On the Limited Impact of Visualizing Encryption: Perceptions of E2E Messaging Security

Christian Stransky, *Leibniz University Hannover;* Dominik Wermke, *CISPA Helmholtz Center for Information Security;* Johanna Schrader, *Leibniz University Hannover;* Nicolas Huaman, *CISPA Helmholtz Center for Information Security;* Yasemin Acar, *Max Planck Institute for Security and Privacy;* Anna Lena Fehlhaber, *Leibniz University Hannover;* Miranda Wei, *University of Washington;* Blase Ur, *University of Chicago;* Sascha Fahl, *Leibniz University Hannover and CISPA Helmholtz Center for Information Security*

https://www.usenix.org/conference/soups2021/presentation/stransky

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

August 9–10, 2021

978-1-939133-25-0

Open access to the Proceedings of the Seventeenth Symposium on Usable Privacy and Security is sponsored by

On the Limited Impact of Visualizing Encryption: Perceptions of E2E Messaging Security

Christian Stransky[†], Dominik Wermke[C], Johanna Schrader[†], Nicolas Huaman[C], Yasemin Acar[‡], Anna Lena Fehlhaber[†], Miranda Wei[*], Blase Ur[◇], Sascha Fahl[†C]
[†] *Leibniz University Hannover;* [C] *CISPA Helmholtz Center for Information Security;*
[‡] *Max Planck Institute for Security and Privacy;* [*] *University of Washington;* [◇] *University of Chicago*

Abstract

Communication tools with end-to-end (E2E) encryption help users maintain their privacy. Although messengers like WhatsApp and Signal bring E2E encryption to a broad audience, past work has documented misconceptions of their security and privacy properties. Through a series of five online studies with 683 total participants, we investigated whether making an app's E2E encryption more visible improves perceptions of trust, security, and privacy. We first investigated why participants use particular messaging tools, validating a prior finding that many users mistakenly think SMS and e-mail are more secure than E2E-encrypted messengers. We then studied the effect of making E2E encryption more visible in a messaging app. We compared six different text disclosures, three different icons, and three different animations of the encryption process. We found that simple text disclosures that messages are "encrypted" are sufficient. Surprisingly, the icons negatively impacted perceptions. While qualitative responses to the animations showed they successfully conveyed and emphasized "security" and "encryption," the animations did not significantly impact participants' quantitative perceptions of the overall trustworthiness, security, and privacy of E2E-encrypted messaging. We confirmed and unpacked this result through a validation study, finding that user perceptions depend more on preconceived expectations and an app's reputation than visualizations of security mechanisms.

1 Introduction

The use of E2E-encrypted communication tools for e-mail (e.g., PGP [70], S/MIME [52]) or for mobile apps (e.g., WhatsApp [66], iMessage [6], Signal [56]) is an effective countermeasure against cybercriminals, nation-state attackers, and other adversaries [36]. Most E2E-encrypted communication tools provide confidentiality, integrity, authenticity, and perfect forward secrecy [20] for message contents, but do not hide metadata like sender/receiver identities or when the message was sent [46]. Many previous studies have documented usability and adoption challenges for encryption tools [8, 12, 15, 62], especially for e-mail encryption [25, 26, 47, 49] and modern E2E-encrypted messaging apps [3, 4].

Of particular concern is that users often have flawed mental models of E2E-encrypted tools' security and privacy properties. This can lead users to mistakenly use less secure alternatives like SMS or e-mail for confidential conversations even when they already have access to E2E encryption through widely used tools like WhatsApp and iMessage [3].

Recent work has highlighted how increasing the visibility of typically invisible security mechanisms can improve user perceptions of trust and security. For example, a qualitative study on e-voting found that displaying security mechanisms improved both user experience and need fulfillment [18]. In the context of E2E encryption on Facebook, another study's qualitative results suggested that visibly transforming Facebook messages to and from ciphertext (an implementation artifact in that work) appeared to increase user trust and perceptions of security [21]. For e-mail security, studies found that clearly labeling PGP-encrypted e-mail differently from unencrypted e-mail improved usability and perceived security, as well as reduced unintentional human error when interacting with PGP-encrypted e-mails [48, 50]. Our work tests these promising results in the space of mobile messaging apps. In an attempt to improve user comprehension and perceptions of security, privacy, and trust for E2E-encrypted mobile messaging apps, we thus investigated visualizing encryption through various text descriptions, icons, and animations of the encryption process.

