Action Needed! Helping Users Find and Complete the Authentication Ceremony in Signal

Elham Vaziripour, Justin Wu, Mark O'Neill, Daniel Metro, Josh Cockrell, Timothy Moffett, Jordan Whitehead, Nick Bonner, Kent Seamons, and Daniel Zappala,
Brigham Young University

https://www.usenix.org/conference/soups2018/presentation/vaziripour

This paper is included in the Proceedings of the Fourteenth Symposium on Usable Privacy and Security.

August 12–14, 2018 • Baltimore, MD, USA

ISBN 978-1-939133-10-6

Open access to the Proceedings of the Fourteenth Symposium on Usable Privacy and Security is sponsored by USENIX.

Action Needed! Helping Users Find and Complete the Authentication Ceremony in Signal

Elham Vaziripour, Justin Wu, Mark O'Neill, Daniel Metro, Josh Cockrell,
Timothy Moffett, Jordan Whitehead, Nick Bonner, Kent Seamons, Daniel Zappala
Brigham Young University
elhamvaziripour@byu.edu, justinwu@byu.edu, mto@byu.edu, joshuakcockrell@gmail.com,
danieltmetro@gmail.com, timothytmoffett@gmail.com, jordan9001@gmail.com, j.nick.bonner@gmail.com,
seamons@cs.byu.edu, zappala@cs.byu.edu

ABSTRACT

The security guarantees of secure messaging applications are contingent upon users performing an authentication ceremony, which typically involves verifying the fingerprints of encryption keys. However, recent lab studies have shown that users are unable to do this without being told in advance about the ceremony and its importance. A recent study showed that even with this instruction, the time it takes users to find and complete the ceremony is excessively long—about 11 minutes. To remedy these problems, we modified Signal to include prompts for the ceremony and also simplified the ceremony itself. To gauge the effect of these changes, we conducted a between-subject user study involving 30 pairs of participants. Our study methodology includes no user training and only a small performance bonus to encourage the secure behavior. Our results show that users are able to both find and complete the ceremony more quickly in our new version of Signal. Despite these improvements, many users are still unsure or confused about the purpose of the authentication ceremony. We discuss the need for better risk communication and methods to promote trust.

1. INTRODUCTION

Numerous secure messaging applications [18] have been developed to provide end-to-end encryption for personal communication. These applications typically automate the encryption process as much as possible, in order to provide a simpler experience for their users. However, the confidentiality provided by these applications relies on the integrity of its central servers, which exchange users' public keys automatically. To protect against a man-in-the-middle attack, either through compromise of the server or other means, users need to verify the exchanged keys with their conversation partners. This is typically done by comparing a fingerprint of the public keys. We refer to this verification process as the *authentication ceremony*, and variations of it have been adopted widely in secure messaging applications.

USENIX Symposium on Usable Privacy and Security (SOUPS) 2018.
August 12–14, 2018, Baltimore, MD, USA.

Research using lab studies has reported that users have difficulty performing the authentication ceremony within secure messaging applications [4], and this makes them susceptible to attack [14]. Two recent papers demonstrated that with some instruction about the ceremony itself [8] or the importance of comparing keys [20], users can successfully find and use the authentication ceremony. However, users still took an inordinate amount of time—over 11 minutes on average—to find and complete the ceremony [20].

In this paper we examine whether *opinionated design* can make it easier for users to find and perform the authentication ceremony, without relying on instruction about the importance of the ceremony or providing any details about how the ceremony works. Our use of opinionated design is inspired by work on the security indicators for the Chrome browser [6], which led to greater adherence to SSL warnings, but not necessarily greater comprehension. We apply opinionated design to the Signal messaging application, seeking to make the minimal set of changes needed to encourage users to find and perform the ceremony. Our design principles follow recommendations from Schröder et al. [14] in their study of the Signal application. We seek to improve both adherence and performance with respect to finding and using the authentication ceremony, with comprehension a secondary goal. We use Signal because it is open source and because it has been at the forefront of this space, having pioneered the Signal protocol that is also used in WhatsApp, Facebook Messenger, Allo, and Skype.

To test the effectiveness of our design, we created two modifications of Signal, which we label Modification 1 and Modification 2. Modification 1 focuses only on helping users find the authentication ceremony, and the ceremony itself is unchanged. Comparing this version to the original version of Signal enables us to test whether it leads to greater adherence, while also providing a baseline for performance with the original ceremony. Modification 2 incorporates all the changes from the first, and also updates the authentication ceremony to make it easier to use. Comparing Modification 2 to Modification 1 enables us to test for differences in performance among the two authentication ceremonies. We used a between-subject lab study to evaluate the impact of these modifications. We encouraged participants to be security minded by promising them a small monetary bonus. We then observed participant actions, measured their accuracy and time to complete the task, and conducted interviews to

understand their comprehension of the ceremony and their opinions regarding the ceremony.

Our findings include:

- Our modifications of the Signal use interface led to 90% of participants finding the authentication ceremony on their own, combining results for Modification 1 and 2. These modifications included visual cues in the Signal conversation screens to indicate the authentication status of users' contacts, with accompanying actions to initiate the authentication ceremony. Most participants found the authentication ceremony in less than a minute, often within a few seconds. This is compared to a 25% discovery rate for the authentication ceremony among those who used the original version of Signal.
- Our redesigned authentication ceremony was successfully completed by 90% of participants who used Modification 2, as compared to 30% for the original ceremony in Modification 1. The new ceremony clearly separates a QR-code method (for in-person authentication) from a phone call method (when contacts are not in the same location), and uses an in-app phone call modeled after Viber's ceremony. The median time to complete the new authentication ceremony was 2 minutes, as compared to 7 minutes for the few who actually completed the original authentication ceremony.
- Our use of opinionated design, combined with an incentive to be security-minded, resulted in equal or better results than the study by Vaziripour et. al [20], which relied on directly instructing users about the importance of comparing keys. The success rate of 90% is better than the 78% who were successful across all participants and applications in their work, and comparable to the 96% success rate they saw with Viber. Moreover, the time to find the ceremony and complete the ceremony in our modifications (less than a minute, median of 2 minutes) is much lower than in their work (3.5 minutes and 7.8 minutes, respectively).
- Comprehension of the purpose of the ceremony is mixed. Many users associate the ceremony with authentication and confidentiality, but express doubts about their answers. Others clearly do not know what the purpose of the ceremony is. Likewise, while many users express trust in Signal, with further probing many indicate a lack of knowledge or experience to really know if they should trust it. When the purpose of the authentication ceremony is explained to participants, they mostly express a desire to use it, though one third would only use it for some content or with some contacts. This leaves room for future work to further improve the authentication ceremony.

Artifacts: We have created a companion website at `https://action.internet.byu.edu` that provides the source code, study materials, and data.

2. RELATED WORK

The usability of the authentication ceremony for secure messaging applications is a relatively new topic in the field usable security. To the best of our knowledge, there are currently only five papers focused on this topic [20, 14, 4, 8, 1]. The common conclusion of these works is that users are vulnerable to attacks and cannot locate or perform the authentication ceremony without sufficient instruction. This is largely due to users' incomplete mental model of threats and usability problems within secure messaging applications.

Our work has been inspired by one of the most recent studies on the usability of the authentication ceremony in secure messaging applications by Vaziripour et al. [20]. In this work, the authors studied users' ability to locate and perform the authentication ceremony in WhatsApp, Facebook Messenger, and Viber. The first phase of this work instructed participants about potential threats, while the second phase added instruction concerning the necessity of the authentication ceremony. From the first to the second phase, the average ceremony success rate increased from 14% to 79%. It took users, on average, over 3 minutes to find the authentication ceremony and over 7.5 minutes to complete it when they succeeded in the second phase. We borrow some of the methodology from this work.

Our Signal modifications are informed by recommendations from a paper by Schröder et al. that studied the usability of Signal under attack conditions. This study revealed that security experts also are susceptible to man-in-the-middle attacks due to usability problems and incomplete mental models of security. Only seven out of 28 (25%) expert participants successfully authenticated their conversation partners [14]. Asal et al. asked 20 participants to complete authentication by available methods (fingerprint, shared secret, and QR code) in ChatSecure. Herzberg and Leibowitz showed in their study that the majority of users fail to perform the authentication ceremony, and that successes were difficult and time-consuming, even when participants were taught how to authenticate [8]. Abu-Salma et al. conducted a usability study on Telegram to show that the UI was a source of confusion when performing the authentication ceremony [1].

