Set-Union Cardinality. In: *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. CCS. ACM, Dallas, TX, USA, 2017, 2295–2312.

[80] Fernandes, E., Jung, J., and Prakash, A. Security Analysis of Emerging Smart Home Applications. In: *2016 IEEE Symposium on Security and Privacy (SP)*. May 2016, 636–654.

[81] Feyerabend, P. *Against Method*. 3rd ed. Verso, London ; New York, 1993.

[82] *Fiddler | Web Debugging Proxy and Troubleshooting Solutions*. URL: `https://www.telerik.com/fiddler` (visited on 06/08/2022).

[83] Fiesler, C. and Proferes, N. "Participant" Perceptions of Twitter Research Ethics. *Social Media + Society* 4, 1 (Jan. 2018).

[84] Filieri, R. What makes online reviews helpful? A diagnosticity-adoption framework to explain informational and normative influences in e-WOM. *Journal of Business Research* 68, 6 (2015), 1261–1270.

[85] Fishbein, M. and Ajzen, I. *Predicting and Changing Behavior: The Reasoned Action Approach*. Psychology Press, New York, 2010.

[86] Forget, A., Pearman, S., Thomas, J., Acquisti, A., Christin, N., Cranor, L. F., Egelman, S., Harbach, M., and Telang, R. Do or do not, there is no try: user engagement may not improve security outcomes. In: *Proceedings of the Twelfth Symposium on Usable Privacy and Security (SOUPS)*. USENIX Association, Denver, CO, USA, 2016.

[87] Forman, G. Counting Positives Accurately Despite Inaccurate Classification. In: *16th European Conference on Machine Learning*. Ed. by Gama, J., Camacho, R., Brazdil, P. B., Jorge, A. M., and Torgo, L. ECML. Springer Berlin Heidelberg, Porto, Portugal, 2005, 564–575.

[88] Franke, T., Attig, C., and Wessel, D. A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *International Journal of Human–Computer Interaction* 35, 6 (2019), 456–467.

[89] Freed, D., Havron, S., Tseng, E., Gallardo, A., Chatterjee, R., Ristenpart, T., and Dell, N. "Is My Phone Hacked?" Analyzing Clinical Computer Security Interventions with Survivors of Intimate Partner Violence. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 202:1–202:24.

[90] Freed, D., Palmer, J., Minchala, D., Levy, K., Ristenpart, T., and Dell, N. "A Stalker's Paradise": How Intimate Partner Abusers Exploit Technology. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. ACM, Montreal, QC, Canada, 2018, 1–13.

[91] *Frida • A World-Class Dynamic Instrumentation Framework*. URL: `https://frida.re` (visited on 06/08/2022).

[92] Friedman, B., Howe, D., and Felten, E. Informed consent in the Mozilla browser: implementing value-sensitive design. en. In: *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*. HICSS. IEEE Comput. Soc, Big Island, HI, USA, 2002, 10.

[93] Frik, A., Malkin, N., Harbach, M., Peer, E., and Egelman, S. A Promise Is A Promise: The Effect of Commitment Devices on Computer Security Intentions. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. ACM, Glasgow, Scotland, UK, 2019, 1–12.

[94] FTC. *FTC Bans SpyFone and CEO from Surveillance Business and Orders Company to Delete All Secretly Stolen Data*. 2021. URL: `https://www.ftc.gov/news-events/press-releases/2021/09/ftc-bans-spyfone-and-ceo-from-surveillance-business` (visited on 06/08/2022).

[95] Fu, B., Lin, J., Li, L., Faloutsos, C., Hong, J., and Sadeh, N. Why People Hate Your App: Making Sense of User Feedback in a Mobile App Store. In: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '13. ACM, Chicago, IL, USA, 2013, 1276–1284.

[96] Gahalaut, A. K. and Khandnor, P. Reverse engineering: an essence for software re-engineering and program analysis. *International Journal of Engineering Science and Technology* 2, 06 (2010), 2296–2303.

[97] Gallagher, K., Patil, S., and Memon, N. New Me: Understanding Expert and Non-Expert Perceptions and Usage of the Tor Anonymity Network. In: *Proceedings of the Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*. USENIX Association, Santa Clara, CA, USA, 2017, 385–398.

[98] Gao, X., Yang, Y., Fu, H., Lindqvist, J., and Wang, Y. Private Browsing: an Inquiry on Usability and Privacy Protection. In: *Proceedings of the 13th Workshop on Privacy in the Electronic Society*. ACM, Scottsdale, AZ, USA, 2014, 97–106.

[99] Garlach, S. and Suthers, D. D. 'I'm supposed to see that?' AdChoices Usability in the Mobile Environment. In: *Proceedings of the 51st Hawaii International Conference on System Sciences*. HICSS. IEEE, Hawaii, USA, 2018, 3779–3788.

[100] Gaw, S., Felten, E. W., and Fernandez-Kelly, P. Secrecy, Flagging, and Paranoia: Adoption Criteria in Encrypted Email. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, 2006, 591–600.

[101] Geeng, C., Harris, M., Redmiles, E., and Roesner, F. "Like Lesbians Walking the Perimeter": Experiences of U.S. LGBTQ+ Folks With Online Security, Safety, and Privacy Advice. In: *31st USENIX Security Symposium (USENIX Security 22)*. USENIX Association, Boston, MA, USA, 2022, 305–322.

[102] Gehrmann, C., Mitchell, C. J., and Nyberg, K. Manual authentication for wireless devices. *RSA Cryptobytes* 7, 1 (2004), 29–37.

[103] Ghosh, A. K., Badillo-Urquiola, K., Guha, S., LaViola Jr, J. J., and Wisniewski, P. J. Safety vs. Surveillance: What Children Have to Say about Mobile Apps for Parental Control. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. ACM, Montreal, QC, Canada, Apr. 2018, 1–14.

[104] Ghosh, A. K. and Wisniewski, P. Understanding User Reviews of Adolescent Mobile Safety Apps: A Thematic Analysis. In: *Proceedings of the 19th International Conference on Supporting Group Work*. GROUP '16. ACM, Sanibel Island, FL, USA, 2016, 417–420.

[105] Gilbert, S., Shilton, K., and Vitak, J. When research is the context: Cross-platform user expectations for social media data reuse. *Big Data & Society* 10, 1 (Jan. 2023).

[106] Giles, D. C. Parasocial Interaction: A Review of the Literature and a Model for Future Research. *Media Psychology* 4, 3 (Aug. 2002), 279–305.

[107] Glaser, B. G. and Strauss, A. L. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Vol. 1. Developmental Psychobiology. Aldine de Gruyter, New York, 1967.

[108] Goodin, D. *Police decrypt 258,000 messages after breaking pricey IronChat crypto app*. Nov. 2018. URL: `https://arstechnica.com/?p=1408441` (visited on 01/07/2021).

[109] Google. *Developer Program Policy: September 16, 2020 Announcement - Play Console Help*. Sept. 2020. URL: `https://support.google.com/googleplay/android-developer/answer/10065487` (visited on 06/08/2022).

[110] Google. *Android Developers: <manifest>*. URL: `https://developer.android.com/guide/topics/manifest/manifest-element.html%5C#package` (visited on 06/08/2022).