We conducted a series of five user studies on MTurk and Prolific to investigate the following three research questions:

USENIX Symposium on Usable Privacy and Security (SOUPS) 2021.
August 8–10, 2021, Virtual Conference.

RQ 1: Which messaging tools do people prefer for confidential communications, and why?

Of participants who had an E2E-encrypted tool installed (80), 62.50% reported they would use a tool without E2E encryption for confidential conversations, echoing prior work [3]. This finding suggests that E2E-encrypted communication tools can do more to discourage users from switching to less-secure tools in situations when security and privacy matter. Our root-cause analysis revealed factors like specific UI features, trust in companies, and security misconceptions contributed to participants' decisions.

RQ 2a: How does visualizing encryption through text, icons, or animations impact perceptions of E2E-encrypted messaging tools' security, trust, and privacy?

RQ 2b: What external factors and expectations mediate encryption visualizations' impact on user perceptions?

In an attempt to highlight tools' E2E encryption, we investigated three types of visualizations: *text disclosures*, *icons*, and *animations*. In our remaining online studies, we investigated the impact of different variants of these visualizations. While some of these disclosures have been investigated previously, the animations are especially novel, as is our application of a consistent human-subjects protocol to study all three types.

We found that perceptions of a tool's security, trust, and privacy increased as soon as there was a simple indicator of encryption, such as a text statement that messages are encrypted (similar to WhatsApp's current interface). Contradicting the recent literature, additional emphasis did not appear to have much impact. More concretely, heavyweight animations and icons did not appear to provide much benefit beyond a lightweight text disclosure in emphasizing E2E-encrypted messengers' security properties to users. While qualitative data suggested that rich visualizations like animations successfully emphasized security and encryption, they did not significantly impact quantitative measures of user perception. Notably, much of the recent literature relies on qualitative observations, whereas our dual use of both perspectives highlights limitations of visualizing security. Through a final study with additional questions, we validated the surprising lack of a quantitative effect and further unpacked users' expectations.

In this paper, we make the following contributions:

- We detail which E2E-encrypted communication tools participants use in different situations, and why.
- We investigate how visualizing encryption through text disclosures, icons, and animations impacts perceptions of security, privacy, and trust.
- We validate our findings and unpack the limitations of visualizing E2E encryption.

The rest of the paper is structured as follows. Section 2 presents previous work relevant to this paper and illustrates the novelty of our research. Section 3 provides detailed information on our methodology, including data quality and data analysis techniques we applied, as well as the ethical considerations and limitations of our work. In Section 4, we discuss the procedure and findings of our first study on the use of communication tools. Section 5 gives a detailed overview of the experiments we conducted on different visualizations of encryption, and Section 6 describes a validation study. Section 7 discusses our results, highlights their implications for secure messaging applications, and outlines possible future work. Finally, we conclude in Section 9.

2 Related Work

We discuss related work on encrypted communication tools' usability, adoption, and perception, and previous attempts to visualize security, especially encryption.

Usability of E2E Encryption The usability of E2E encryption has been a research focus since at least 1999, when Whitten and Tygar evaluated PGP with cognitive walkthroughs in a landmark paper [67]. One-third of participants failed to sign and encrypt an e-mail message within 90 minutes.

More recent work observes similar barriers. In two-person lab sessions, Ruoti et al. examined initial user experiences for three secure e-mail systems (Pwm, Tutanota, Virtru) through role-play scenarios with 50 participants. They found that participants were interested in secure e-mail in the abstract, but were unsure when and how they actually would use it. Only a few participants desired to use secure e-mail regularly [47]. De Luca et al. conducted online studies and interviews to investigate the role of security and privacy in people's decisions to use secure messaging apps. They reported that peer influence primarily drove decisions to use a particular secure messaging app; security and privacy were minor factors [14].

A number of prior research studies utilized interviews [4, 5, 9, 27, 68] or surveys [3, 5] to investigate users' mental models of E2E encryption. Similar to the findings of our first of five studies, these works identified a number of misconceptions regarding the security properties of E2E encryption. We based some of our survey questions on this prior work in an attempt to gain deeper insight into the root causes of users' security misconceptions and to try to mitigate such misconceptions.

