Exploring User-Centered Security Design for Usable Authentication Ceremonies

Matthias Fassl
matthias.fassl@cispa.saarland
CISPA Helmholtz Center for Information Security
Saarland University

Lea Gröber
lea.groeber@cispa.saarland
CISPA Helmholtz Center for Information Security
Saarland University

Katharina Krombholz
krombholz@cispa.saarland
CISPA Helmholtz Center for Information Security

ABSTRACT

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.

CCS CONCEPTS

• **Security and privacy → Authentication**; **Usability in security and privacy**; *Key management*; • **Human-centered computing → Participatory design**; **Interface design prototyping**.

KEYWORDS

Instant Messaging; Man-in-the-Middle (MitM); Authentication; Usability; User-Centered Design

ACM Reference Format:
Matthias Fassl, Lea Gröber, and Katharina Krombholz. 2021. Exploring User-Centered Security Design for Usable Authentication Ceremonies. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3411764.3445164

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

ACM ISBN 978-1-4503-8096-6/21/05…$15.00
https://doi.org/10.1145/3411764.3445164

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 [31]. Investigations by a Dutch news network [19, 36] 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 [41], and those who are aware have problems conducting them correctly [22, 37, 44].

Most of these authentication ceremonies are based on much older methods to securely pair multiple devices owned by the same user [4, 18, 23, 25]. 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 [40]. After discovering that existing ceremonies do not work well, prior work focused primarily on incremental improvements [43, 46] or proposed to remove users from the loop [29, 42] (which potentially leads to a lack of trust [26, 32]). 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.

2 RELATED WORK AND BACKGROUND

Designing secure experiences. Dan Saffer [34] 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.

Using this categorization, we classify the creation of the original authentication ceremonies as a mixture of systems design and genius design. The adoption of device pairing mechanisms that were not intended for this use case corresponds to a systems design approach, and the lack of considerations for the users' needs is common in genius design. The suggestions for improvements that we presented above [39, 42, 43, 46] mainly employ activity-centered design. They all closely examine the process of authentication and try to remove barriers, which is important and effective work. However, more fundamental questions about users' needs and goals can easily be overlooked when continuous improvement of existing systems is the goal.

Zurko et al. [47] already advocated for user-centered security design in the 1990s. In 2017, Dodier-Lazaro et al. [14] condemned that many security "improvements" stem from a paternalistic mindset that ignores users' values. Mathiasen et al. [26, 27] proposed an experience-based design approach to enable secure experiences. Weber et al. [45] used participatory design to draft SSL warning messages. Gorski et al. [20] used a similar approach to create warning messages for developers that use cryptographic libraries.

History of pairing mechanisms. In the early 2000s, the secure pairing of devices emerged as a heavily studied research topic [4, 18, 23, 28]. 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. [25] 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.

In 2011, Uzun et al. [40] criticized that: (1) existing pairing methods have been devised by security professionals with little regard for their usability, and (2) that it is not possible to reduce the problem of social pairing to personal pairing of devices. Social pairing introduces an additional layer of necessary interaction and adds social context such as potential embarrassment or discomfort to the problem. Ignoring those social aspects of security effects may have severe downsides, as suggested by the increasing body of research about their importance for security adoption. Gaw et al. [17] found that even employees of a security-concerned activist group considered the social implications before sending encrypted mails. Having the recipients invest extra effort to decrypt "unimportant" mails was considered rude. Das et al. present qualitative [11] and quantitative [12] evidence that social processes affect the (non-)adoption of security features. Exposure to many feature-adopting friends increases the likelihood of adoption and vice versa. Abu-Salma et al. [1] found that an important factor for the adopting secure messengers is the friends' opinion about its security. Ruoti et al. [32] showed that hiding security mechanisms can decrease users' trust in them.

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. [22] 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. [37] found that the majority of participating CS students failed to detect and mitigate MitM attacks using Signal's authentication ceremony. Vaziripour et al. [44] 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. [41] 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. [39] focused on improving success rates by identifying a suitable key representation and mode of comparison. Vaziripour et al. [43] streamlined Signal's ceremony by providing easy access and additional guidance. Wu et al. [46] produced new visual indicators, new notification dialogs, and a new simplified notification flow. In the second category, Vaziripour et al. [42] partially removed users from the loop by using Keybase for an authentication method based on social media. Melara et al. [29] 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 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.