[111] Google. *Android Developers: Configure Your Build*. URL: `https://developer.android.com/studio/build` (visited on 06/08/2022).

[112] Google. *Android Developers: PackageManager*. URL: `https://developer.android.com/reference/android/content/pm/PackageManager` (visited on 06/08/2022).

[113] Google Play Store. *Anti Spy Mobile PRO*. 2021. URL: `https://play.google.com/store/apps/details?id=com.antispycell` (visited on 06/08/2022).

[114] Google Play Store. *Mobile Security - Lookout*. 2021. URL: `https://play.google.com/store/apps/details?id=com.lookout` (visited on 06/08/2022).

[115] Gorski, P. L., Acar, Y., Iacono, L. L., and Fahl, S. Listen to Developers! A Participatory Design Study on Security Warnings for Cryptographic APIs. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM, 2020, 1–13.

[116] Gratian, M., Bandi, S., Cukier, M., Dykstra, J., and Ginther, A. Correlating human traits and cyber security behavior intentions. *Computers & Security* 73 (Mar. 2018), 345–358.

[117] Green, M. and Smith, M. Developers are Not the Enemy!: The Need for Usable Security APIs. *IEEE Security & Privacy* 14, 5 (Sept. 2016), 40–46.

[118] Greenberg, S. and Buxton, B. Usability evaluation considered harmful (some of the time). In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. ACM, 2008, 111–120.

[119] Grossman, J. Feeling secure. *interactions* 13, 3 (2006), 50.

[120] Gu, X. and Kim, S. "What Parts of Your Apps are Loved by Users?" (T). In: *2015 30th IEEE/ACM International Conference on Automated Software Engineering (ASE)*. IEEE, Lincoln, NE, USA, 2015, 760–770.

[121] Gunkel, D. J. The Relational Turn: Third Wave HCI and Phenomenology. In: *New Directions in Third Wave Human-Computer Interaction: Volume 1 - Technologies*. Ed. by Filimowicz, M. and Tzankova, V. Springer International Publishing, Cham, 2018, 11–24.

[122] Guzman, E. and Maalej, W. How Do Users Like This Feature? A Fine Grained Sentiment Analysis of App Reviews. In: *2014 IEEE 22nd International Requirements Engineering Conference (RE)*. IEEE, Karlskrona, Sweden, 2014, 153–162.

[123] Ha, E. and Wagner, D. Do Android users write about electric sheep? Examining consumer reviews in Google Play. In: *2013 IEEE 10th Consumer Communications and Networking Conference (CCNC)*. IEEE, Las Vegas, NV, USA, 2013, 149–157.

[124] Habib, H., Colnago, J., Gopalakrishnan, V., Pearman, S., Thomas, J., Acquisti, A., Christin, N., and Cranor, L. F. Away From Prying Eyes: Analyzing Usage and Understanding of Private Browsing. In: *Proceedings of the Fourteenth Symposium on Usable Privacy and Security*. SOUPS. USENIX Association, Baltimore, MD, USA, 2018, 159–175.

[125] Hamilton, M. L., Smith, L., and Worthington, K. Fitting the Methodology with the Research: An exploration of narrative, self-study and auto-ethnography. *Studying Teacher Education* 4, 1 (May 2008), 17–28.

[126] Hammad, M., Garcia, J., and Malek, S. A Large-Scale Empirical Study on the Effects of Code Obfuscations on Android Apps and Anti-Malware Products. In: *Proceedings of the 40th International Conference on Software Engineering*. ICSE '18. ACM, Gothenburg, Sweden, 2018, 421–431.

[127] Harborth, D., Pape, S., and Rannenberg, K. Explaining the Technology Use Behavior of Privacy-Enhancing Technologies: The Case of Tor and JonDonym. *Proceedings on Privacy Enhancing Technologies* 2020, 2 (Apr. 2020), 111–128.

[128] Harkin, D. and Molnár, A. Operating-System Design and Its Implications for Victims of Family Violence: The Comparative Threat of Smart Phone Spyware for Android Versus iPhone Users. *Violence Against Women* 27, 6-7 (May 2021), 851–875.

[129] Hasan, R., Weil, R., Siegel, R., and Krombholz, K. A Psychometric Scale to Measure Individuals' Value of Other People's Privacy (VOPP). In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg, Germany, Apr. 2023, 1–14.

[130] Hassenzahl, M. The Effect of Perceived Hedonic Quality on Product Appealingness. *International Journal of Human-Computer Interaction* 13, 4 (Dec. 2001), 481–499.

[131] Hassenzahl, M. Experience Design: Technology for All the Right Reasons. *Synthesis Lectures on Human-Centered Informatics* 3, 1 (Jan. 2010), 1–95.

[132] Hassenzahl, M., Burmester, M., and Koller, F. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: *Mensch & Computer 2003*. Ed. by Szwillus, G. and Ziegler, J. Vol. 57. Vieweg+Teubner Verlag, Wiesbaden, 2003, 187–196.

[133] Hassoun, A., Beacock, I., Consolvo, S., Goldberg, B., Kelley, P. G., and Russell, D. M. Practicing information sensibility: how gen z engages with online information. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI '23. ACM, Hamburg, Germany, 2023.

[134] Havron, S., Freed, D., Chatterjee, R., McCoy, D., Dell, N., and Ristenpart, T. Clinical Computer Security for Victims of Intimate Partner Violence. In: *Proceedings of the 28th USENIX Conference on Security Symposium*. SEC '19. USENIX Association, Santa Clara, CA, USA, 2019, 105–122.

[135] Hedegaard, S. and Simonsen, J. G. Extracting Usability and User Experience Information from Online User Reviews. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '13. ACM, Paris, France, 2013, 2089–2098.

[136] Hern, A. *'Fake' Android Antivirus App Developer Says Virus Shield Was a 'Foolish Mistake'*. Apr. 2014. URL: `http://www.theguardian.com/technology/2014/apr/10/fake-android-antivirus-app-developer-virus-shield` (visited on 06/08/2022).

[137] Herzberg, A. and Leibowitz, H. Can Johnny finally encrypt?: evaluating E2E-encryption in popular IM applications. In: *Proceedings of the 6th Workshop on Socio-Technical Aspects in Security and Trust*. STAST '16. ACM, Los Angeles, CA, USA, Dec. 2016, 17–28.

[138] Herzberg, A. and Leibowitz, H. Can Johnny finally encrypt?: evaluating E2E-encryption in popular IM applications. In: *Proceedings of the 6th Workshop on Socio-Technical Aspects in Security and Trust (STAST '16)*. ACM, 2016, 17–28.

[139] Herzberg, A., Leibowitz, H., Seamons, K., Vaziripour, E., Wu, J., and Zappala, D. Secure Messaging Authentication Ceremonies Are Broken. *IEEE Security & Privacy* 19, 2 (Mar. 2021), 29–37.

[140] Holmquist, L. E., Mattern, F., Schiele, B., Alahuhta, P., Beigl, M., and Gellersen, H.-W. Smart-Its Friends: A Technique for Users to Easily Establish Connections between Smart Artefacts. In: *Proceedings of the 3rd international conference on Ubiquitous Computing (UbiComp '01)*. Springer-Verlag, 2001, 116–122.