There are several works on the usability of the verification mechanism itself. Shirvanian et al. studied key verification performance by users performing authentication on remote and local conversation partners. They showed that users perform poorly under most key verification methods, especially in the remote case [15]. Independent of a particular application, Tan et al. compared eight representations of authentication material, including textual and graphical representations, with varying degrees of structure, in a simulated attack scenario [17]. They showed that graphical representations were relatively more susceptible to attack but were easy to use, and comparison of graphical forms was quick. Dechand et al. studied textual key verification methods, finding that users are more resistant to attacks when using sentence-based encoding as compared to hexadecimal, alphanumeric, or numeric representations [5]. Sentence-based encoding rated high on usability but low on trustworthiness on a post-study Likert scale.

Another important aspect of our work is the qualitative analysis of users' comments and thinking process to inspect their decision-making processes. A study by Google shows that redesign of Chrome's SSL warnings to promote safe decisions resulted in 30% more users making correct decisions, but found that user comprehension of threats remained low. The authors hypothesized that if users understood the risks better, they would not ignore warnings. [6]. Cormac Herley calculated that the economic cost of time users spend on

Design Principle	Modification 1	Modification 2
Awareness of security status of conversations	Added verification status in conversation list and view (Figures 2a, 2b)	Same as Modification 1
Comprehensible instructions for recommended actions	Added instruction to visit verification screen via button (Figure 2b)	Same as Modification 1 + Separate in-person and remote authentication walkthroughs (Figures 3, 4)
Clear risk communication	None	Inform users of additional actions needed to secure conversations (Figures 3a, 4d)
Easily accessible verification	Clickable action bar in conversations (Figure 2b)	Same as Modification 1 + Clickable action bar in conversations (Figure 3a) and walkthrough (Figures 3, 4)

Table 1: Description of our application of Schröder's design principle recommendations

standard security is substantially higher than the benefits they incur. He argues that users' rejection of security advice is therefore rational economically [7]. Implications for nudging users toward more beneficial and secure choices have been considered recently [3]. Angela Sasse argues that security mechanisms with a high false-positive rate undermine the credibility of security and train users to ignore them [13].

3. MODIFYING SIGNAL

Schröder et al. found several problems with the usability of Signal under attack conditions [14]. They recommend four design principles to overcome these obstacles: awareness of conversation security status, comprehensible instructions for recommended actions, clear risk communication, and easily accessible verification. We applied these principles to redesigning the Signal application and evaluated their effect with a user study. Table 1 outlines our modifications and how they correspond to Schröder's recommendations. We created custom implementations of both the iOS and Android versions of Signal with these changes

We began by creating visual mockups of our modifications to Signal's interface that would employ three of our target design principles. In particular, we provided visual cues to the Signal conversation screens to indicate the verification status of users' contacts, with accompanying actions to initiate the authentication ceremony. Signal already employs a rudimentary indication of verification status in the form of a (hardly noticeable) checkmark under the names of verified contacts, but this is not easily associated with verification status and nothing is shown in the case where a user has not yet verified a contact. We were also careful not to overstate vulnerabilities in our visual cues, in line with recommendations from Sasse [13]. We showed these mockups to 40 university students to gather feedback for various designs, which varied in their use of icons, colors, phrasing, and position of verification status cues and options. We settled on the design as shown because it performed best in our mockups and provided clear warnings. We also used the Signal color scheme and terminology (e.g., *safety number*) for consistency with the original version. Next, we performed a cognitive walkthrough on the modified application to make sure the language used in the interface was clear. Once we were confident in our design, we made the necessary modifications to Signal to implement it. These changes comprise our first modification of Signal (Modification 1).

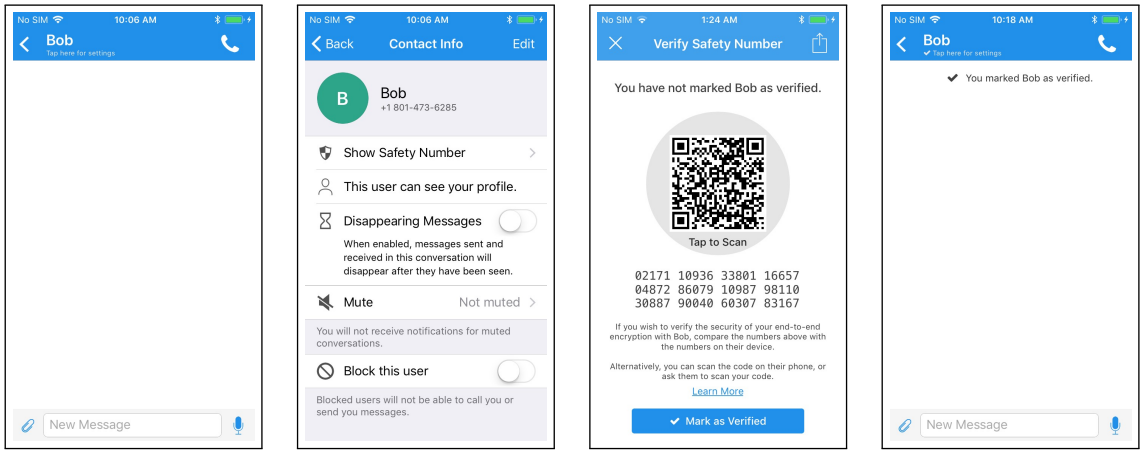
Our second modification of Signal (Modification 2) incorporated all of the changes of the first, but added a set of instructions for users to follow that streamline the authentication ceremony process. In a study by Vaziripour et al. [20], users were more successful performing the authentication ceremony in Viber, and did so in less time compared to other apps. We hypothesize that this was due to Viber providing an in-app phone call that presented encryption keys to users for verification on the same screen. Accordingly, we separate the QR code and phone call verification options in Signal, provide in-app functionality for verification phone calls, and incorporate guiding dialogue to successfully perform verification in each scenario. To develop this second variant of Signal, we conducted a set of pilot studies. We learned that users expect to be able to scan the QR code on each other's phone simultaneously, which could not be done using the original version of Signal. As a result, we modified the application to use a new dual camera/QR code screen.

3.1 Original Signal

Signal [16] uses a Double Ratchet algorithm [12] to update session keys with each exchanged message, which provides forward secrecy for the conversation. Before initiating the ratchet, it uses a triple Diffie-Hellman (3-DH) handshake to exchange public keys. This exchange is automated using a central server. To avoid a man-in-the-middle attack, users must verify the authenticity of the public keys that have been exchanged by the central server. Under Signal, the authentication ceremony is performed using fingerprints from a combination of a user's public key and his/her contact's public key. This fingerprint is called a *safety number*.

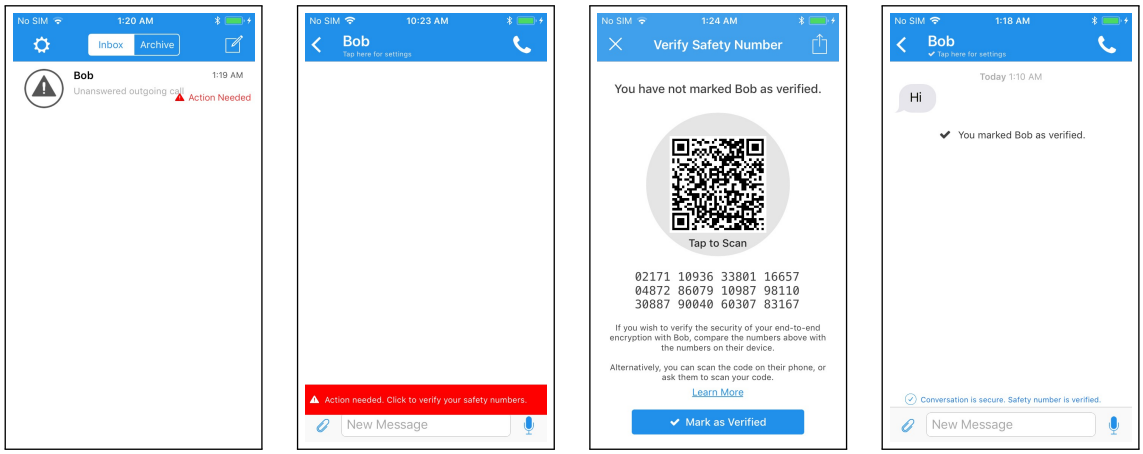
Figure 1 shows the workflow for the authentication ceremony in the current version of Signal. In the conversation screen, after a conversation is initiated with any contact, users can tap on a contact's name in the conversation screen shown in Figure 1a. At this point an option labeled *Show Safety Number* is found, shown in Figure 1b. By selecting this option, users will be transferred to the screen shown in Figure 1c, wherein two options are given to perform the authentication.