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 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 [38]. We structured these workshops similar to the participatory design studies by Weber et al. [45] and Gorski et al. [20]: (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 4 provides a detailed description of the collaborative design workshops.

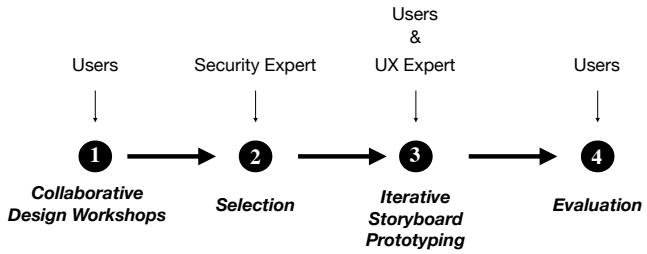


Figure 1: Overview of the design-process and the involved parties. We describe each of the steps in a separate section in the remaining paper.

(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 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 also be used to analyze the work process of prospective users and simulate possible future work [8]. 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 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 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.

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. [45] 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 (included in the supplementary material) 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 intercoder 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 is presented in the supplementary material. 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.

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 foolproof 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.

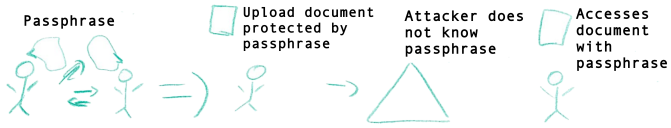


Figure 2: In-person meeting for exchanging a passphrase that protects against attackers.

Figure 3: Conversation partners need a previously exchanged password to access the conversation.

Figure 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 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 communication partner performing a specified task. Figure 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 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

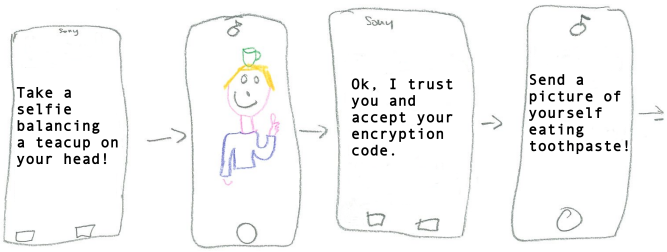


Figure 4: Conversation partners authenticate each other by taking pictures of themselves executing a task defined by the other partner.

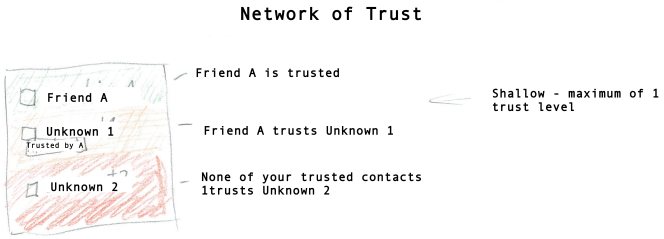


Figure 5: A contact list showing the origin of trust information and color-coded entries based on the trust status.

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.

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 [2] 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. [2] mentioned, even secrets with a very low entropy are secure against offline dictionary or brute-force attack. Boudot et al. [6] 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 [19]. 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 6). In this case, they meet in person and the compare keys automatically as suggested by Tan et al. [39]. 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.

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 [16]).

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 [16]).

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.

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' 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.[13]. 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.

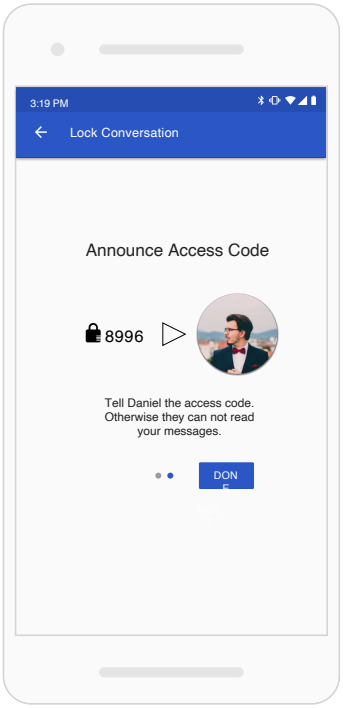
At the end of the second meeting, the UX expert suggested several sources of design inspiration, helpful books, and design tools that could help in the future. Figure 6 shows impressions of the three resulting prototypes.