Visualizing Encryption We discuss literature on visualizing encryption in three areas: web, e-mail, and messaging.

Visualizing and highlighting whether or not webpages are SSL/TLS-encrypted was historically a major focus of usable security research [58, 61]. In a lab setting, Whalen et al. conducted an eye-tracking study with 16 participants to test visual cues for SSL warnings, finding that icons provide prominent visual cues, yet they must be large and prominently placed [65]. Accordingly, we designed sufficiently large cues and placed them prominently in the center of our messaging app. Both Sobey et al. [57] and Maurer et al. [35] investigated alternative display methods, including full-browser themes, as security indicators of extended validity certificates. They

found that additional indicators of the level of security improved user confidence, the ease of finding information, and user understanding. Based on their work, we tested a number of variations for each type of visual cue. Schechter et al. conducted a qualitative lab study with 67 participants about the effect of removing security indicators on a banking website [54], finding that users ignore security indicators and that study designs incorporating role-playing reduce participants' security behaviours. More recently, in 2016 Felt et al. conducted a large quantitative online survey with 1329 participants, testing multiple cryptography-related labels and icons. They arrived at three indicators consisting of icons and labels for valid and invalid HTTPS and HTTP certificates to visualize the security level of the connection [23]. We built on this prior work by applying a similar but extended approach to the area of encrypted messaging apps, including the addition of qualitative elements and a validation study.

In the context of encrypting e-mail, related work investigates how user errors can be prevented and perceptions of security can be improved using security indicators. Two recent connected studies from 2013 and 2015 by Ruoti et al. proposed a web interface to support PGP encryption [48, 50]. They found that visualizing encryption using labels and adding scrambled text as an indicator of encrypted text helped to reduce user error when using PGP and supports trust in e-mail encryption. They proposed to further improve trust by letting users copy and paste e-mail ciphertext, but in a followup study found that doing so had no measurable effect on usability or security perceptions. Garfinkel et al. found that Key Continuity Management (KCM) systems with color-coded messages could improve e-mail security and effectively help novice users identify signed e-mails [26]. In 2015, Atwater et al. conducted a lab study investigating how a web interface can support e-mail encryption [8]. They found that participants prefer PGP to be integrated into their existing tool (e.g., Gmail). Participants' trust perceptions were based not on the tool's design, but rather the tool's overall reputation. Based on this finding that encryption should integrate into existing and well-known tools, we chose to test our own indicators using a modified version of the highly popular, E2E-encrypted WhatsApp Messenger.

Finally, we discuss related work regarding instant messaging and mobile apps. In 2012, Fahl et al. designed a tool for E2E encryption of private Facebook messages, evaluating the tool through lab and interview studies [21]. An artifact of their tool's implementation was that participants would see plaintext Facebook messages being translated to and from ciphertext. Their qualitative results implied that participants seeing the ciphertext upon sending or receiving messages was viewed positively and seemed to increase trust in the tool's security properties. In a lab study of the SELENE electronic voting protocol, Distler et al. [18] investigated how users reacted to seeing an explanation of encryption during the voting process. They found that overall perspicuity and users' perceptions of security increased due to the added waiting screen. In a followup online survey [17], they also tested different wordings of encryption in the scenarios of e-voting, online pharmacies, and online banking. They concluded that explanations of encryption should consist of short text without many elements, underpinning the design of the text disclosures we tested in Section 5.2.

In 2018, Demjaha et al. conducted an online study with 96 participants investigating metaphors to explain E2E encryption to users [16]. They concluded that wordings like "encryption" might be overloaded for end users and alternative metaphors might better explain the strengths and weaknesses of E2E encryption. While we focus on differences in structural explanations, we implement some metaphorical approaches in our icons and animations, measuring their effects compared to more straightforward labels and icons. Schröder et al. investigated authenticity-related error messages for the Signal [56] Android app [55]. They conducted a mostly qualitative study with 28 participants, finding that Signal needs to improve the awareness and verification of authenticity in conversations, as well as to communicate risks more clearly (e.g., providing guidelines for handling potential MITM attacks). Their findings suggest that the security perceptions of Signal could be improved in general.