Users can either compare their safety numbers directly using their numeric representations, or by scanning an equivalent QR code displayed on their contact's device. After users verify that their safety numbers are equivalent, they are ex-



(a) (b) (c) (d)

Figure 1: Authentication ceremony within the current Signal application



(a) (b) (c) (d)

Figure 2: User interface for finding the authentication ceremony and showing successful verification (Modification 1)

pected to toggle a UI switch captioned *verified*, also shown in Figure 1c, to indicate that they manually verified the numbers to be identical. If users choose to scan the QR code and the result is a successful match, the *verified* switch is changed automatically. Next to the name of verified contacts, the interface places a check mark, shown in Figure 1d, which confirms that the contact has been verified and can be trusted to have a secure conversation with, through the Signal application.

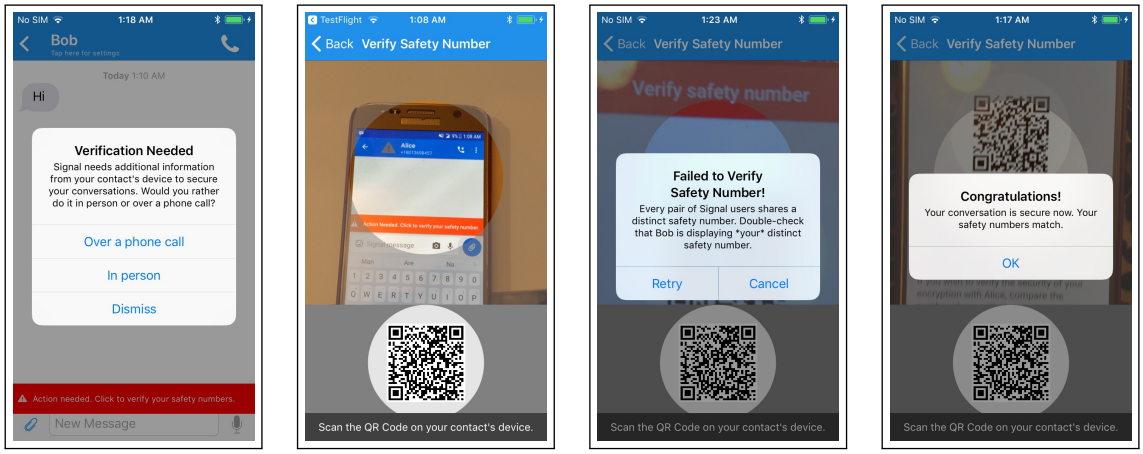
If the encryption keys change for this contact, due to reinstalling the application or a man-in-the-middle attack, users will be prompted to redo the verification process.

3.2 Modification 1

Modification 1 was designed to facilitate the process of finding the authentication ceremony. Users are prompted to perform the authentication ceremony in two locations, as shown in Figure 2. First, in the list of contacts, shown in Figure 2a, any unverified contact has a warning tag indicating *Action Needed*. We also replaced the profile image of unverified contacts with a warning icon until they are verified. Second, in the conversation view depicted in Figure 2b, if the contact is not verified, the bottom of the screen contains a red warning banner with the text *Action needed! Click to verify your safety numbers*. If users notice the red warnings and press either of them, they are directed to the original authentication ceremony screen, shown in Figure 2c. After successful verification, a check mark appears next to the contact name, the red warning band disappears, and each are replaced by a blue message which indicates that the contact has been verified. This text for a conversation window is shown in Figure 2d. The profile image of this contact also shown in favor of the alert icon. Note that in this version users still use the original authentication ceremony.

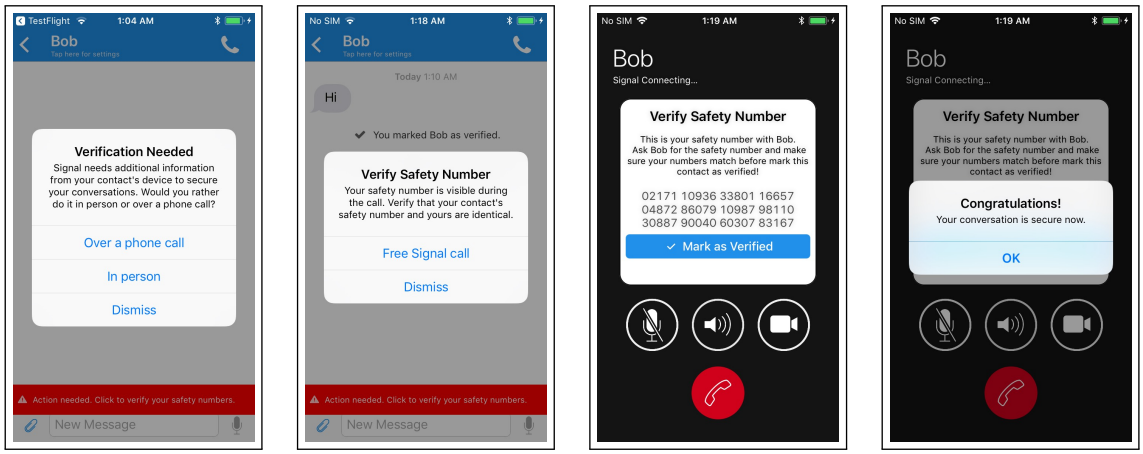
3.3 Modification 2

Our second variant of Signal, Modification 2, was designed to reduce the time required to perform the authentication ceremony. We also attempted to enhance participant under-



(a) (b) (c) (d)

Figure 3: Authentication ceremony for scanning the QR code (Modification 2)



(a) (b) (c) (d)

Figure 4: Authentication ceremony for comparing safety numbers using a phone call (Modification 2)

standing of the purpose of the ceremony, while not necessarily understanding the details of its inner workings.

We separated the two options of scanning the QR code and verifying the safety numbers. Figure 3 and 4 show these modifications. When users press the red warning within the conversation windows, a small dialog appears, shown in Figure 3a, informing users that the verification is necessary for the security of their conversation. They are given two choices of performing the authentication: over a free phone call (via Signal) or in person (QR code scan).

If users choose to verify the safety number in person, as shown in Figure 3, they will be directed to the screen shown in Figure 3b, with the camera activated. In this screen, the local QR code is also shown, allowing the user and his/her contact to scan and verify the safety numbers simultaneously. If the authentication fails, users are given another chance to scan the correct QR code, shown in Figure 3c.

If users instead choose to verify the safety number over a phone call, they will be informed that the call will be free, shown in Figure 4b. We modified the call screen such that immediately after initiating the phone call, users see their safety number with a very brief instruction, shown in Figure 4c. Users are expected to read their safety numbers and ensure they have an identical sequence of numbers. We use a phone call from within Signal because this allows users to see the safety numbers while making a call. Afterward, users press the *Mark as verified* button (iOS) or flip the toggle (Android).

We noticed during pilot studies that users lacked feedback after a successful verification. As a result, contacts who have been verified by the user have a *Verified* tag next to their names in the conversation list, instead of an *Action needed* tag. In addition, the profile image is loaded. During the pilot studies we noticed that users also need feedback to make sure they completed the ceremony correctly, so we created a short congratulation message, shown in Figures 3d and 4d.

4. METHODOLOGY

We conducted an IRB-approved, between-subject user study, examining how participant pairs locate and complete the authentication ceremony across three versions of the Signal secure messaging application. These three versions are the current version of Signal, Modification 1 (with changes to prompt the user to find the authentication ceremony), and Modification 2 (with additional changes to improve the usability of the authentication ceremony). Our study materials are shown in Appendix B.

In the study, we asked participants to complete a scenario wherein one participant needed to send a credit card number to the other participant. The base pay for the study was $7 per participant, with a $3 bonus if they performed the task safely. To avoid any hurt feelings, all participants were given the bonus, but in our observations the bonus served to sufficiently motivate participants to act securely.

We also wanted to test whether we followed Krug's first law of usability—*Don't make me think!* [10]. Thus, we did not provide participants with any instructions on the necessity of performing the authentication (in contrast to [20]), nor did we give them instructions on how to find or complete the authentication ceremony. We gave each participant a time limit of 10 minutes to complete the task, though they were not aware of this limit in advance.

To test each version of Signal equally, we assigned each pair of participants to one of the versions in a round robin manner. Prior to conducting the study a power analysis (described in Appendix A) indicated we needed 10 pairs of participants for each version. During the study subjects installed and used the Signal version to be evaluated on their own mobile devices. The original version of Signal was retrieved from the relevant official app stores for iOS and Android. We uploaded the Android versions of Modification 1 and Modification 2 to Google Play, and we used TestFlight for evaluating our Signal modifications on iOS.