7 EVALUATION

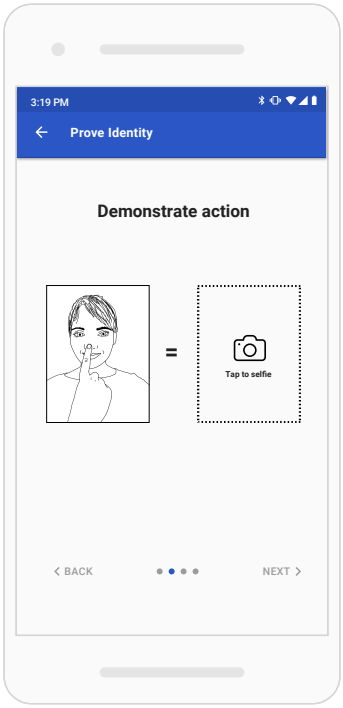
During the collaborative design workshops, we found that (1) participants who understood the purpose and consequence of the ceremony were willing to invest an additional effort for some of their contacts, and (2) some participants felt reassured about the security if they were able to participate in 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 include the three developed prototypes and Signal's current authentication ceremony in the same presentation format and design language as the other prototypes. All click-through prototypes are provided in the supplementary material.

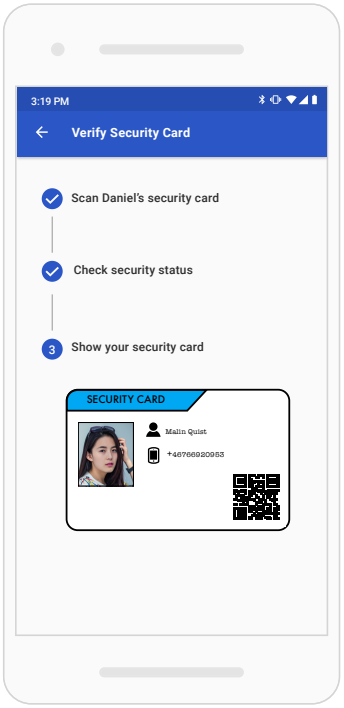
At the start of the survey, we presented the same messaging-related scenario to all participants. This scenario introduces a threat model: the potential risk of losing one's job if the messages are intercepted or sent to the wrong person – this was chosen since (a)



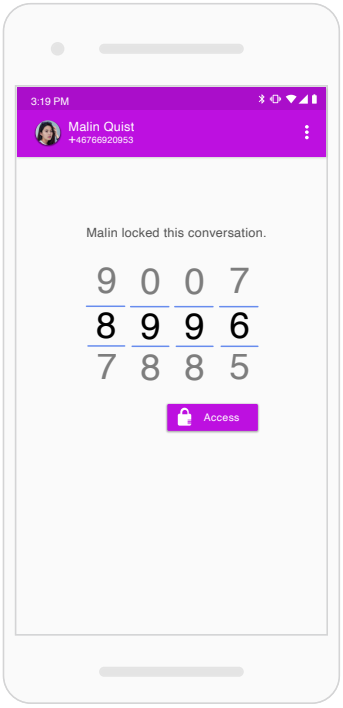
(a) Combination lock based prototype (Malin)



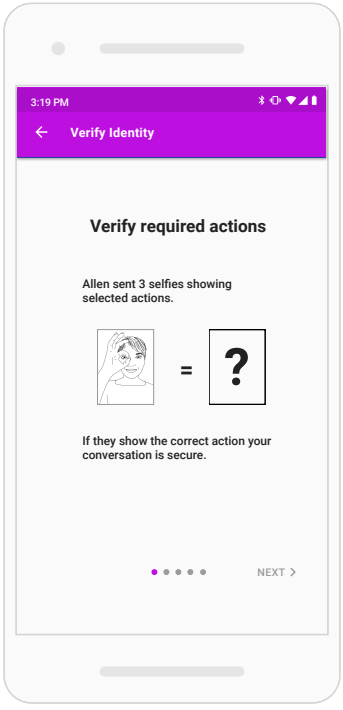
(b) Selfies based prototype (Malin)