[141] Hong, S.-M. and Faedda, S. Refinement of the Hong Psychological Reactance Scale. *Educational and Psychological Measurement* (1996).

[142] Höök, K. Transferring qualities from horseback riding to design. In: *Proceedings of the 6th Nordic Conference on Human-Computer Interaction Extending Boundaries*. NordiCHI '10. ACM, Reykjavik, Iceland, 2010, 226–235.

[143] Huang, J. L., Bowling, N. A., Liu, M., and Li, Y. Detecting Insufficient Effort Responding with an Infrequency Scale: Evaluating Validity and Participant Reactions. *Journal of Business and Psychology* 30, 2 (2015), 299–311.

[144] Iacob, C. and Harrison, R. Retrieving and analyzing mobile apps feature requests from online reviews. In: *2013 10th Working Conference on Mining Software Repositories (MSR)*. IEEE, San Francisco, CA, USA, 2013, 41–44.

[145] Ikram, M., Vallina-Rodriguez, N., Seneviratne, S., Kaafar, M. A., and Paxson, V. An Analysis of the Privacy and Security Risks of Android VPN Permission-enabled Apps. In: *Proceedings of the 2016 Internet Measurement Conference*. ACM, Santa Monica, CA, USA, Nov. 2016, 349–364.

[146] Ion, I., Reeder, R., and Consolvo, S. "... No one Can Hack My Mind": Comparing Expert and Non-Expert Security Practices. In: *Proceedings of the Eleventh Symposium on Usable Privacy and Security*. 2015, 327–346.

[147] *Jadx*. URL: `https://github.com/skylot/jadx` (visited on 06/08/2022).

[148] Jain, D., Desjardins, A., Findlater, L., and Froehlich, J. E. Autoethnography of a Hard of Hearing Traveler. In: *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*. ASSETS '19. ACM, Pittsburgh, PA, USA, Oct. 2019, 236–248.

[149] Jansen, R. and Johnson, A. Safely Measuring Tor. In: *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. CCS '16. ACM, Vienna, Austria, 2016, 1553–1567.

[150] Kang, R., Brown, S., Dabbish, L., and Kiesler, S. Privacy Attitudes of Mechanical Turk Workers and the U.S. Public. In: *Proceedings of the Tenth Symposium On Usable Privacy and Security*. SOUPS. USENIX Association, Menlo Park, CA, USA, 2014, 37–49.

[151] Kang, R., Dabbish, L., Fruchter, N., and Kiesler, S. "My data just goes everywhere": User mental models of the internet and implications for privacy and security. In: *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. SOUPS. USENIX Association, Ottawa, ON, Canada, 2015, 39–52.

[152] Kelley, P. G., Cesca, L., Bresee, J., and Cranor, L. F. Standardizing privacy notices: an online study of the nutrition label approach. en. In: *Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI '10*. CHI. ACM, Atlanta, Georgia, USA, 2010, 1573–1582.

[153] Keyes, O., Hoy, J., and Drouhard, M. Human-Computer Insurrection: Notes on an Anarchist HCI. en. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow, Scotland, UK, May 2019, 1–13.

[154] Khalid, H. On identifying user complaints of iOS apps. In: *2013 35th International Conference on Software Engineering (ICSE)*. IEEE, San Francisco, CA, USA, 2013, 1474–1476.

[155] Khan, M. T., DeBlasio, J., Voelker, G. M., Snoeren, A. C., Kanich, C., and Vallina-Rodriguez, N. An Empirical Analysis of the Commercial VPN Ecosystem. In: *Proceedings of the Internet Measurement Conference 2018*. IMC '18. ACM, Boston, MA, USA, 2018, 443–456.

[156] Kiko. *Firefox Browser Add-On – "I Don't Care about Cookies"*. Jan. 2021. URL: `https://addons.mozilla.org/de/firefox/addon/i-dont-care-about-cookies/` (visited on 02/11/2021).

[157] Kobsa, A., Sonawalla, R., Tsudik, G., Uzun, E., and Wang, Y. Serial Hook-Ups: A Comparative Usability Study of Secure Device Pairing Methods. In: *Proceedings of the 5th Symposium on Usable Privacy and Security (SOUPS '09)*. USENIX Association, 2009, 12.

[158] Krombholz, K., Mayer, W., Schmiedecker, M., and Weippl, E. "I Have No Idea What I'm Doing" – On the Usability of Deploying HTTPS. In: *26th USENIX Security Symposium (USENIX Security 17)*. USENIX Security. 2017, 1–18.

[159] Kulyk, O., Hilt, A., Gerber, N., and Volkamer, M. "This Website Uses Cookies": Users' Perceptions and Reactions to the Cookie Disclaimer. In: *The 3rd European Workshop on Usable Security (EuroUSEC)*. Internet Society, London, UK, 2018, 1–11.

[160] Kulyk, O., Mayer, P., Volkamer, M., and Kafer, O. A Concept and Evaluation of Usable and Fine-Grained Privacy-Friendly Cookie Settings Interface. In: *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering*. TrustCom/BigDataSE. IEEE, New York, NY, 2018, 1058–1063.

[161] Law, E. L.-C., Vermeeren, A. P., Hassenzahl, M., and Blythe, M. Towards a UX Manifesto. In: *Proceedings of HCI 2007 The 21st British HCI Group Annual Conference University of Lancaster, UK*. Sept. 2007.

[162] Lee, Y. and Kozar, K. A. An empirical investigation of anti-spyware software adoption: A multitheoretical perspective. *Information & Management* 45, 2 (2008), 109–119.

[163] Leitão, R. Technology-Facilitated Intimate Partner Abuse: a qualitative analysis of data from online domestic abuse forums. *Human–Computer Interaction* 36, 3 (2021), 203–242.

[164] Levy, K. and Schneier, B. Privacy threats in intimate relationships. *Journal of Cybersecurity* 6, 1 (May 2020).

[165] Li, F., Rogers, L., Mathur, A., Malkin, N., and Chetty, M. Keepers of the Machines: Examining How System Administrators Manage Software Updates. In: *Proceedings of the Fifteenth Symposium on Usable Privacy and Security*. 2019.

[166] Lindorfer, M., Neugschwandtner, M., and Platzer, C. MARVIN: Efficient and Comprehensive Mobile App Classification through Static and Dynamic Analysis. In: *2015 IEEE 39th Annual Computer Software and Applications Conference*. COMPSAC. IEEE, Taichung, Taiwan, 2015, 422–433.

[167] Lindorfer, M., Neugschwandtner, M., Weichselbaum, L., Fratantonio, Y., Veen, V. van der, and Platzer, C. ANDRUBIS - 1,000,000 Apps Later: A View on Current Android Malware Behaviors. In: *2014 Third International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS)*. IEEE, Wrocław, Poland, 2014, 3–17.

[168] Lingel, J., Trammell, A., Sanchez, J., and Naaman, M. Practices of information and secrecy in a punk rock subculture. In: *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*. ACM, Seattle, WA, USA, 2012, 157–166.

[169] Lockton, D., Zea-Wolfson, T., Chou, J., Song, Y., Ryan, E., and Walsh, C. Sleep Ecologies: Tools for Snoozy Autoethnography. In: *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. DIS '20. ACM, Eindhoven, Netherlands, July 2020, 1579–1591.