4.1 Task design

In each experiment, the task provided to participants was as follows:

> *You left your credit card at home! You are going to be using the Signal app to ask your friend to send you the credit card number.*
>
> *This is the message you should send to your friend:*
>
> *"Hi! Can you send me my credit card number? I left my card on my desk at home."*
>
> *You can both earn a bonus of $3 for this study if you make sure that nobody can steal this information while your friend is sending it.*

Participant B was instructed similarly:

> *Your friend is going to use the Signal app to ask you for their credit card number. Use the credit card given to you by the study coordinator.*
>
> *You can both earn a bonus of $3 for this study if you make sure that nobody can steal this information while you're sending it.*

Despite a difference in roles, our intention was for both participants to complete the authentication ceremony. Participants were instructed to "talk aloud" as they performed the task, explaining their observations, actions, and reasoning.

Participants failed the task if they sent the credit card number before performing the authentication ceremony correctly, or ten minutes elapsed before completion of the task. In failure cases, participants still performed post-task duties such as responding to questionnaires and interview questions.

During the study, the coordinators checked whether participants had performed the authentication ceremony correctly. If the participants were successful, the coordinators recorded the method used (QR code or comparing the fingerprints verbally in a phone call). If the participants were not successful, the coordinators recorded the reason why.

4.2 Study questionnaire

Participants used a web-based Qualtrics survey on a laptop during the study. This survey both recorded participant answers to various questions both before and after the task, and also briefed them on the task itself. The survey contained:

- A standard set of demographic questions.
- A description of the primary study task, involving the exchange of a credit card number.
- A question asking if the participant believed they had exchanged the credit card number safely, followed by a free-response question to explain the answer.
- A question asking if the participant had seen the authentication ceremony screen (depicted by a screenshot in the survey) during the task. If so, the survey asked the participant several followup questions.
- A question asking if the participant had previously used secure messaging applications to send sensitive information, and the nature of that information.
- A question asking if the participant trusted Signal to be secure, followed by an open-response question to explain the answer.
- A question to rank participant knowledge of computer security.

4.3 Post-study interview

At the conclusion of each study, the coordinators verbally asked each individual participant the following questions:

- We asked participants what features they were looking for to aid in accomplishing the task. This provided us with insight into reasons for success and failure.
- We showed participants how to find the authentication ceremony and asked them to explain how they thought this ceremony helped them (or would have helped them) accomplish the task.
- We asked participants whether they were willing to perform the authentication ceremony before exchanging information with their friends in the future.

We recorded the audio of each study and transcribed the post-study interviews. To analyze the data for open-response questions in the survey and interviews, two authors coded the data together using conventional content analysis. Any disagreements were resolved via discussion. First, we reviewed qualitative comments phrase-by-phrase and word-by-word to assign codes that classified users' comments with regards to a particular topic. Then, we used the constant comparative method to group codes into concepts and organized related categories by merging related codes.

4.4 Study recruitment and design

We recruited pairs of participants on our campus, telling them that each person needed to bring a friend, and that both participants needed to have smartphones. Recruitment proceeded from November 14, 2017, to January 28, 2018, with 41 unique participant pairs recruited in total: 10 pairs for testing each version, eight pairs for pilot studies, and three pairs for replacement. We had to replace the data for three studies, two because the participants had participated in similar studies recently and one because a participant's device had security software that warned them against using our modified version of Signal.

When participants arrived for their scheduled appointment, we presented them with the requisite forms for consent and compensation. We instructed them to download and install the Signal application being tested. We then read them a brief introduction describing the study conditions and their rights as study participants. We informed them that they would be placed in separate rooms. We also informed participants that a study coordinator would be with them at all times and would answer any questions they might have. We let participants choose the study coordinator they would be comfortable working with.

We led the participants to their respective rooms, initiated audio recording, and instructed them to begin the survey. Throughout the study, coordinators were available to answer general questions but were careful not to provide any instructions that would aid in the use of the applications. Sometimes, participants asked if they could meet, and we told them they could. The nature of the scenario led most participants to assume they would not meet.

4.5 Limitations

The scenario we gave participants to exchange a credit card number included telling participants to make sure that no one could steal their information. This caused confusion in one case, when the participants made a phone call through the app in order to perform the authentication ceremony, when they noticed that they could use the same phone call to exchange the credit card number. It may be better to create a scenario where users first validate the safety numbers, then are given a task to exchange the credit card number.

The iOS and Android versions are slightly different. The Android version in Modifications 1 and 2 tells the user that they need to send a message in Signal before they can verify safety numbers. This message appears because the safety number is generated from a combination of local identities and remote identity public keys, and on Android the remote identity key is only received after exchanging the first message. For iOS, this is not the case, and safety number is available before any message exchange.

Due to our method of recruitment, our participants were largely students and their acquaintances, and subsequently exhibited some degree of homogeneity. All participants were between 18 and 34 years of age and had received at least some college education. This could cause absolute success rates or usability scores to be higher than in a broader population, though it should not affect comparisons among different versions of the application.

4.6 Demographics

Our participants were not balanced with respect to gender—50.0% (10) of our participants for the original Signal, 70.0% (14) of participants for Modification 1, and 35.0% (7) of participants for Modification 2 were male.

Since we distributed recruitment flyers on the university campus, most of our participants were undergrads, between 18 and 24—90.0% (18), 100.0% (20), and 90.0% (18) for each of the three versions. Most participants had some college but not yet earned a diploma—90.0% (18), 75.0% (15), and 65.0% (13) for the three versions.

Participants had a variety of backgrounds, skewed toward fields with non-technical backgrounds and less explicitly IT-related. Participants were asked to place themselves into categories of "beginner," "intermediate," and "advanced" regarding their security expertise. Most participants regarded themselves as beginners—85.0%(17), 85.0%(17), and 70.0% (14) for the three versions. None of our participants classified themselves as advanced, including the four participants from computer science or computer engineering.

5. RESULTS

In this section, we discuss the quantitative and qualitative results regarding the use of the authentication ceremony by participants. Details of our statistical methods are given in Appendix A.

5.1 Adherence and Completion

Participants who completed the ceremony compared their safety numbers by either scanning the QR code or by comparing the numbers over a phone call. We recorded a failure when participants transmitted sensitive data before verifying safety numbers, or if they failed to locate and validate safety numbers within ten minutes of launching the application. We also asked participants whether they felt they had safely exchanged the credit card number. Success and failure reports from both participants and the study coordinators are shown in Table 2.

Half of the participants who used the original Signal, and the majority of participants who used the modified versions, believed that they completed the task safely. However, none of the participants who used the original Signal version successfully performed the authentication ceremony. Only five participants even located the screen where safety numbers were displayed. In one of these cases, the participant ignored the instructions on the screen and simply pressed the *Mark as verified* button. In the other cases, participants ignored the screen entirely and immediately dismissed it. Participants tried several methods to deliver the message securely, including using various forms of primitive coding (e.g. developing their own substitution cipher), or enabling Signal's message impermanence feature.

	Participant self-report			*Study coordinator report*				
				Yes		**No**		
Application	**Yes**	**No**	**Not sure**	**QR code**	**Phone call**	**Not found**	**Ignored**	**Toggled**
Original	10	3	7	0	0	15	4	1
Modification 1	18	0	2	4	2	0	2	12
Modification 2	12	1	7	0	18	1	1	0

Table 2: Did the participants safely exchange the credit card number?

Application	**Time to locate authentication ceremony**	**Time to complete authentication**	**Time to complete the task**
Original	3.5	N/A	5
Modification 1	<1	7	8
Modification 2	<1	2	4

Table 3: Median time, in minutes, for finding and using the authentication ceremony.

All of the participants who used Modification 1 located the authentication ceremony screen, a large increase over the original Signal. However, while six participants correctly verified their safety numbers, the remaining 14 did not. Two of these latter participants ignored the screen and dismissed it, and the other 12 simply toggled the *Mark as verified* switch without comparing numbers. In successful cases, participants met to scan the QR code on each other's phone and in one case they wrote the safety numbers on paper and then made a phone call to verify them. We also notice that under this version of Signal, nearly all of the participants (18) believed they had safely performed the task. Only one of the participants who toggled the switch claimed to be unsure about the safety of the exchange.

Participant performance with Modification 2 was drastically better when compared to both the original Signal and Modification 1. Under Modification 2, 18 (90%) participants successfully performed the authentication ceremony, all of whom elected to do it over a phone call. The two failures were from the same pair of participants. In this case, Participant A erroneously informed his partner that the information had been transmitted safely, which caused Participant B to abandon his viewing of the authentication ceremony. However, Participant B did note that he was unsure the information was transferred safely in the post-task survey.