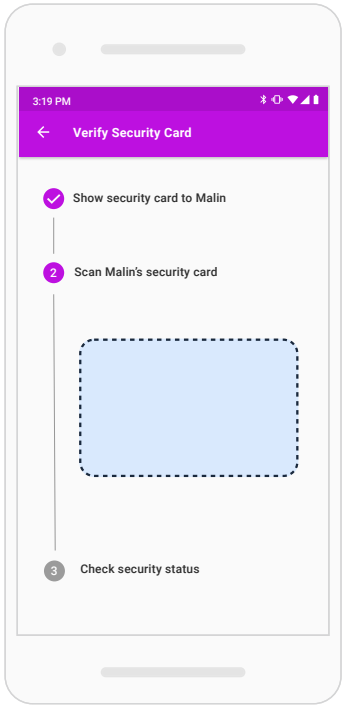
(c) ID card based prototype (Malin)



(d) Combination lock based prototype (Daniel)



(e) Selfies based prototype (Daniel)



(f) ID card based prototype (Daniel)

Figure 6: Impression of the prototypes' authentication interaction between the fictional characters Malin (blue) and Daniel (magenta).

Table 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 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

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 in the supplementary material). 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 the supplementary material.

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. [24]). 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 [16]. 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. The supplementary material includes a table with all collected demographic information.

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 [16].

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* [5, 33]. 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 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 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 3 shows the result of all planned contrasts on the outcome measure *Threat Models*.

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

Table 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

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 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 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 4 and 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 (included in the supplementary material) 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 [3, 7] 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. [46] 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.

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.

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 [9]. Similarly, it is a commonly observed issue in User-Centered Design that designers tend to imagine users that are similar to themselves [10, 30]. 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. [15] 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 [35]. 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. [21] 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.

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 the supplementary material) 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. [44] 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 [29]) 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.

10 CONCLUSION

Authentication ceremonies in secure instant messaging are a well-researched security problem [22, 37, 39, 41–44, 46]. 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.

REFERENCES

[1] Ruba Abu-Salma, M. Angela Sasse, Joseph Bonneau, Anastasia Danilova, Alena Naiakshina, and Matthew Smith. Obstacles to the Adoption of Secure Communication Tools. In *2017 IEEE Symposium on Security and Privacy (S&P 2017)*, pages 137–153. IEEE, 2017.

[2] Chris Alexander and Ian Goldberg. Improved user authentication in off-the-record messaging. In *Proceedings of the 2007 ACM Workshop on Privacy in Electronic Society (WPES 2007)*, pages 41–47. ACM, 2007.

[3] Morten Sieker Andreasen, Henrik Villemann Nielsen, Simon Ormholt Schrøder, and Jan Stage. 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)*, pages 1405–1414. ACM, 2007.

[4] Dirk Balfanz, DK Smetters, Paul Stewart, and H Chi Wong. Talking To Strangers: Authentication in Ad-Hoc Wireless Networks. In *Proceedings of Network and Distributed System Security Symposium 2002 (NDSS '02)*. Internet Society, 2002.

[5] Aaron Bangor, Philip T. Kortum, and James T. Miller. An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction*, 24(6):574–594, 2008.

[6] Fabrice Boudot, Berry Schoenmakers, and Jacques Traoré. A fair and efficient solution to the socialist millionaires' problem. *Discrete Applied Mathematics*, 111(1-2):23—36, 2001.

[7] Anders Bruun, Peter Gull, Lene Hofmeister, and Jan Stage. 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)*, pages 1619–1628. ACM, 2009.

[8] Susanne Bødker and Kaj Grønbæk. Cooperative Prototyping: Users and Designers in Mutual Activity. *International Journal of Man-Machine Studies*, 34:453 — 478, 1991.