In a recent study Akgul et al. evaluated if in-workflow messages in a messenger could improved the mental models of E2E encryption and found that while participants noticed them, they did not pay much attention to it, which limited the effect [5].

3 Methodology

We conducted a series of online studies on MTurk and Prolific (cf. Figure 1). This section gives a high-level overview of our approach. Section 4 details our study investigating current use of communication tools. Section 5 describes our studies on how different designs of encryption visualizations impact user perceptions.

Overall, we conducted five different user studies with 683 participants. For the first four, we recruited on MTurk. For the fifth, which was our validation study, we recruited on Prolific. We required participants in studies 2–5 be experienced WhatsApp users, enforcing this requirement through a qualification task on MTurk (cf. Section 5) and Prolific's built-in participant filters. We decided to use WhatsApp for our studies, since it is the most commonly used messenger with E2E encryption enabled by default in the US that is available on multiple platforms [59]. We estimated required participant numbers for each survey using power analysis and were limited by the total number of available WhatsApp users.

Each study had a distinct purpose:

Study 1: Use of Communication Tools The purpose of this study was to gain insight into the selection of communication tools for both day-to-day and confidential conversations. We aimed to understand how and why users decide to use certain

tools in particular circumstances. Table 4 illustrates messengers that were considered in this paper, and their features. Based on previous work [26, 50, 67], the results of Study 1, and the visual design of modern secure messaging apps, we then implemented potential encryption visualizations in a modern secure messaging app. Our goal was to investigate whether adding encryption visualizations to E2E-encrypted messaging app's UI would increase perceptions of trust, security, and privacy without sacrificing usability. The results for this study can be found in section 4.

Study 2: Disclosures Current secure messaging apps use specific textual framing (disclosures) to inform their users that conversations are E2E-encrypted. For example, WhatsApp displays *"Messages to this chat and calls are now secured with end-to-end encryption."*. However, prior studies on private browsing modes [69] and security warnings [22] have illustrated users' confusion about analogous disclosures. Therefore, we aimed to investigate whether a more detailed and technically correct ("end-to-end encrypted") disclosure had a different contribution to perceived security than more generous disclosures that are still connected to messaging security and comparatively tested six different versions. The results for this study can be found in section 5.2.

Study 3: Icons In addition to disclosures, a common approach is the use of security icons (e. g., lock symbols) [23] to indicate the presence of encryption or other security mechanisms. Similar to Study 2, we based our analysis on current secure messaging apps' security icons and icons discussed in previous usable security papers [23, 54]. We investigated three different icons, studying their impact on perceived trust, security and privacy, and usability. The results for this study can be found in section 5.3.

Study 4: Animations Additionally, we implemented and studied three animations of encryption. Prior work [18, 21] and the results of Study 1 implied that dynamic animations of the encryption process (e. g. disappearing messages or animations of plaintext turning into ciphertext) might increase perceptions of trust, security, and privacy. The results for this study can be found in section 5.4.

Study 5: Validation To validate and clarify the findings from studies 1–4, we performed a fifth study that addresses limitations of the previous four. One key challenge of studies 1–4 is the demographic bias of Amazon MTurk. Recent research identified generalizability and data quality issues on MTurk [31]. To account for this, we switched recruitment platforms, choosing Prolific [42]. Prolific provides strong tools to obtain a more diverse sample. Additionally, we performed the validation study to investigate root causes of particular results of Studies 2–4, so we also added an additional control condition and qualitative questions. To remove a potential confound suggested by the results of Studies 2–4, we also changed the messaging app from WhatsApp to a fictitious app we called Erebus. The results for this study can be found in section 6.