[170] Lucero, A. Living Without a Mobile Phone: An Autoethnography. In: *Proceedings of the 2018 Designing Interactive Systems Conference*. DIS '18. ACM, Hong Kong, China, June 2018, 765–776.

[171] Maalej, W. and Nabil, H. Bug report, feature request, or simply praise? On automatically classifying app reviews. In: *2015 IEEE 23rd International Requirements Engineering Conference (RE)*. IEEE, Ottawa, ON, Canada, Aug. 2015, 116–125.

[172] Machuletz, D. and Böhme, R. Multiple Purposes, Multiple Problems: A User Study of Consent Dialogs after GDPR. In: *Proceedings on Privacy Enhancing Technologies*. Vol. 2. PETS. Sciendo, Berlin, DE, 2020, 481–498.

[173] Mai, A., Pfeffer, K., Gusenbauer, M., Weippl, E., and Krombholz, K. User Mental Models of Cryptocurrency Systems - A Grounded Theory Approach. In: *Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020)*. USENIX Association, 2020, 341–358.

[174] Mani, A., Wilson-Brown, T., Jansen, R., Johnson, A., and Sherr, M. Understanding Tor Usage with Privacy-Preserving Measurement. In: *Proceedings of the Internet Measurement Conference 2018*. IMC. ACM, Boston, MA, USA, Oct. 2018, 175–187.

[175] Manovich, L. *The Language of New Media*. The MIT Press, Cambridge, MA, USA, 2002.

[176] Marques, D., Muslukhov, I., Guerreiro, T., Beznosov, K., and Carriço, L. Snooping on Mobile Phones: Prevalence and Trends. In: *Proceedings of the Twelfth USENIX Conference on Usable Privacy and Security*. SOUPS '16. USENIX Association, Denver, CO, USA, 2016, 159–174.

[177] Marwick, A. Privacy Without Power: What Privacy Research Can Learn from Surveillance Studies. *Surveillance & Society* 20, 4 (Dec. 2022), 397–405.

[178] Mather, M. and Johnson, M. K. Choice-supportive source monitoring: Do our decisions seem better to us as we age? *Psychology and Aging* 15, 4 (2000), 596–606.

[179] Mathiasen, N. R. and Bødker, S. Threats or Threads: From Usable Security to Secure Experience? In: *Proceedings of the 5th Nordic Conference on Human-Computer Interaction: Building Bridges*. NordiCHI '08. ACM, Lund, Sweden, 2008, 283–289.

[180] Mathiasen, N. R. and Bødker, S. Experiencing Security in Interaction Design. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '11. ACM, Vancouver, BC, Canada, 2011, 2325–2334.

[181] Mathur, A. and Chetty, M. Impact of user characteristics on attitudes towards automatic mobile application updates. In: *Proceedings of the Thirteenth Symposium on Usable Privacy and Security*. 2017, 175–193.

[182] Mathur, A., Engel, J., Sobti, S., Chang, V., and Chetty, M. "They Keep Coming Back Like Zombies": Improving Software Updating Interfaces. In: *Proceedings of the Twelfth Symposium on Usable Privacy and Security*. 2016, 43–58.

[183] Mathur, A., Malkin, N., Harbach, M., Peer, E., and Egelman, S. Quantifying Users' Beliefs about Software Updates. In: *Workshop on Usable Security (USEC) 2018*. 2018.

[184] Matte, C., Bielova, N., and Santos, C. Do Cookie Banners Respect my Choice? In: *2020 IEEE Symposium on Security and Privacy*. SP. IEEE, San Francisco, CA, USA, 2020, 791–809.

[185] Matthews, T., O'Leary, K., Turner, A., Sleeper, M., Woelfer, J. P., Shelton, M., Manthorne, C., Churchill, E. F., and Consolvo, S. Stories from Survivors: Privacy & Security Practices when Coping with Intimate Partner Abuse. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI '17. ACM, Denver, CO, USA, 2017, 2189–2201.

[186] McCarthy, J. and Wright, P. Technology as experience. *interactions* 11, 5 (2004), 42–43.

[187] McCune, J. M., Perrig, A., and Reiter, M. K. Seeing-Is-Believing: Using Camera Phones for Human-Verifiable Authentication. In: *2005 IEEE Symposium on Security and Privacy (S&P '05)*. IEEE, 2005.

[188] McDonald, A. M. and Cranor, L. F. Beliefs and Behaviors: Internet Users' Understanding of Behavioral Advertising. In: *38th Research Conference on Communication, Information, and Internet Policy*. TPRC. 2010, 1–31.

[189] McDonald, N., Schoenebeck, S., and Forte, A. Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 1–23.

[190] McIlroy, S., Ali, N., Khalid, H., and E. Hassan, A. Analyzing and automatically labelling the types of user issues that are raised in mobile app reviews. *Empirical Software Engineering* 21, 3 (June 2016), 1067–1106.

[191] Mehrnezhad, M. A Cross-Platform Evaluation of Privacy Notices and Tracking Practices. en. In: *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. EuroS&P Workshops. IEEE, Genoa, Italy, Sept. 2020, 97–106.

[192] Melara, M. S., Blankstein, A., Bonneau, J., Felten, E. W., and Freedman, M. J. CONIKS: Bringing Key Transparency to End Users. In: *24th USENIX Security Symposium (USENIX Security 15)*. USENIX Association, 2015, 383–398.

[193] Melicher, W., Sharif, M., Tan, J., Bauer, L., Christodorescu, M., and Leon, P. G. (Do Not) Track Me Sometimes: Users' Contextual Preferences for Web Tracking. In: *Proceedings on Privacy Enhancing Technologies*. Vol. 2. PETS. Sciendo, Berlin, Germany, 2016, 135–154.

[194] Meme, K. Y. *Good Luck, I'm Behind 7 Proxies*. 2010. URL: `https://knowyourmeme.com/memes/good-luck-im-behind-7-proxies` (visited on 11/22/2022).

[195] Merten, D. E. ENCULTURATION INTO SECRECY AMONG JUNIOR HIGH SCHOOL GIRLS. *Journal of Contemporary Ethnography* 28, 2 (Apr. 1999), 107–137.

[196] Meulen, R. van der. *Gartner Says 6.4 Billion Connected "Things" Will Be in Use in 2016, Up 30 Percent From 2015*. URL: `https://www.gartner.com/en/newsroom/press-releases/2015-11-10-gartner-says-6-billion-connected-things-will-be-in-use-in-2016-up-30-percent-from-2015` (visited on 01/26/2024).

[197] Millett, L. I., Friedman, B., and Felten, E. Cookies and Web browser design: toward realizing informed consent online. en. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '01)*. CHI. ACM, Seattle, WA, USA, 2001, 46–52.

[198] Möller, A., Diewald, S., Roalter, L., Michahelles, F., and Kranz, M. Update Behavior in App Markets and Security Implications: A Case Study in Google Play. In: *Research in the LARGE: Proceedings of the 3rd International Workshop. Held in Conjunction with Mobile HCI*. 2012, 3–6.

[199] Moody, R. *VPN Market Report 2022: Who's Got the Biggest VPN Market Share?* Feb. 2020. URL: `https://www.comparitech.com/blog/vpn-privacy/vpn-market-share-report/` (visited on 11/22/2022).