[9] Geoff Cooper and John Bowers. Representing the user: Notes on the disciplinary rhetoric of human-computer interaction. In Peter J. Thomas, editor, *The Social and Interactional Dimensions of Human-Computer Interfaces*, Cambridge Series on Human-Computer Interaction, pages 48—66. Cambridge University Press, 1995.

[10] Sasha Costanza-Chock. *Design Justice: Community-Led Practices to Build the Worlds We Need*. Information Policy. MIT Press, 2020.

[11] Sauvik Das, Tiffany Hyun-Jin Kim, Laura A Dabbish, and Jason I Hong. The Effect of Social Influence on Security Sensitivity. In *10th Symposium On Usable Privacy and Security (SOUPS 2014)*, pages 143–157. USENIX Association, 2014.

[12] Sauvik Das, Adam D.I. Kramer, Laura A. Dabbish, and Jason I. Hong. 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)*, pages 1416–1426. ACM, 2015.

[13] Verena Distler, Marie-Laure Zollinger, Carine Lallemand, Peter B Roenne, Peter Y A Ryan, and Vincent Koenig. Security - Visible, Yet Unseen? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2019)*, pages 1–13. ACM, 2019.

[14] Steve Dodier-Lazaro, Ruba Abu-Salma, Ingolf Becker, and M. Angela Sasse. From Paternalistic to User-Centred Security: Putting Users First with Value-Sensitive Design. In *Workshop on Values in Computing*. ACM, 2017.

[15] Ksenia Ermoshina, Harry Halpin, and Francesca Musiani. Can Johnny build a protocol? Co-ordinating developer and user intentions for privacy-enhanced secure messaging protocols. In *The 2nd European Workshop on Usable Security (EuroUSEC)*. IEEE, 2017.

[16] Thomas Franke, Christiane Attig, and Daniel Wessel. 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):456—467, 2019.

[17] Shirley Gaw, Edward W Felten, and Patricia Fernandez-Kelly. Secrecy, Flagging, and Paranoia: Adoption Criteria in Encrypted Email. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*, pages 591–600. ACM, 2006.

[18] Christian Gehrmann, Chris J. Mitchell, and Kaisa Nyberg. Manual authentication for wireless devices. *RSA Cryptobytes*, 7(1):29—37, 2004.

[19] Dan Goodin. Police decrypt 258,000 messages after breaking pricey IronChat crypto app. https://arstechnica.com/?p=1408441, 2018. Accessed: 2021-1-7.

[20] Peter Leo Gorski, Yasemin Acar, Luigi Lo Iacono, and Sascha Fahl. 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)*, pages 1—13. ACM, 2020.

[21] Saul Greenberg and Bill Buxton. Usability evaluation considered harmful (some of the time). In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*, pages 111–120. ACM, 2008.

[22] Amir Herzberg and Hemi Leibowitz. 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)*, pages 17–28. ACM, 2016.

[23] Lars Erik Holmquist, Friedemann Mattern, Bernt Schiele, Petteri Alahuhta, Michael Beigl, and Hans-W. Gellersen. 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)*, pages 116—122. Springer-Verlag, 2001.

[24] Jason L. Huang, Nathan A. Bowling, Mengqiao Liu, and Yuhui Li. Detecting Insufficient Effort Responding with an Infrequency Scale: Evaluating Validity and Participant Reactions. *Journal of Business and Psychology*, 30(2):299–311, 2015.

[25] Alfred Kobsa, Rahim Sonawalla, Gene Tsudik, Ersin Uzun, and Yang Wang. 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)*, page 12. USENIX Association, 2009.

[26] Niels Raabjerg Mathiasen and Susanne Bødker. Threats or threads: from usable security to secure experience? In *Proceedings of the 5th Nordic Conference on Human-Computer Interaction: Building Bridges (NordiCHI '08)*, pages 283–289. ACM, 2008.

[27] Niels Raabjerg Mathiasen and Susanne Bødker. Experiencing Security in Interaction Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*, pages 2325–2334. ACM, 2011.

[28] Jonathan M. McCune, Adrian Perrig, and Michael K. Reiter. Seeing-Is-Believing: Using Camera Phones for Human-Verifiable Authentication. In *2005 IEEE Symposium on Security and Privacy (S&P '05)*. IEEE, 2005.