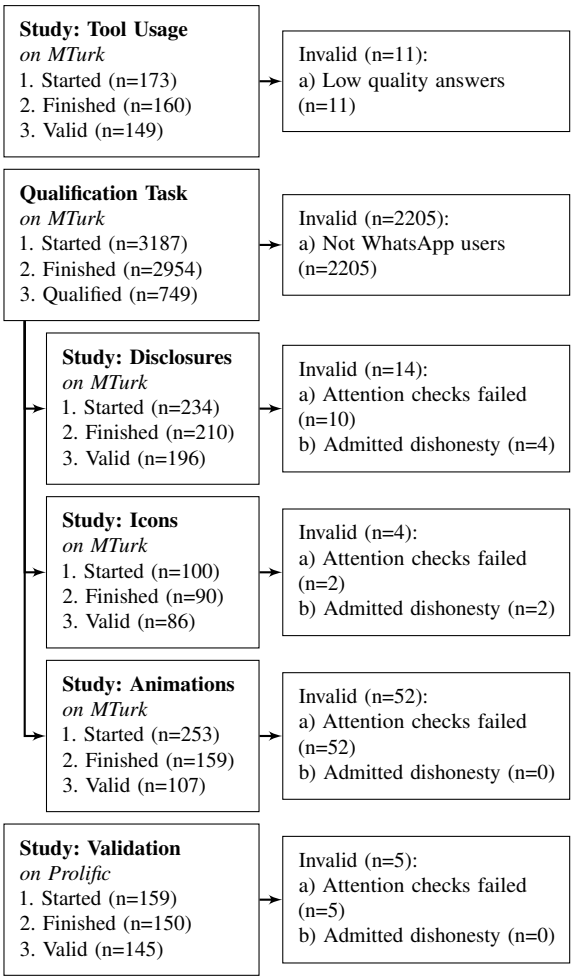


Figure 1: Illustration of our research procedure including *survey platform*, number of participants, and dropouts.

3.1 Study Procedure

We conducted all five studies sequentially to allow the findings of preceding studies to inform the design of later studies. For example, we used the most promising text disclosure from Study 2 in Studies 3–5.

Lab vs. Online Study Across studies 2–5 we investigated six different text disclosures (cf. Section 5.2), three different icons (cf. Section 5.3) and three different dynamic animations (cf. Section 5.4). Consequently, we recruited a rather high number of participants ($n = 534$ total participants). This made a laboratory experiment infeasible. Hence, we decided to conduct our experiments online using Amazon Mechanical Turk and Prolific Academic. Both platforms are popular amongst usable security and privacy user studies [1, 29, 37, 63].

Mockups vs. Real App We aimed for high internal validity to ensure that font sizes, types, positions of icons, animations, and the content of conversations (cf. Figure 7) remained consistent for all participants. Hence, we decided to use mockups

Factor	Description
Required	
Condition	Disclosures, icons, or animations (baseline: Control)
Optional	
CS Edu	Has CS education (self-reported, baseline: No)
CS Job	Has CS job (self-reported, baseline: No)
Age	Age in years (self-reported)

Table 1: Factors used in regression models. Model candidates were defined using all possible combinations of optional factors, with the required factors included in every candidate. Final models were selected by minimum AIC. Categorical factors are individually compared to the baseline.

instead of asking participants to install a real app on their devices. For the mockups, we created screencasts by forking the Signal Android app [56], since it implements the same encryption workflow that WhatsApp uses. We implemented the WhatsApp look and feel and all encryption visualizations. We recorded screencasts using the app and the conversation in Figure 7. During the conversation, we presented the different visualizations in each condition.

Additionally, each study had an online survey questionnaire at the end. The survey questionnaire addressed the perceived usability, trust, security, privacy and satisfaction with the tool, tool preference for both day-to-day and confidential conversations and demographic information about our participants.

Because we expected significant learning effect across conditions, Studies 2–5 followed a between-groups design.

Pre-Testing Before we conducted the studies, we pre-tested our questionnaires and screencasts, following best practices for cognitive interviews [24]. To glean insights into how survey respondents might interpret and answer questions and how they perceive the screencasts, we asked participants to share their thoughts as they answered each survey question and watched the screencasts. We used the findings to iteratively revise and rewrite our survey questions to minimize bias and maximize validity and modify the screencasts based on the feedback. We conducted cognitive interviews with members of our research group and university students, and performed a pre-test on MTurk to evaluate our survey questions under realistic conditions and to calibrate compensation relative to the time required. Pilots took an average of 15 minutes, so we compensated participants $2.50 (an hourly wage of $10).

Data Analysis Prior to data analysis, we took measures to ensure data quality (cf. A.5).