[200] Muslukhov, I., Boshmaf, Y., Kuo, C., Lester, J., and Beznosov, K. Know Your Enemy: The Risk of Unauthorized Access in Smartphones by Insiders. In: *Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services*. MobileHCI '13. ACM, Munich, Germany, 2013, 271–280.

[201] Namara, M., Wilkinson, D., Caine, K., and Knijnenburg, B. P. Emotional and Practical Considerations Towards the Adoption and Abandonment of VPNs as a Privacy-Enhancing Technology. *Proceedings on Privacy Enhancing Technologies*. PETS 2020, 1 (2020), 83–102.

[202] Newman, N., Fletcher, R., Robertson, C. T., Eddy, K., and Nielsen, R. K. *Reuters Institute Digital News Report 2022*. 2022. URL: `https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2022-06/Digital_News-Report_2022.pdf` (visited on 11/22/2022).

[203] Nguyen, D. C., Derr, E., Backes, M., and Bugiel, S. Short Text, Large Effect: Measuring the Impact of User Reviews on Android App Security & Privacy. In: *2019 IEEE Symposium on Security and Privacy (SP)*. IEEE, San Francisco, CA, USA, 2019, 555–569.

[204] Nouwens, M., Liccardi, I., Veale, M., Karger, D., and Kagal, L. Dark Patterns after the GDPR: Scraping Consent Pop-ups and Demonstrating their Influence. en. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI. ACM, Honolulu, HI, USA, Apr. 2020, 1–13.

[205] nusenu. *OrNetStats*. 2021. URL: `https://nusenu.github.io/OrNetStats/` (visited on 07/05/2021).

[206] O'Kane, A. A., Rogers, Y., and Blandford, A. E. Gaining empathy for non-routine mobile device use through autoethnography. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '14. ACM, Toronto, ON, Canada, Apr. 2014, 987–990.

[207] Oudshoorn, N., Rommes, E., and Stienstra, M. Configuring the User as Everybody: Gender and Design Cultures in Information and Communication Technologies. *Science, Technology & Human Values* 29, 1 (2004), 30–63.

[208] OWASP. *Mobile Security Testing Guide (MSTG)*. 2021. URL: `https://mobile-security.gitbook.io/` (visited on 06/08/2022).

[209] Palmer, D. *Can You Trust Your Android Antivirus Software? Malicious Fake Protection Apps Flood Google Play Store*. June 2017. URL: `https://www.zdnet.com/article/can-you-trust-your-mobile-antivirus-software-malicious-fake-protection-apps-flood-google-play-store/` (visited on 06/08/2022).

[210] Panichella, S., Di Sorbo, A., Guzman, E., Visaggio, C. A., Canfora, G., and Gall, H. C. How can i improve my app? Classifying user reviews for software maintenance and evolution. In: *2015 IEEE International Conference on Software Maintenance and Evolution (ICSME)*. IEEE, Bremen, Germany, Sept. 2015, 281–290.

[211] Park, C. Y., Faklaris, C., Zhao, S., Sciuto, A., Dabbish, L., and Hong, J. Share and Share Alike? An Exploration of Secure Behaviors in Romantic Relationships. In: *Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018)*. 2018, 83–102.

[212] Park, S. and Albert, K. *A Researcher's Guide to Some Legal Risks of Security Research*. 2020. URL: https://clinic.cyber.harvard.edu/files/2020/10/Security%5C_Researchers%5C_Guide-2.pdf (visited on 06/08/2022).

[213] Parliament, E. and Council, E. *Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 Concerning the Processing of Personal Data and the Protection of Privacy in the Electronic Communications Sector (Directive on Privacy and Electronic Communications)*. Dec. 2002. URL: https://eur-lex.europa.eu/le%20agal-content/EN/ALL/?uri=CELEX%5C%3A32002L0058 (visited on 02/11/2021).

[214] Parsons, C., Molnar, A., Dalek, J., Kenyon, M., Haselton, B., Khoo, C., and Deibert, R. *The Predator in Your Pocket: A Multidisciplinary Assessment of the Stalkerware Application Industry*. 2019. URL: https://citizenlab.ca/docs/stalkerware-holistic.pdf (visited on 06/08/2022).

[215] Patil, S., Hoyle, R., Schlegel, R., Kapadia, A., and Lee, A. J. Interrupt Now or Inform Later? Comparing Immediate and Delayed Privacy Feedback. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI '15. ACM, Seoul, Republic of Korea, Apr. 2015, 1415–1418.

[216] PenzeyMoog, E. *Design for Safety*. A Book Apart, Aug. 2021.

[217] Permisio. *Permisio – Customise Your Privacy Preferences in Just a Few Clicks*. 2021. URL: https://www.permisio.com/ (visited on 02/11/2021).

[218] Pfeffer, K., Mai, A., Weippl, E., Rader, E., and Krombholz, K. Replication: Stories as Informal Lessons about Security. In: *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*. USENIX Association, Boston, MA, 2022, 1–18.

[219] Phong, M. V., Nguyen, T. T., Pham, H. V., and Nguyen, T. T. Mining User Opinions in Mobile App Reviews: A Keyword-Based Approach (T). In: *2015 30th IEEE/ACM International Conference on Automated Software Engineering (ASE)*. IEEE, Lincoln, NE, USA, 2015, 749–759.

[220] Pina, Á., Lobato, E. M., Santander, J. G., Roa, P., Sancristóbal, F., Morales, E., Barrocal (Writers), D., and Quintas (Director), J. *Money Heist - A Matter of Efficiency (Season 2, Part 3)*. 2017. URL: https://www.imdb.com/title/tt6851508/ (visited on 01/26/2024).

[221] Portokalidis, G., Homburg, P., Anagnostakis, K., and Bos, H. Paranoid Android: Versatile Protection for Smartphones. In: *Proceedings of the 26th Annual Computer Security Applications Conference*. ACSAC '10. ACM, Austin, TX, USA, 2010.

[222] Preda, M. D. and Maggi, F. Testing android malware detectors against code obfuscation: A systematization of knowledge and unified methodology. *Journal of Computer Virology and Hacking Techniques* 13, 3 (Aug. 2017), 209–232.

[223] Press Statement of the Dutch Police. *Police have achieved a breakthrough in the interception and decryption of crypto communication.* 2018. URL: `https://www.politie.nl/en/news/2018/november/02-apeldoorn-police-have-achieved-a-breakthrough-in-the-interception-and-decryption-of-crypto-communication.html` (visited on 01/07/2021).

[224] Pu, Y. and Grossklags, J. Valuating Friends' Privacy: Does Anonymity of Sharing Personal Data Matter? In: *Thirteenth Symposium on Usable Privacy and Security*. SOUPS 2017. USENIX Association, Santa Clara, CA, USA, 2017, 339–355.

[225] Rader, E. and Slaker, J. The importance of visibility for folk theories of sensor data. In: *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*. USENIX Association, Santa Clara, CA, 2017, 257–270.

[226] Rader, E. and Wash, R. Identifying patterns in informal sources of security information. *Journal of Cybersecurity* 1, 1 (Sept. 2015), 121–144.