To test whether there are any differences between the versions of Signal, we used Cochran's Q test. We found that the success rate was statistically different for the applications ($\chi^2(2) = 27.11$, $p<.0005$). We then ran Barnard's exact test to find the significant differences among the pairs of applications. This test shows the differences among all the pairs are significant (Signal vs. Modification 1, $p = 0.0165$; Signal vs. Modification 2, $p = 1.15E-05$; Modification 1 vs. Modification 2, $p = 0.0163$).

5.2 Timing

The study coordinators timed each of participants and obtained three metrics, all with a granularity of minutes. First, the time required to locate the authentication ceremony was measured from the time that participants launch the application to the time where they first find the screen wherein the safety numbers reside. Second, the time for authentication completion was measured from the time users find the safety number screen to the time they verify their partner's safety number matches their own. Third, task completion time was measured from the time participants launch the application to the time they send (or receive) the credit card number from their partner.

Table 3 shows median times for each of the discussed metrics. For studies involving Modification 1, no one performed the authentication; thus we did not include this data in the table. In all of the studies with Modification 1 and Modification 2, all participants except for two discovered the authentication ceremony in less than 1 minute, with many taking just a few seconds. For the original version, only 5 (25%) of the participants found the screen, with a median of 3.5 minutes. Note the average discovery time in [20] was 3.2 minutes.

Participants correctly performed the authentication ceremony in 3 out of the 10 experiments with Modification 1, taking a median of 7 minutes. Participants correctly performed the authentication ceremony in 9 out of the 10 experiments with Modification 2, finishing in a median of 2 minutes. Note that the average time to complete the ceremony in [20] was 7.8 minutes.

For finding the ceremony, a two-tailed, two-sample t-test with equal variance shows there is no significant difference between Modification 1 and Modification 2 (p=0.484, 95% CI: [-0.37, 0.759] minutes). This is expected since the interfaces for finding the ceremony are identical in these two versions. For completing the authentication ceremony, a two-tailed, two-sample t-test with equal variance shows there is a significant difference between Modification 1 and Modification 2 (p=7.849E-05, 95% CI: [1.937, 6.16] minutes).

5.3 Usability

We asked participants who found the authentication ceremony to rank the usability of the ceremony on a five-point Likert scale, from *Extremely easy* to *Extremely difficult*. Table 4 shows the participant responses to this question. No one reported the task as extremely difficult and the majority of participants found it easy or somewhat easy to work with the authentication ceremony.

Note that the one who ranked the ceremony in the original Signal as *Extremely easy* to use simply toggled the *Mark as verified* switch. Of the nine participants using Modification 1 who reported it was extremely easy for them to use the ceremony, 5 either ignored it or toggled the *Mark as verified*

Application	Extremely easy	Somewhat easy	Neither easy nor difficult	Somewhat difficult	Extremely difficult
Original	1	0	2	1	0
Modification 1	9	5	5	1	0
Modification 2	7	10	1	1	0

Table 4: Responses to: "How difficult or easy was it to use this screen to verify the safety number?"

switch, with the rest successfully completing the ceremony. All of the participants using Modification 2 saw the authentication ceremony screen and the majority believed it was easy to use. Because many of the participants either didn't use the ceremony or didn't complete it properly, we didn't run any statistical comparisons among the different versions.

We also asked these same participants what they liked or disliked about verifying their safety number, in an open-response question. Interestingly, some users felt the length of the safety numbers improved the security of the task, while others felt they were too long or hard to keep track of. This is well illustrated by the comment from one participant who used Modification 2:

> *"I liked that it came up on the middle of the phone call screen rather than being sent through a text message that I would have to pull up during the conversation. There were a lot of numbers, which could be hard to keep track of if you were reading them over the phone, but the amount of numbers ensures greater safety."*

The confusion regarding the original authentication ceremony is well illustrated by this comment from a participant who used Modification 1:

> *"I was a little confused at first and I wondered if we needed to be in the same room to scan the QR code to make sure our conversation was secure. At the bottom it just asked if I could switch the conversation to verified and so I did."*

Another participant who used Modification 2 stated: *"I liked how the numbers were large and visible but I didn't like that the numbers had to be read on speaker phone so everyone could have heard them.* This indicated some confusion about the role that safety numbers play in securing the conversation.

Note that we didn't make any statistical comparisons for this or other qualitative data in the paper. Our qualitative data is noisy, meaning some users may not have offered all their reasoning in a particular answer, while other users gave multiple reasons and were coded into multiple categories. In addition, because of a large number of categories, the values of many cells in the tables are small. These factors make statistical comparisons problematic.

5.4 Comprehension

During the post-task survey we also asked participants who found the authentication ceremony what they thought the screen did. Overall, 43 out of 60 participants answered this question. We also showed the screen to participants during the interview portion of the study, asking them how they thought the screen helped them with the task. We coded this data, with the results shown in Table 5.

Code	Original	M1	M2
(A) Survey			
Authentication	3	4	6
Confidentiality	2	2	3
Security	0	6	7
Trust	0	2	1
Didn't know	0	7	5
(B) Interview			
Authentication	7	7	6
Confidentiality	3	6	7
Security	2	1	2
Trust	2	0	0
Didn't know	5	7	5

Table 5: Coded responses to: (A) "What does this screen do?" (shown if they saw the ceremony during the study), (B) "How does this screen help you to accomplish the task?"

Many participants believed the authentication ceremony was involved with either authentication or confidentiality. Typically when mentioning authentication they discussed making sure they were talking to the right person and not an impostor. Some participants indicated a good level of understanding. For example, one participant who used Modification 2 said:

> *"That made sure that you weren't talking to someone pretending to be your friend or someone who had hacked her number and was answering the phone for her. Because there was not picture of her, no live stream video. So it could have been someone that sounded like her really closely. So I think that's what the numbers did...If numbers didn't match.It would mean that I would send him a message, and then his phone would try to unencrypt it, and it would just get garbage information."*

A participant who scanned QR codes with Modification 1 said:

> *"I think it helps, that, there were so many of them that its hard to replicate so I don't think that it would be easy for someone to just steal them or come up with them and so, cause there were enough of them that when I saw that (A) had all the same ones I was like 'Cool I am definitely talking to the person I think I am.' "*

When mentioning confidentiality, participants discussed making sure nobody else could read their conversation. A participant who used Modification 2 said: *"I was thinking for safety reasons. To make sure that the information we're telling to each other is just between the two of us."*

Participants also often mentioned security, generally, without any additional clarification about what it meant to have a

Code	Original	M1	M2
Positive impressions			
Use of primitive cipher	6	3	0
Trust in the application	4	9	1
Message impermanence	3	3	0
Use of other security features	2	0	0
Successful message delivery	1	0	0
Contact list synchronization	1	0	1
Trust voice call	1	3	2
Absence of physical threats	1	2	1
Authentication ceremony	0	7	11
Negative impressions			
Lack trust in the application	7	1	1
Lack of knowledge	4	2	4
Time cutoff reached	1	0	0
Lack of transparency	1	0	1
Lack of trust in mobile apps	0	1	0
Lack of trust in text	0	0	1
Possible physical threat	0	0	1

Table 6: Coded responses to: "Do you think you have safely exchanged the credit card number with your friend? Explain your answer."

"secure connection". There are also significant numbers who didn't know and, by their own admission, could not make a guess, or who were clearly making up an answer on the fly.

Note that many participants, across all codings, expressed doubt about their answers, as is typical in lab studies involving technical topics. The vast majority of participants were not entirely sure about the role that safety numbers played.

5.5 Participant Report on Success or Failure

During the post-study survey, participants were asked: "*Do you think you have safely exchanged the credit card number with your friend? Explain your answer.*" The available discrete responses were *Yes*, *No*, and *Not sure* (reported previously in Table 2), and an adjacent free response field required respondents to explain in their own words. We coded the free response portion of these answers into two groups: positive impressions and negative impressions. Positive impressions were used to support claims of success and negative impressions support claims of possible failure. Note that these categories are not mutually exclusive. For example, a persons unsure of their success in the task sometimes provided both positive and negative impressions. The number of responses in each identified category across all variants of the Signal application tested are shown in Table 6.

The use of a primitive cipher, such as writing the credit card number backwards, sending a screenshot instead of textual data, or mapping numbers to letters in the recipients name, was the most popular positive impressions for tasks under the original Signal. This was followed by trust in the application and the use of message impermanence settings. We see an increase in mentions (from 0 to 7) of the authentication ceremony from study participants who used Modification 1, which is then further increased by participants using Modification 2, with over half of participants mentioning this in their response. However, we also note that lack of knowledge seemed relatively unaffected by Modification 2 with respect to the original Signal.