We perform both quantitative and qualitative data analysis. Throughout the paper, we measure usability using the UMUX Lite questionnaire [34]. We compare responses to the UMUX Lite across conditions with Pearson's chi-squared test (χ^2). We also collect net promoter scores, which are a quantitative measure of willingness to recommend a product. As these scores are continuous, we use the non-parametric Kruskal-Wallis H test (KW-H) for comparing conditions.

Because they might be influenced by multiple distinct factors, we analyze participants' perceptions of trust, security, and privacy by fitting linear regression models. For each regression analysis, we consider a set of candidate models and select the model with the lowest Akaike Information Criterion (AIC) [10]. We consider candidate models consisting of the "condition" (indicating the particular text disclosure, icon, or animation tested) plus every possible combination of optional factors. Required factors, optional factors, and corresponding baseline values are described in Table 1.

We present the outcomes of our regressions in tables where each row contains a factor and the corresponding change of the analyzed outcome in relation to the baseline of the given factor. Linear regression models measure change from baseline factors with a coefficient (*Coef.*) of zero for the value of the outcome. For each factor of a model, we also list a 95% confidence interval (*C.I.*) and a p-value indicating statistical significance. Also, we highlight p-values below $\alpha = 0.05$ with an asterisk (*).

We analyzed all free-text responses in an open-coding process [13, 60]. Two researchers iteratively developed a codebook [11], then used this initial codebook to code all free-text responses simultaneously, resolved coding conflicts, and incrementally updated the codebook until they were able to code open-ended questions without modifications to the codebook. The codebook remained stable once both researchers were satisfied that all important themes and concepts in the responses could be captured with the codes. Since the researchers resolved conflicts immediately as they emerged, we do not calculate inter-coder agreement [32].

3.2 Limitations

As with most self-reported online studies, our work has several limitations. In general, self-report studies may suffer from several biases, including over- and under-reporting, sampling bias, and social desirability bias. While we utilize self-report data, our central claims are not about the accuracy of respondents' answers to a given question, but rather about whether and how responses from different conditions differ from each other. Consequently, the threats to validity caused by those biases should apply equally across all conditions.

Conducting user studies on Amazon MTurk and Prolific is a widely used and accepted procedure for this type of research [39, 43]. However, MTurkers are known to be younger and more tech-savvy than the average population [43]. Additionally, our study focuses on the responses of U.S. Internet users, and thus, we can offer no insight into the generalizability of results for international participants.

Recently, the frequency of low data quality on MTurk has been increasing [31]. Therefore, we implemented a number of countermeasures (cf. Section 3.1). During data cleaning,

we identified several participants who did not pass our quality measures (Figure 1) and excluded them from further analysis.

We cannot guarantee that no participants were both registered MTurk and Prolific users and took more than one study, since there is no way to track people across both services. However, this is unlikely, since MTurk and Prolific target different geographic regions, and we conducted only the validation study on Prolific. Hence, we can guarantee that no participant took the same study twice.

Studies 2–5 tested a small set of different text disclosures, icons, and animations. While we based our designs on previous work and the results of our first study, we cannot guarantee that there are not other variants that work even better. Individual studies transpired in a somewhat isolated context, potentially missing certain effects of long-time exposure. We deliberately focused on multiple shorter studies, instead of one single in-depth, long-term study, to gather wider insights with different elements.

We showed our participants short screencasts (videos) in studies 2–5 instead of letting them use a real messaging application on their own devices. We aimed for high internal validity, so we wanted to ensure all participants would receive the same treatment. Comparable related work also worked with mockups instead of real applications for the same reason [18, 21]. While this experimental design results in lower external validity, we consider this tradeoff acceptable.

We decided to use a widely-deployed tool instead of a fictitious app mockup to study the challenges of visualizing E2E-encryption for an existing service provider and user base. We think our research provides valuable insights for a large set of users, although findings may not generalize to other E2E-encrypted messaging tools.

3.3 Ethical Considerations

We designed our studies with privacy in mind and followed best practices concerning data collection to ensure that we adhere to the German data- and privacy-protection laws aswell as the European General Data Protection Regulation. Our institution does not require a formal IRB, but we designed the study protocol based on a previous IRB approved study. All surveys started with a consent form to inform participants about the purpose of the study and about the data we would collect and store. The consent form also contained contact information to reach the PI in case of questions or concerns.