[227] Rader, E., Wash, R., and Brooks, B. Stories as Informal Lessons about Security. In: *Proceedings of the Eighth Symposium on Usable Privacy and Security*. SOUPS '12. ACM, Washington D.C., USA, 2012.

[228] Rahman, M., Hernandez, N., Recabarren, R., Ahmed, S. I., and Carbunar, B. The Art and Craft of Fraudulent App Promotion in Google Play. In: *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*. CCS '19. ACM, London, UK, 2019, 2437–2454.

[229] Raja, F., Hawkey, K., Jaferian, P., Beznosov, K., and Booth, K. S. It's Too Complicated, so i Turned It off! Expectations, Perceptions, and Misconceptions of Personal Firewalls. In: *Proceedings of the 3rd ACM Workshop on Assurable and Usable Security Configuration*. SafeConfig '10. ACM, Chicago, IL, USA, 2010, 53–62.

[230] Ramesh, R., Evdokimov, L., Xue, D., and Ensafi, R. VPNalyzer: Systematic Investigation of the VPN Ecosystem. In: *Proceedings 2022 Network and Distributed System Security Symposium*. Internet Society, San Diego, CA, USA, 2022.

[231] Ramesh, R., Vyas, A., and Ensafi, R. "All of them claim to be the best": Multi-perspective study of VPN users and VPN providers. In: *32nd USENIX Security Symposium (USENIX Security 23)*. USENIX Association, Anaheim, CA, 2023.

[232] Rammstedt, B. and John, O. P. Kurzversion des Big Five Inventory (BFI-K): *Diagnostica* 51, 4 (2005), 195–206.

[233] Randall, A., Liu, E., Akiwate, G., Padmanabhan, R., Voelker, G. M., Savage, S., and Schulman, A. Trufflehunter: Cache Snooping Rare Domains at Large Public DNS Resolvers. In: *Proceedings of the ACM Internet Measurement Conference*. IMC '20. ACM, Virtual Event, USA, 2020, 50–64.

[234] Rastogi, V., Chen, Y., and Jiang, X. DroidChameleon: Evaluating Android Anti-Malware against Transformation Attacks. In: *Proceedings of the 8th ACM SIGSAC Symposium on Information, Computer and Communications Security*. ASIA CCS '13. ACM, Hangzhou, China, 2013, 329–334.

[235] Reason, J. The Contribution of Latent Human Failures to the Breakdown of Complex Systems. *Philosophical Transactions of the Royal Society B* 327 (1990), 475–484.

[236] Reddit, T. *FAQ: "Should I Use a VPN with Tor?"* 2021. URL: `https://old.reddit.com/r/TOR/wiki/index#wiki_should_i_use_a_vpn_with_tor.3F_tor_over_vpn.2C_or_vpn_over_tor.3F` (visited on 11/22/2022).

[237] Redmiles, E. M., Kross, S., and Mazurek, M. L. How I Learned to be Secure: a Census-Representative Survey of Security Advice Sources and Behavior. In: *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. CCS '16. ACM, Vienna, Austria, 2016, 666–677.

[238] Redmiles, E. M., Kross, S., and Mazurek, M. L. 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)*. 2019, 1326–1343.

[239] Redmiles, E. M., Malone, A. R., and Mazurek, M. L. I Think They're Trying to Tell Me Something: Advice Sources and Selection for Digital Security. In: *2016 IEEE Symposium on Security and Privacy (SP)*. IEEE, San Jose, CA, USA, 2016, 272–288.

[240] Redmiles, E. M., Warford, N., Jayanti, A., and Koneru, A. A Comprehensive Quality Evaluation of Security and Privacy Advice on the Web. In: *Proceedings of the 29th USENIX Security Symposium*. USENIX Security 2020. USENIX Association, 2020, 89–108.

[241] Reeder, R. W., Ion, I., and Consolvo, S. 152 Simple Steps to Stay Safe Online: Security Advice for Non-Tech-Savvy Users. *IEEE Security & Privacy* 15, 5 (2017), 55–64.

[242] Renaud, K., Volkamer, M., and Renkema-Padmos, A. Why Doesn't Jane Protect Her Privacy? In: *International Symposium on Privacy Enhancing Technologies (PETS 2014)*. Springer International Publishing, 2014, 244–262.

[243] Renaud, K. and Warkentin, M. Risk Homeostasis in Information Security: Challenges in Confirming Existence and Verifying Impact. In: *Proceedings of the 2017 New Security Paradigms Workshop*. ACM, Santa Cruz, CA, USA, 2017, 57–69.

[244] Roundy, K. A., Mendelberg, P. B., Dell, N., McCoy, D., Nissani, D., Ristenpart, T., and Tamersoy, A. The Many Kinds of Creepware Used for Interpersonal Attacks. In: *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE, San Francisco, CA, USA, 2020, 626–643.

[245] Ruoti, S., Andersen, J., Heidbrink, S., O'Neill, M., Vaziripour, E., Wu, J., Zappala, D., and Seamons, K. We're on the Same Page: A Usability Study of Secure Email Using Pairs of Novice Users. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, 2016, 4298–4308.

[246] Ruoti, S. and Seamons, K. Standard Metrics and Scenarios for Usable Authentication. In: *Who Are You?! Adventures in Authentication Workshop (WAY)*. USENIX Association, 2016.

[247] Saffer, D. *Designing for Interaction, Second Edition: Creating Innovative Applications and Devices*. New Riders, 2010.

[248] Saltzer, J. and Schroeder, M. The protection of information in computer systems. *Proceedings of the IEEE* 63, 9 (1975), 1278–1308.

[249] Sanchez-Rola, I., Dell'Amico, M., Kotzias, P., Balzarotti, D., Bilge, L., Vervier, P.-A., and Santos, I. Can I Opt Out Yet?: GDPR and the Global Illusion of Cookie Control. en. In: *Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security*. AsiaCCS. ACM, Auckland New Zealand, July 2019, 340–351.

[250] Satchell, C. and Dourish, P. Beyond the user: use and non-use in HCI. In: *Proceedings of the 21st conference of the computer-human interaction special interest group of Australia on Computer-human interaction: design (OZCHI '09)*. ACM, 2009, 9–16.

[251] Schellevis, J. *Beveiliging door politie gekraakte 'cryptofoons' was twijfelachtig*. Dutch. 2018. URL: `https://www.nos.nl/l/2258309` (visited on 01/07/2021).

[252] Schneier, B. *Beyond Fear: Thinking Sensibly about Security in an Uncertain World*. Copernicus Books, New York, NY, USA, 2003.

[253] Schneier, B. The Psychology of Security. In: *AFRICACRYPT*. 2008, 30.

[254] Schröder, S., Huber, M., Wind, D., and Rottermanner, C. When SIGNAL hits the Fan: On the Usability and Security of State-of-the-Art Secure Mobile Messaging. In: *European Workshop on Usable Security*. EuroUSEC 2016. IEEE, Darmstadt, Germany, 2016, 1–7.

[255] Shankar, U. and Karlof, C. Doppelganger: Better browser privacy without the bother. en. In: *Proceedings of the 13th ACM Conference on Computer and Communications Security (CCS '06)*. CCS. ACM, Alexandria, VA, USA, 2006, 154–167.