[29] Marcela S Melara, Aaron Blankstein, Joseph Bonneau, Edward W Felten, and Michael J Freedman. CONIKS: Bringing Key Transparency to End Users. In *24th USENIX Security Symposium (USENIX Security 15)*, pages 383–398. USENIX Association, 2015.

[30] Nelly Oudshoorn, Els Rommes, and Marcelle Stienstra. Configuring the User as Everybody: Gender and Design Cultures in Information and Communication Technologies. *Science, Technology & Human Values*, 29(1):30–63, 2004.

[31] Press Statement of the Dutch Police. Police have achieved a breakthrough in the interception and decryption of crypto communication. https://www.politie.nl/en/news/2018/november/02-apeldoorn-police-have-achieved-a-breakthrough-in-the-interception-and-decryption-of-crypto-communication.html, 2018. Accessed: 2021-1-7.

[32] Scott Ruoti, Jeff Andersen, Scott Heidbrink, Mark O'Neill, Elham Vaziripour, Justin Wu, Daniel Zappala, and Kent Seamons. 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)*, pages 4298–4308. ACM, 2016.

[33] Scott Ruoti and Kent Seamons. Standard Metrics and Scenarios for Usable Authentication. In *Who Are You?! Adventures in Authentication Workshop (WAY)*. USENIX Association, 2016.

[34] Dan Saffer. *Designing for Interaction, Second Edition: Creating Innovative Applications and Devices*. New Riders, 2010.

[35] Christine Satchell and Paul Dourish. 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)*, pages 9–16. ACM, 2009.

[36] Joost Schellevis. Beveiliging door politie gekraakte 'cryptofoons' was twijfelachtig. https://www.nos.nl/l/2258309, 2018. Accessed: 2021-1-7.

[37] Svenja Schröder, Markus Huber, David Wind, and Christoph Rottermanner. 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, 2016.

[38] Clay Spinuzzi. The Methodology of Participatory Design. *Technical Communication*, 52(2):163—174, 2005.

[39] Joshua Tan, Lujo Bauer, Joseph Bonneau, Lorrie Faith Cranor, Jeremy Thomas, and Blase Ur. Can Unicorns Help Users Compare Crypto Key Fingerprints? In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*, pages 3787–3798. ACM, 2017.

[40] Ersin Uzun, Nitesh Saxena, and Arun Kumar. Pairing Devices for Social Interactions: A Comparative Usability Evaluation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*, pages 2315–2324. ACM, 2011.

[41] Elham Vaziripour, Reza Farahbakhsh, Mark O'Neill, Justin Wu, Kent Seamons, and Daniel Zappala. A Survey Of the Privacy Preferences and Practices of Iranian Users of Telegram. In *Workshop on Usable Security (USEC 2018)*. Internet Society, 2018.

[42] Elham Vaziripour, Devon Howard, Jake Tyler, Mark O'Neill, Justin Wu, Kent Seamons, and Daniel Zappala. 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)*, pages 1–12. ACM, 2019.

[43] Elham Vaziripour, Justin Wu, Mark O'Neill, Daniel Metro, Josh Cockrell, Timothy Moffett, Jornad Whitehead, Nick Bonner, Kent Seamons, and Daniel Zappala. 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)*, pages 47–62. USENIX Association, 2018.

[44] Elham Vaziripour, Justin Wu, Mark O'Neill, Ray Clinton, Jordan Whitehead, Scott Heidbrink, Kent Seamons, and Daniel Zappala. 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)*, pages 29—47. USENIX Association, 2017.

[45] Susanne Weber, Marian Harbach, and Matthew Smith. Participatory Design for Security-Related User Interfaces. In *Workshop on Usable Security (USEC '15)*. Internet Society, 2015.

[46] Justin Wu, Cyrus Gattrell, Devon Howard, Jake Tyler, Elham Vaziripour, Kent Seamons, and Daniel Zappala. "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, 2019.

[47] Mary Ellen Zurko and Richard T. Simon. User-Centered Security. In *Proceedings of the 1996 Workshop on New Security Paradigms (NSPW 1996)*, pages 27—33, 1996.