In cases of failure to find the authentication ceremony or perform it correctly, participants were asked: "*What were you looking for to accomplish the task?*" By this time we had already showed them the authentication ceremony screen and its purpose, so this question allowed them to provide us with insight to what information they lacked during the study that would have helped them find and use the ceremony.

For users of the original Signal, this responses were largely aimed at explaining why they did not locate the ceremony. These reasons varied wildly, which in itself became the overall theme: participants lacked sufficient direction under the original Signal. One participant said *"I had no idea what I needed to do"* and another said he was *"just looking for any sort of security setting or application."* Some explained their method for ad-hoc cipher use, implying that they didn't look for built-in functionality to provide safety and instead resorted to their own means. Others explained that they got caught up experimenting with other security features of the application, such as message destruction, or blamed their own laziness for not finding the ceremony.

Since modified Signal versions effectively led the participants to the ceremony screen, responses provided insight into why the authentication ceremony may not have been performed properly. The primary difficulty in using Modification 2 was that participants had difficulty knowing what to do with the authentication ceremony screen and its "Mark as verified" button. For example, one participant remarked,

> *"I hit [the button] and then I was like, 'well that did nothing' and so I hit it again and nothing happened...I hit verify and then it says that I just unhit it immediately afterwards...I was just like, 'verify what? What am I verifying?' It didn't really tell me...Honestly it meant nothing."*

Our second modification was designed to deal with these problems by guiding users through the authentication ceremony. Only one pair was unsuccessful in properly performing the ceremony. The response to this question from that pair explained that the participants felt comfortable once they identified each other on the call and thus didn't proceed.

5.6 Trust

Participants were presented with the statement *I trust that Signal is secure*, and asked to rank their agreement with the statement on a five-point Likert scale ranging from *Strongly agree* to *Strongly disagree*. Table 7 shows responses to this question for the different versions of Signal. A one-way ANOVA shows that there are no significant differences between the different versions (p=0.143). We also asked participants to explain their answer in an open response question. We coded their positive and negative impressions, and this data is shown in Table 8.

The majority of participants somewhat or strongly agreed with the statement, with more users of Modification 1 and 2 expressing these sentiments as compared to the original version. Note that many of the people expressing trust in the application had no specific reason other than that the application seemed secure, or that it seemed more secure than other applications they had used. A number of people pointed to the authentication ceremony as a reason to trust the application, but this could be because they sensed this

Application	Strongly agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Strongly disagree
Original	4	9	6	1	0
Modification 1	5	13	2	0	0
Modification 2	6	12	2	0	0

Table 7: Responses to: "I trust that Signal is secure."

Code	Original	M1	M2
Positive impressions			
Seem secure	6	3	7
Relatively secure	1	2	0
Authentication ceremony	2	4	6
Security settings	1	1	0
No evidence to contrary	2	0	2
Trust university	0	3	0
Message impermanence	2	0	0
Few people using it so far	0	0	1
User interface	0	1	0
Negative impressions			
Lack of knowledge/experience	9	7	8
Lack of reputation	1	1	4
Lack of transparency	1	1	0
Lack of trust	1	1	0
Confusing user interface	0	2	0

Table 8: Coded responses to: "Please explain your answer" (regarding whether they trust Signal to be secure)

was the purpose of the study. One participant who used Modification 2 stated: *"I think this app is strongly agree because once you are verified with others you can actually trust the person on the call and exchange your information."* Several participants indicated they trusted the application because they assumed it had been made by developers at our university.

The only person who chose *Somewhat disagree*, for the original version of Signal, couldn't find the authentication ceremony and referred to lack of transparency as the reason for this choice. This participant said: *"There is no proof of this at all. It says it is secure but does not give me any information."* The main negative impressions expressed by participants were lack of lack of knowledge about the application or experience using it, and lack of reputation.

5.7 Adoption

During the interview portion of the study, we read participants the following statement:

> *"It is possible for someone to intercept your messages. These screens we have been showing you are called an authentication ceremony. Using the authentication ceremony ensures that nobody, not Signal, not hackers, and not even the government, is able to intercept your messages. You only need to do this once (or if your friend reinstalls the app). Now that you know this, are you willing to use the authentication ceremony before you exchange messages with a friend the first time?"*

We then asked participants if they would be willing to use the authentication ceremony in secure messaging applications in the future. Of the participants who answered this question, 32 said yes, 4 said no, 14 said only if they were exchanging confidential information, and 6 said only with certain contacts. We emphasize that participants were likely to say yes to this question, due to the nature of the study.

As an example of how we rated someone who would use the ceremony only when sending certain content, one participant who used Modification 1 said:

> *"Am I willing? Yes. Will I? No. Because here is the thing, I don't really care if my messages get intercepted because most of the time I am not sending my credit card number or social security numbers. Will I use it for things that are really important? For sure."*

6. DISCUSSION

In this section we discuss the significance and shortcomings of our results.

6.1 Adherence, Timing, and Comprehension

One of the primary contributions of this work is that the modifications that we made to Signal result in a higher success rate and lower task completion time in comparison to the original version. With Modification 1 and Modification 2 combined, 97.5% of participants found the ceremony, compared to only 25% for the original version of Signal. In addition, the changes made to the authentication ceremony in Modification 2 resulted in a success rate of 90% for completion of the ceremony, as compared to 30% for Modification 1. Numerous participants were confused by the *Mark as Verified* toggle in the ceremony for Modification 1 (the same as Signal's current ceremony), and assumed that flipping this switch would activate some kind of automatic verification.

Our results improve on prior work by Vaziripour et al. [20]. Participants found the authentication ceremony in an average of less than a minute (and often seconds), as compared to 3.5 minutes for [20]. Likewise, the average time to complete the authentication ceremony was 2.11 minutes, as compared to 7.8 minutes across the three applications [20] studied.

These advances were made using opinionated design to encourage participants to use the authentication ceremony, combined with a small monetary incentive to be security-minded. Our methodology included no instruction on finding or completing the ceremony, as in prior work [8, 20]. This indicates that the interface changes were enough to lead to the desired behavior, once participants had a security mindset.

Despite these results, participants did not demonstrate a strong comprehension of the purpose of the authentication ceremony. Although some participants believed the ceremony had something to do with authentication or confidentiality, many expressed doubts about their opinions. Still others either directly or indirectly admitted they didn't know what

it was for. As one participant who used Modification 2 stated, *"I don't know. I'm not really sure, actually, how it helped."*

Overall, these results are similar to a recent Google study on SSL warnings [6]. This study found that design of the warnings enhanced secure behavior from users and boosted threat understanding, but did not necessarily improve user comprehension of the warnings. This indicates that more work is needed to help users understand what they are doing in the authentication ceremony, and why they are doing it.

6.2 Adoption, Risk Communication, and Trust

Our interviews with participants indicate more work is needed within secure messaging applications to explain the purpose of the ceremony and to help users make choices about when it is necessary. Once the purpose of the authentication ceremony was explained to users, they readily understood it. However, a third of participants indicated that they would only want to use it certain in cases when they were sending sensitive information, and their responses indicated that they viewed the risk as acceptable when sending ordinary information. Others indicated they would never see the need, or said they would have trouble convincing their contacts to adopt secure messaging apps or use the ceremony.

A review of terminology used in Signal and in our modifications illustrates the difficulty. Our warning message to users reads "*Action needed! Click to verify your safety numbers.*" There is no indication of what comparing these numbers will do for users, nor what risks occur if they don't. Likewise, in the current Signal ceremony, it tells users that:

> *"If you wish to verify the security of your end-to-end encryption with Bob, compare the numbers above with the numbers on their device."*

Many users may not know what end-to-end encryption is, why comparing these numbers helps, nor what risks occur if they do not do this. Similar criticisms are valid for our modified ceremony.

In addition, users make rational tradeoffs between security and convenience [7]. Even if the ceremony is highly usable, users may still not adopt it, since usability is not the primary obstacle to adoption of secure messaging applications [2]. Rather, users may perceive the ceremony as "geeky" [9], they may not be convinced there is a need for it, or they may not be able to convince their contacts to use it.