[256] Simpson, A. K., Roesner, F., and Kohno, T. Securing vulnerable home IoT devices with an in-hub security manager. In: *2017 IEEE International Conference on Pervasive Computing and Communications Workshops*. 2017, 551–556.

[257] Slupska, J. and Tanczer, L. M. Threat Modeling Intimate Partner Violence: Tech Abuse as a Cybersecurity Challenge in the Internet of Things. In: *The Emerald International Handbook of Technology-Facilitated Violence and Abuse*. Ed. by Bailey, J., Flynn, A., and Henry, N. Emerald Publishing Limited, June 2021, 663–688.

[258] Son, J. Y., Doumas, L. A. A., and Goldstone, R. L. When Do Words Promote Analogical Transfer? *The Journal of Problem Solving* 3, 1 (2010).

[259] Spencer, R. The Streamlined Cognitive Walkthrough Method, Working around Social Constraints Encountered in a Software Development Company. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '00. ACM, The Hague, The Netherlands, 2000, 353–359.

[260] Spero, E. and Biddle, R. Out of Sight, Out of Mind: UI Design and the Inhibition of Mental Models of Security. In: *New Security Paradigms Workshop 2020*. NSPW '20. ACM, Online, USA, 2020, 127–143.

[261] Spiel, K. "Why are they all obsessed with Gender?" — (Non)binary Navigations through Technological Infrastructures. In: *Designing Interactive Systems Conference 2021*. DIS '21. ACM, Virtual Event, USA, June 2021, 478–494.

[262] Spinuzzi, C. The Methodology of Participatory Design. *Technical Communication* 52, 2 (2005), 163–174.

[263] Stephens, K., Butler, M., Holloway, L. M., Goncu, C., and Marriott, K. Smooth Sailing? Autoethnography of Recreational Travel by a Blind Person. In: *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. ASSETS '20. ACM, Virtual Event, Greece, Oct. 2020, 1–12.

[264] Story, P., Smullen, D., Yao, Y., Acquisti, A., Cranor, L. F., Sadeh, N., and Schaub, F. Awareness, Adoption, and Misconceptions of Web Privacy Tools. *Proceedings on Privacy Enhancing Technologies* 2021, 3 (2021), 308–333.

[265] Strohmayer, A., Bellini, R., and Slupska, J. Safety as a Grand Challenge in Pervasive Computing: Using Feminist Epistemologies to Shift the Paradigm From Security to Safety. *IEEE Pervasive Computing* 21, 3 (July 2022), 61–69.

[266] Tabriz, P. *Optimistic Dissatisfaction with the Status Quo: Steps We Must Take to Improve Security in Complex Landscapes*. en. Aug. 2018. URL: `https://www.youtube.com/watch?v=py2qmGbyhlw` (visited on 02/11/2021).

[267] Tan, J., Bauer, L., Bonneau, J., Cranor, L. F., Thomas, J., and Ur, B. Can Unicorns Help Users Compare Crypto Key Fingerprints? In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, 2017, 3787–3798.

[268] Tang, J., Birrell, E., and Lerner, A. Replication: How Well Do My Results Generalize Now? The External Validity of Online Privacy and Security Surveys. In: *Proceedings of the Eighteenth Symposium on Usable Privacy and Security*. SOUPS. USENIX Association, Boston, MA, USA, 2022.

[269] Tian, Y., Liu, B., Dai, W., Ur, B., Tague, P., and Cranor, L. F. Supporting Privacy-Conscious App Update Decisions with User Reviews. In: *Proceedings of the 5th Annual ACM CCS Workshop on Security and Privacy in Smartphones and Mobile Devices*. SPSM '15. ACM, Denver, CO, USA, 2015, 51–61.

[270] Tiefenau, C., Häring, M., and Krombholz, K. Security, Availability, and Multiple Information Sources: Exploring Update Behavior of System Administrators. In: *Proceedings of the Sixteenth Symposium on Usable Privacy and Security*. 2020, 239–258.

[271] TOP10VPN and globalwebindex. *Global VPN Usage Report 2020: An Exploration of VPNs and Their Users around the World*. 2020. URL: https://www.top10vpn.com/assets/2020/03/Top10VPN-GWI-Global-VPN-Usage-Report-2020.pdf (visited on 11/22/2022).

[272] TorProject Trac. *Tor + VPN*. June 2022. URL: https://trac.torproject.org/projects/tor/wiki/doc/TorPlusVPN (visited on 11/22/2022).

[273] Traudt, M. *VPN + Tor: Not Necessarily a Net Gain*. Nov. 2016. URL: https://matt.traudt.xyz/posts/2016-11-12-vpn-tor-not-net-gain/ (visited on 11/22/2022).

[274] Trevisan, M., Traverso, S., Bassi, E., and Mellia, M. 4 Years of EU Cookie Law: Results and Lessons Learned. In: *Proceedings on Privacy Enhancing Technologies*. Vol. 2. PETS. Sciendo, Berlin, Germany, 2019, 126–145.

[275] Tseng, E., Bellini, R., McDonald, N., Danos, M., Greenstadt, R., McCoy, D., Dell, N., and Ristenpart, T. The Tools and Tactics Used in Intimate Partner Surveillance: An Analysis of Online Infidelity Forums. In: *Proceedings of the 29th USENIX Conference on Security Symposium*. USENIX Association, 2020, 1893–1909.

[276] Turner, S., Nurse, J. R., and Li, S. "It was hard to find the words": Using an Autoethnographic Diary Study to Understand the Difficulties of Smart Home Cyber Security Practices. In: *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*. CHI '22 Extended Abstracts. ACM, New Orleans, LA, USA, Apr. 2022, 1–8.

[277] Ur, B., Leon, P. G., Cranor, L. F., Shay, R., and Wang, Y. Smart, useful, scary, creepy: perceptions of online behavioral advertising. en. In: *Proceedings of the Eighth Symposium on Usable Privacy and Security - SOUPS '12*. SOUPS. ACM Press, Washington, D.C., 2012, 1–15.

[278] Ur, B., Noma, F., Bees, J., Segreti, S. M., Shay, R., Bauer, L., Christin, N., and Cranor, L. F. "I Added '!' at the End to Make It Secure": Observing Password Creation in the Lab. In: *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. USENIX Association, Ottawa, ON, Canada, 2015, 123–140.

[279] Uzun, E., Saxena, N., and Kumar, A. Pairing Devices for Social Interactions: A Comparative Usability Evaluation. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '11. ACM, Vancouver, BC, Canada, 2011, 2315–2324.

[280] Vaniea, K. and Rashidi, Y. Tales of Software Updates: The process of updating software. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2016, 3215–3226.

[281] Vaniea, K. E., Rader, E., and Wash, R. Betrayed by Updates: How Negative Experiences Affect Future Security. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '14. ACM, Toronto, ON, Canada, 2014, 2671–2674.

[282] Vaziripour, E., Farahbakhsh, R., O'Neill, M., Wu, J., Seamons, K., and Zappala, D. A Survey Of the Privacy Preferences and Practices of Iranian Users of Telegram. In: *Workshop on Usable Security*. USEC 2018. Internet Society, San Diego, CA, USA, 2018, 1–20.