4 Use of Communication Tools (Study 1)

The main goal of Study 1 was to learn which communication tools our participants used and preferred for everyday and confidential conversations, as well as to learn about the decisions they made when using specific communication tools for particular conversations. In particular, we were interested in how many participants already used tools that provide E2E encryption by default for everyday conversations. We were especially interested in what fraction of them preferred less secure alternatives to E2E encrypted messengers for confidential conversations, and why. The questionnaire consisted of both closed- and open-ended questions. We followed the methodology described in Section 3 and developed the survey questionnaire in an iterative process, using pre-tests to improve the questionnaire, data quality, and determine appropriate compensation. We recruited 149 U.S.-based participants on MTurk.

4.1 Questionnaire Structure

We asked our participants to answer questions about their current use of communication tools for day-to-day and confidential conversations, as well as decisions they make when they choose one of the tools they have available for communicating with a single person or with groups of people. We decided to ask for specific tools or tool providers to glean insights into real behaviors and decision processes. We administered demographic questions at the end of the questionnaire to minimize stereotype bias [33, 53].

Past Tool Usage We asked participants which communication tools they have used in the last six months. The list of tools included the ten most popular tools in the U.S. [59]. We added iMessage, e-mail, and SMS to the list as popular messaging services that are pre-installed on many mobile devices by default. To better understand participants' choices and glean insights into their underlying mental models, we asked open-ended questions to explain their choices.

Security Assessments We asked participants to rate their perceived level of security when using personal e-mail, Facebook Messenger, WhatsApp, Snapchat, and SMS in the presence of different attackers. We chose these tools based on their popularity [2] and security properties (cf. Table 4 in the appendix).

Demographics We included several demographic questions about gender, age, ethnicity, education level, employment status, mobile device use, and the Security Behaviors Intentions Scale [19] for each participant. We aimed to assess whether demographic information would affect respondents' answers to the survey questionnaire. We also asked respondents for general feedback on the survey questionnaire.

4.2 Findings

We present both quantitative as well as qualitative results for the 149 valid respondents. The reporting of our findings focuses on actual tool usage in the past, insights into the perceptions, and decisions our participants made and their assessment of the security they think popular tools provide. Table 3 provides an overview of demographic characteristics of the participants in all studies.

Tool Usage Of the 149 participants in this study, the majority used regular e-mail (133; 89.26%), SMS/Text Messages (123; 82.55%) or the Facebook Messenger (114; 76.51%) that do not provide E2E encryption by default (cf. Figure 6). Only a few participants (7) reported having used PGP, S/MIME, or a provider supporting E2E encryption to secure e-mail conversations. Few participants (1) indicated prior use of Facebook's "Secret Conversation" feature. Overall, more than half of participants (80; 53.69%) reported use of an E2E-encrypted communication tool, with WhatsApp (47; 31.54%) being the most popular by far.

Tools that support E2E encryption as an optional feature, such as Facebook Messenger, Skype and Telegram (cf. Table 4), were also widely used (81.88%). However, only a few participants (9.02%) reported having used their E2E features.

While e-mail, SMS/text message, Facebook, WhatsApp, iMessage, and Skype are the most popular tools for both day-to-day and confidential conversations (cf. Figure 4 and Figure 5), a minority of participants (32; 21.48%) preferred none of the given tools for confidential conversations[1]. They only trusted non-digital forms of communication.

Even though they were users of E2E-encrypted communication tools for day-to-day conversations, many participants preferred e-mail and SMS for confidential conversations. Of the 80 participants who used E2E-encrypted tools for communication in general, the majority (50; 62.50%) preferred the use of insecure alternatives for confidential conversations. In particular, most (32; 68.09%) of the 47 WhatsApp users prefer less secure alternatives for confidential conversations.

Reasons for Using a Tool for Day-to-Day Conversations The main reason for people to use a certain communication tool for day-to-day conversation is ease of use (61.49%) followed by the availability of contacts in this tool (49.32%) and convenience (28.38%). One out of four (25.00%) participants also mentioned the delivery speed of text messages or instant messaging services and few (15.54%) mentioned the provided functionality. Some stated they are using a specific tool for a particular circle of people (11.49%) as mentioned by few participants: "*I belong to an online community for work and our main line of communication is through Facebook's messaging service.*" (P157), "*My husband uses Google hangouts too, and since I talk to him the most, this is the app I use most often.*" (P23), "*This is a group of family that has them, when I just need to relay info to that group I get on Telegram.*" (P27).