Finally, many users readily admitted that they lacked the knowledge and experience necessary to know whether to trust Signal. The difficulty this poses for users was expressed well by one participant who used Modification 2:

> *"I don't know that there is anything that would make be sure that no one else is listening in. I don't know if whoever has developed Signal has someone set it up so that they can listen in. I would assume that they don't because it seems like their purpose is security. But I guess it might be possible for someone to be listening in. I don't know how I would know that that isn't happening."*

It's not clear how to give users a sense of trust in secure applications, especially when there are regular breaches of security that they hear about in the news.

6.3 Generality

Our results on finding and using the authentication ceremony should generalize to other secure messaging applications. We examined several major messaging applications to identify how our research would apply to them.

- *Finding the Authentication Ceremony:* WhatsApp, Telegram, Facebook Messenger, and Viber all require multiple clicks to find the authentication ceremony within the menu system, similar to Signal. With both Telegram and Facebook Messenger, encrypted chats are optional, so additional steps are needed to initiate a secure chat. We expect our improvements for finding the ceremony would be applicable to all of these applications.
- *Using the Authentication Ceremony:* The ceremonies in WhatsApp, Telegram, and Facebook Messenger differ in varying degrees from Signal. WhatsApp is nearly identical, with options for scanning a QR code or comparing an alphanumeric fingerprint, and no integrated phone call. Telegram allows the user to compare either a graphical or alphanumeric fingerprint, with no integrated QR scanning or phone call and few instructions. Facebook Messenger only offers the option to compare an alphanumeric fingerprint, and there are separate keys for each device, again with no integrated phone call. We expect our improvements for using the ceremony will be applicable to all of these applications. Viber is unique in that it integrates a phone call into their application to make the ceremony easier to use. Thus it is likely that Viber's ceremony would have similar success as our design. In prior work [20] Viber had the highest success rate for the authentication ceremony once people were directed to find it.

7. CONCLUSION

Our study indicates that users can find and complete the authentication ceremony in secure messaging applications, provided they have a security mindset and the application is designed to help them easily accomplish these tasks. This raises numerous open questions for further study. First, comprehension is still somewhat low, and additional design is needed to help users understand why they should perform the ceremony and when it is necessary. Second, it is not clear whether users will be security-minded without encouragement, such as a small monetary reward in the case of our study. More work is needed to determine if user interface changes alone can encourage use of the ceremony. Third, work is needed to determine if these advances can be applied to helping users cope with an attack scenario or when a contact re-installs Signal. Both of these situations will cause the security numbers to change, alerting users to a possible attack, and evidence to date shows that users do not cope well. Fourth, it may be possible to fully automate the authentication ceremony, using social authentication [19] or CONIKS [11]. Finally, work is needed to help users make good choices about which secure messaging applications are safe to use.

8. ACKNOWLEDGMENTS

The authors thank the anonymous reviewers and our shepherd, Apu Kapadia, for their helpful feedback. This material is based upon work supported by the National Science Foundation under Grant No. CNS-1528022.

9. REFERENCES

[1] R. Abu-Salma, K. Krol, S. Parkin, V. Koh, K. Kwan, J. Mahboob, Z. Traboulsi, and M. A. Sasse. The security blanket of the chat world: An analytic evaluation and a user study of telegram. In *European Workshop on Usable Security (EuroUSEC)*. Internet Society, 2017.

[2] R. Abu-Salma, M. A. Sasse, J. Bonneau, A. Danilova, A. Naiakshina, and M. Smith. Obstacles to the adoption of secure communication tools. In *IEEE Symposium on Security and Privacy (SP 2017)*, pages 137–153. IEEE, 2017.

[3] A. Acquisti, I. Adjerid, R. Balebako, L. Brandimarte, L. F. Cranor, S. Komanduri, P. G. Leon, N. Sadeh, F. Schaub, M. Sleeper, et al. Nudges for privacy and security: Understanding and assisting users' choices online. *ACM Computing Surveys (CSUR)*, 50(3):44, 2017.

[4] H. Assal, S. Hurtado, A. Imran, and S. Chiasson. What's the deal with privacy apps?: A comprehensive exploration of user perception and usability. In *International Conference on Mobile and Ubiquitous Multimedia (MUM)*. ACM, 2015.

[5] S. Dechand, D. Schürmann, T. IBR, K. Busse, Y. Acar, S. Fahl, and M. Smith. An empirical study of textual key-fingerprint representations. In *USENIX Security Symposium*. USENIX Association, 2016.

[6] A. P. Felt, A. Ainslie, R. W. Reeder, S. Consolvo, S. Thyagaraja, A. Bettes, H. Harris, and J. Grimes. Improving SSL warnings: Comprehension and adherence. In *Conference on Human Factors in Computing Systems (CHI)*, pages 2893–2902. ACM, 2015.

[7] C. Herley. So long, and no thanks for the externalities: the rational rejection of security advice by users. In *New Security Paradigms Workshop*, pages 133–144. ACM, 2009.

[8] A. Herzberg and H. Leibowitz. Can Johnny finally encrypt? Evaluating E2E-encryption in popular IM applications. In *Workshop on Socio-Technical Aspects in Security and Trust (STAST)*, Los Angeles, California, USA, 2016.

[9] A. Kapadia. A case (study) for usability in secure email communication. *IEEE Security & Privacy*, 5(2), 2007.

[10] S. Krug. *Don't make me think!: a common sense approach to web usability*. Pearson Education India, 2000.

[11] M. S. Melara, A. Blankstein, J. Bonneau, E. W. Felten, and M. J. Freedman. CONIKS: Bringing key transparency to end users. In *USENIX Security Symposium*, pages 383–398. USENIX Association, 2015.

[12] T. Perrin and M. Marlinspike. The double ratchet algorithm. `https://signal.org/docs/specifications/doubleratchet/doubleratchet.pdf`, 2016.

[13] A. Sasse. Scaring and bullying people into security won't work. *IEEE Symposium on Security and Privacy (S&P)*, 13(3):80–83, 2015.

[14] S. Schröder, M. Huber, D. Wind, and C. 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.

[15] M. Shirvanian, N. Saxena, and J. J. George. On the pitfalls of end-to-end encrypted communications: A study of remote key-fingerprint verification. In *Annual Computer Security Applications Conference (ACSAC)*, pages 499–511. ACM, 2017.

[16] O. W. Systems. Signal. `https://signal.org/`. Accessed: 2018-02-10.

[17] J. Tan, L. Bauer, J. Bonneau, L. F. Cranor, J. Thomas, and B. Ur. Can unicorns help users compare crypto key fingerprints? In *Conference on Human Factors and Computing Systems (CHI)*, pages 3787–3798. ACM, 2017.

[18] N. Unger, S. Dechand, J. Bonneau, S. Fahl, H. Perl, I. Goldberg, and M. Smith. SoK: secure messaging. In *IEEE Symposium on Security and Privacy (S&P)*, pages 232–249. IEEE, 2015.

[19] E. Vaziripour, M. O'Neill, J. Wu, S. Heidbrink, K. Seamons, and D. Zappala. Social authentication for end-to-end encryption. In *Who Are You?! Adventures in Authentication Workshop (WAY)*. USENIX Association, 2016.

[20] E. Vaziripour, J. Wu, M. O'Neill, R. Clinton, J. Whitehead, S. Heidbrink, K. Seamons, and D. Zappala. Is that you, Alice? a usability study of the authentication ceremony of secure messaging applications. In *Symposium on Usable Privacy and Security (SOUPS)*, 2017.

APPENDIX

A. STATISTICAL TESTS

This section contains the details of the statistical tests we ran.

A.1 Sample Size

We calculated the necessary sample size to compare two sample proportions (for comparing success rates) and two sample means (for comparing task times). With a 95% confidence interval, 80% power, and an expected success rate for the two samples (15% and 80%, based on our previous work [20]), the required sample size is 6. With a 95% confidence interval and 80% power, the hypothesized difference in timing completing the ceremony (4 minutes), and our previous measurements of variance for the task (9 minutes), the required sample size is 9. We rounded up to 10.

A.2 Success and Failure Rates

This data measures whether the participants were successful in using the authentication ceremony for the original Signal and each of the modifications. We want to test whether there are any differences in the success rate between the three versions of the Signal application.

Because the data is dichotomous we used Cochran's Q Test and found that the success rate was statistically different for the applications ($\chi^2(2) = 27.11$, $p<.0005$).

Since we used a between-subject study design, we performed Barnard's exact test to find the significant differences among the pairs of applications. This test shows the differences among all the pairs are significant (Signal vs. Modification 1, $p = 0.0165$; Signal vs. Modification 2, $p = 1.15E-05$; Modification 1 vs. Modification 2, $p = 0.0163$).

A.3 Task Completion Times

This data measures the time taken by participants to (a) find the authentication ceremony and (b) complete the authentication ceremony. We want to test whether there are any differences between the three versions of the Signal application, in finding and task completion time.