[283] Vaziripour, E., Howard, D., Tyler, J., O'Neill, M., Wu, J., Seamons, K., and Zappala, D. I Don't Even Have to Bother Them!: Using Social Media to Automate the Authentication Ceremony in Secure Messaging. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. ACM, Glasgow, Scotland, UK, 2019, 1–12.

[284] Vaziripour, E., Wu, J., O'Neill, M., Clinton, R., Whitehead, J., Heidbrink, S., Seamons, K., and Zappala, D. Is that you, Alice? A Usability Study of the Authentication Ceremony of Secure Messaging Applications. In: *Thirteenth Symposium on Usable Privacy and Security*. SOUPS 2017. USENIX Association, Santa Clara, CA, USA, 2017, 29–47.

[285] Vaziripour, E., Wu, J., O'Neill, M., Metro, D., Cockrell, J., Moffett, T., Whitehead, J., Bonner, N., Seamons, K., and Zappala, D. Action Needed! Helping Users Find and Complete the Authentication Ceremony in Signal. In: *Proceedings of the Fourteenth Symposium on Usable Privacy and Security*. SOUPS 2018. USENIX Association, Baltimore, MD, USA, 2018, 47–62.

[286] Verizon. *Verizon Wireless' Use of a Unique Identifier Header (UIDH)*. 2021. URL: `https://www.verizon.com/support/unique-identifier-header-faqs/` (visited on 02/11/2021).

[287] Vidstrom, A. *The Legal Boundaries of Reverse Engineering in the EU*. May 2019. URL: `https://vidstromlabs.com/blog/the-legal-boundaries-of-reverse-engineering-in-the-eu/` (visited on 06/08/2022).

[288] Voskobojnikov, A., Wiese, O., Mehrabi Koushki, M., Roth, V., and Beznosov, K. The U in Crypto Stands for Usable: An Empirical Study of User Experience with Mobile Cryptocurrency Wallets. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama, Japan, 2021.

[289] Wash, R. Folk Models of Home Computer Security. In: *Proceedings of the Sixth Symposium on Usable Privacy and Security*. SOUPS '10. ACM, Redmond, WA, USA, 2010, 1–16.

[290] Wash, R., Rader, E., Vaniea, K., and Rizor, M. Out of the loop: How automated software updates cause unintended security consequences. In: *Symposium on Usable Privacy and Security*. 2014, 89–104.

[291] Weber, S., Harbach, M., and Smith, M. Participatory Design for Security-Related User Interfaces. In: *Workshop on Usable Security (USEC '15)*. Internet Society, 2015.

[292] Wei, M., Zeng, E., Kohno, T., and Roesner, F. Anti-Privacy and Anti-Security Advice on TikTok: Case Studies of Technology-Enabled Surveillance and Control in Intimate Partner and Parent-Child Relationships. In: *Proceedings of the Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*. USENIX Association, 2022, 447–462.

[293] Weinberg, Z., Cho, S., Christin, N., Sekar, V., and Gill, P. How to Catch when Proxies Lie: Verifying the Physical Locations of Network Proxies with Active Geolocation. In: *Proceedings of the Internet Measurement Conference 2018*. ACM, Boston, MA, USA, 2018, 203–217.

[294] Wharton, C., Rieman, J., Lewis, C., and Polson, P. The Cognitive Walkthrough Method: A Practitioner's Guide. In: *Usability Inspection Methods*. John Wiley & Sons, Inc., 1994, 105–140.

[295] Whitten, A. and Tygar, D. J. Why johnny can't encrypt: A usability evaluation of PGP 5.0. In: USENIX Security Symposium. Vol. 348. 1999.

[296] Wiegers, K. *Designing around Bad Actors and Dangerous Actions*. Feb. 2021. URL: `https://uxdesign.cc/designing-around-bad-actors-and-dangerous-actions-8fc7984c510d` (visited on 06/08/2022).

[297] Willcox, G. The Feeling Wheel: A Tool for Expanding Awareness of Emotions and Increasing Spontaneity and Intimacy. *Transactional Analysis Journal* 12, 4 (Oct. 1982), 274–276.

[298] Wu, J., Gattrell, C., Howard, D., Tyler, J., Vaziripour, E., Seamons, K., and Zappala, D. "Something isn't secure, but I'm not sure how that translates into a problem": Promoting autonomy by designing for understanding in Signal. In: *Fifteenth Symposium on Usable Privacy and Security*. SOUPS 2019. USENIX Association, Santa Clara, CA, USA, 2019, 137–154.

[299] Wu, Y., Edwards, W. K., and Das, S. SoK: Social Cybersecurity. In: *2022 IEEE Symposium on Security and Privacy (SP)*. IEEE, San Francisco, CA, USA, 2022, 1863–1879.

[300] Xie, Z. and Zhu, S. AppWatcher: Unveiling the Underground Market of Trading Mobile App Reviews. In: *Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks*. WiSec '15. ACM, New York, NY, USA, 2015, 1–11.

[301] Xie, Z., Zhu, S., Li, Q., and Wang, W. You Can Promote, but You Can't Hide: Large-Scale Abused App Detection in Mobile App Stores. In: *Proceedings of the 32nd Annual Conference on Computer Security Applications*. ACSAC '16. ACM, Los Angeles, CA, USA, 2016, 374–385.

[302] Yao, Y., Lo Re, D., and Wang, Y. Folk Models of Online Behavioral Advertising. en. In: *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. CSCW. ACM, Portland, OR, USA, 2017, 1957–1969.

[303] Yue, C., Xie, M., and Wang, H. An automatic HTTP cookie management system. *Computer Networks* 54, 13 (2010), 2182–2198.

[304] Zeng, E. and Rösner, F. Understanding and Improving Security and Privacy in Multi-User Smart Homes: A Design Exploration and In-Home User Study. In: *Proceedings of the 28th USENIX Security Symposium*. USENIX Association, Santa Clara, CA, USA, 2019, 159–176.

[305] Zhai, J., Ajmal, H., and Su, J. *Preying on Insecurity: Placebo Applications With No Functionality on Google Play and Amazon.Com*. June 2014. URL: `https://www.fireeye.com/blog/threat-research/2014/06/preying-on-insecurity-placebo-applications-with-no-functionality-on-google-play-and-amazon-com.html` (visited on 06/08/2022).

[306] Zheng, M., Lee, P. P. C., and Lui, J. C. S. ADAM: An automatic and extensible platform to stress test android anti-virus systems. In: *Detection of Intrusions and Malware, and Vulnerability Assessment*. Ed. by Flegel, U., Markatos, E., and Robertson, W. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, 82–101.

[307] Zichermann, G. and Cunningham, C. *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps*. 1st. O'Reilly Media, Sebastopol, CA, USA, 2011.

[308] Zou, Y., Roundy, K., Tamersoy, A., Shintre, S., Roturier, J., and Schaub, F. Examining the Adoption and Abandonment of Security, Privacy, and Identity Theft Protection Practices. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu, HI, USA, 2020, 1–15.

[309] Zuboff, S. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile books, London, UK, 2018.

[310] Zurko, M. E. and Simon, R. T. User-Centered Security. In: *Proceedings of the 1996 Workshop on New Security Paradigms (NSPW 1996)*. NSPW. 1996, 27–33.