We did not perform a multiple samples comparison test because of the high failure rate with the original version of Signal. Since the studies are between subject, we ran a two-tailed two-sample t-test between Modification 1 and Modification 2.

For finding the authentication ceremony, a two-tailed, two-sample t-test with equal variance shows there is no significant difference between Modification 1 and Modification 2 (p=0.484, 95% CI: [-0.37, 0.759] minutes). This is expected since the interfaces for finding the ceremony are identical in these two versions. For completing the authentication ceremony, a two-tailed, two-sample t-test with equal variance shows there is a significant difference between Modification 1 and Modification 2 (p=7.849E-05, 95% CI: [1.937, 6.16] minutes). For the total time to find and complete the ceremony, a two-tailed, two-sample t-test with equal variance shows there is a significant difference between Modification 1 and Modification 2 (p=1.05E-05, 95% CI: [2.982, 6.518] minutes).

A.4 Trust Scores

A one-way ANOVA shows that there are no significant differences between the different versions (p=0.143).

B. STUDY MATERIALS

This section contains the study materials we used. The study coordinators used the interview guide to ensure that each pair of participants experienced an identical study. The study participants used the questionnaire to guide them through the study.

B.1 Interview Guide

Make sure to complete the following steps:

- When the two users arrive, read them the following:

  Welcome to our secure messaging application study. We are the study coordinators and are here to assist you as needed.

  Before we start the study, we need you to let us install an application called Signal on your phone. You will use this application during the study, and then we will delete it for you when we are done.

- Install the Signal application on their phone.
- Now read the following:

  In this study, the two of you will be in different rooms and will use the Signal app to communicate with each other.

  You will be asked to *think aloud* during the study. This means that you should explain everything you are thinking and feeling during the study so we can understand how you interact with the Signal application.

  During the course of this study we will be making an audio recording of what you say. We will transcribe these recordings and may publish them as part of our study, but we will not identify you in any way. We will destroy the audio recordings and will publish only transcripts so that you will be anonymous. We will not collect any personally identifying information about you.

  You will also take a survey during the study, and we will publish your answers, but without any information that can identify you.

  You will each receive $7 cash as compensation for your participation in this study. You will also have an opportunity during the study to earn a bonus of $3 cash, based on your performance. The expected time commitment is approximately 30 minutes.

  If you have any questions or concerns, feel free to ask us. You can end participation in this survey at any time and we will delete all data collected at your request. A study coordinator will be with you at all times to observe the study and also to answer any questions you may have.

- Before going to the study rooms, make sure the participants sign the audio recording consent form.
- Flip a coin and choose one participant to be Person A and one person to be Person B. Take the participant with whom you will work to the study room. Ask the participant to sit down.
- Start the audio recording using the equipment in the study room.
- Read the following instructions to your participant:

  We are going to ask you to do a series of tasks. Once you are done with each step, let the study coordinator know you have finished the task. You will then fill out a questionnaire and go to the next step. We need you to think out loud while you are doing the tasks in this study, meaning you are supposed to talk about how you are accomplishing the task and express any feelings you have. If you have any questions about the study ask the study coordinator. Remember you are allowed to talk to or meet your friend during the study.

  Please do not forget to think out loud.

- On the chromebook, load the survey from Qualtrics.
- Before using Signal, the survey will instruct the participant to tell you they are ready to begin the next task.

  During the course of the task pay attention to what user is doing and fill out one of the attached sheets.

  The user is supposed to think aloud while doing the tasks. If she forgets, gently remind her.

  Do not answer any questions from the participants.

  The participants have 10 minutes to complete the primary task, which is using Signal to exchange credit card information. If they do not finish the task on time, guide them to the next part of the survey. If you end

the task, inform the other study coordinator that you have done so, so that he catches up with you.

If it takes the pair too long to complete authentication or if they sent a credit card number before performing the authentication, then record that as a failure.

- When the survey is finished, ask the participant about their experience.

  Use the situations you noted while they took the study or interesting things they said on the survey. If they had any problems during the study, ask them to use their own words to describe the problem. Ask them how they would like to see it resolved.

- When the participant is finished, ask his/her opinion on the following questions:
  - Ask user if they trust the voice or text messaging for secure conversation?
  - If they did not use the authentication ceremony:
    - Ask them what they were looking for.
    - Show them how to find the application ceremony. Why did they not find it?
    - How do you think this screen would have helped you accomplish the task?
  - If they did use the authentication ceremony, show them the screen(s).
    - How do you think this screen helped you accomplish the task?
  - Explain the following:
    It is possible for someone to intercept your messages. These screens we have been showing you are called an authentication ceremony. Using the authentication ceremony ensures that nobody, not Signal, not hackers, and not even the government, is able to intercept your messages. You only need to do this once (or if your friend reinstalls the app). Now that you know this, are you willing to use the authentication ceremony before you exchange messages with a friend the first time?
- Stop the audio recording.
- Return to the study room. Thank the participants for their time. Ask them not to invite their friends to participate. Help them fill out the compensation forms and give them compensation.

B.2 Study Questionnaire

Signal study

1. Please enter whether you are Participant A or B.
   - A
   - B
2. What is your gender?
   - Male
   - Female
   - I prefer not to answer
3. What is your age?
   - 18-24
   - 25-34
   - 35-45
   - 46-64
   - 65 and over
   - I prefer not to answer
4. What is the highest degree or level of schooling you have completed?
   - None
   - Primary/grade school
   - Some high school, no diploma
   - High school graduate: diploma or equivalent (e.g., GED)
   - Some college, no diploma
   - Associate's or technical degree
   - Bachelor's degree
   - Graduate/professional degree
   - I prefer not to answer
5. What is your major, or if employed, your occupation?
6. Tell the study coordinator that you are ready for the next task to begin.
7. **For Person A**

   You left your credit card at home! You are going to be using the Signal app to ask your friend to send you the credit card number.

   This is the message you should send to your friend:

   "Hi! Can you send me my credit card number? I left my card on my desk at home."

   You can both earn a bonus of $3 for this study if you make sure that nobody can steal this information while your friend is sending it.

   Talk out loud as you do this task.

   For Person B

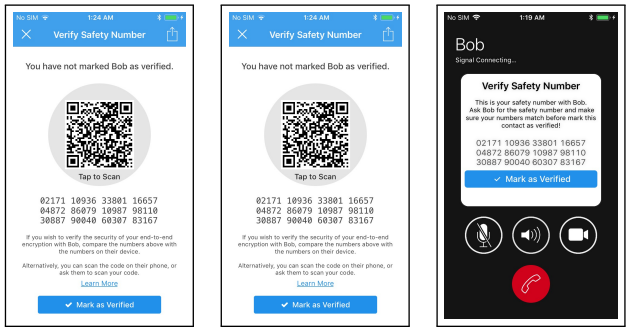
   Your friend is going to use the Signal app to ask you for their credit card number. Use the credit card given to you by the study coordinator.

   You can both earn a bonus of $3 for this study if you make sure that nobody can steal this information while you're sending it.

   Talk out loud as you do this task.
8. You will now be asked several questions concerning your experience with Signal.
9. Do you think you have safely exchanged the credit card number with your friend?
   - No
   - Yes
   - Not sure
10. Please explain your answer:
11. Did you see this screen during the study?
    (showed Figure 5)
    - No
    - Yes

    If (Yes), ask the following three questions.
12. What do you think this screen does?
13. Overall, how difficult or easy was it to use this screen to verify the safety number?

    (Extremely easy to extremely difficult)



(a) Original Signal (b) Modification 1 (c) Modification 2

Figure 5: Authentication ceremony screen

14. When you used this screen during the study to verify the safety number, what did you like or dislike about this? Please explain why.
(showed Figure 5)
15. Before this study, have you ever tried to send sensitive information when you use a secure messaging application like Signal?
    - Yes
    - No
16. (If Yes), Explain what kind of sensitive information you have sent.
17. I trust that Signal is secure.
    - Strongly agree
    - Somewhat agree
    - Neither agree nor disagree
    - Somewhat disagree
    - Strongly disagree
18. Please explain your answer to the above question.
19. How would you rate your knowledge of computer security?
    - Beginner
    - Intermediate
    - Advanced
20. Which of the following applications have you ever used? Select as many options that applies to you.
    - ☐ WhatsApp
    - ☐ Signal
    - ☐ Telegram
    - ☐ Line
    - ☐ Allo
    - ☐ Facebook Messenger
    - ☐ iMessage
    - ☐ Skype
    - ☐ Viber
    - ☐ Other