Few participants mentioned that they like a tool for storing a conversation history (5.41%), group chats (4.05%), message read info (5.40%) and disappearing messages (1.35%).

E-mail was an outlier as a preferred communication tool in many ways. Some participants (17.86% of e-mail users) prefer e-mail over other tools because they did not feel forced to reply to e-mails immediately:

"*It's more low key. There are no read receipts and you aren't expected to make a response immediately. You get to take your time.*" - P151.

E-mail has a professional reputation as it is often used in the workplace, which 16.06% of e-mail users noted. For 26.79% of e-mail users, a key reason to use e-mail is the support for large attachments and long text. This differs from all other tools, which are primarily instant-messaging services.

Reasons for Using a Tool for Confidential Conversations In two open-ended questions, we asked participants to elaborate on their preference for a specific tool for sensitive or confidential conversations and how they can tell that a specific tool keeps conversations confidential.

Almost half of participants (45.54%) mentioned a gut instinct that leads to a security belief as their main reason to prefer a specific tool for confidential conversations, e. g. "*I feel that it is safe.*" (P36).

A quarter of our participants (25.00%) assumed a tool to be confidential when they send messages directly to their intended contact and had their own name and the name(s) of the communication partner(s) being shown in the user interface. 16.96% mentioned access control and strong passwords as reasons to prefer a particular tool as mentioned by one participant: "*I have a secure E-mail that is guarded by a good strong password.*" (P123).

One out of four (26.79%) assumed a tool to be acceptable for confidential conversations because they thought it uses some form of encryption. However, 14.29% made wrong assumptions and thought encryption was being deployed on unencrypted channels (e. g., for SMS/text messages). Interestingly, only a few (8.04%) referenced "secret mode" or "secure chat" options in their decision.

6% of our participants also reported using SMS as a confidential channel because it is not an internet service: "*It is sent from me to another person, not on the internet.*" (P31) and "*It feels off the grid, away from the dangers of the internet.*" (P66)

For a few (3.57%), visual indicators like colors or icons earned trust even if they did not directly relate to security or privacy e. g. "*If the message is blue it should be encrypted.*" (P148). In the iOS messenger, a blue message indicates that a message was sent via iMessage and a green message indicates that it was sent as a Text Message.

Few participants mentioned self-destructing and disappearing messages (4.46%), as in SnapChat, or the ability to delete messages manually (3.57%), as offered in WhatsApp, as influencing their preference:

"*I know that gmail for example encrypts messages and I trust google to be safe.*" (P77)

At the same time, half of the participants (50%) could not report specific reasons for their trust in a particular tool.

> Key Insights: Tool Usage, Decisions and Security Beliefs.
>
> - E-mail and SMS/text messages are the most popular tools for both day-to-day and confidential conversations.

[1] None is an exclusive option and deselected the other fields.

- 53.69% of participants use a communication tool with E2E encryption enabled by default.
- 62.50% of participants who use E2E-encrypted tools prefer less secure alternatives for confidential conversations.
- Participants reported a gut instinct that made them believe a tool to be secure.

5 Visualizing Encryption (Study 2–4)

Both previous work and the findings of our first study illustrate that the situation around E2E-encrypted communication tools is complicated. Many users will avoid installing a new, more secure messaging tool [2] only because it provides better security [51, 67]. Instead, most users only consider messaging tools if their contacts (i. e. friends, family, and colleagues) also use the tools [14]. Additionally, previous work [14, 68], and our first study show many people suffer from misunderstandings and misconceptions of encryption.

Instead of propagating the more widespread use of such niche tools or working on correcting users' misunderstandings and misconceptions alone, we followed a different route. Depending on geographic region, between half of users (cf. Section 4) and 90% [2] of users *already* have tools that support E2E encryption by default, with WhatsApp being the most popular. However, our findings (cf. Section 4) suggest that many users are not aware of these security properties. More than half of our participants who use WhatsApp prefer less secure alternatives such as e-mail or SMS/text messages for confidential conversations. Therefore, the remainder of our studies investigate how visualizing encryption impacts perceptions of E2E messaging security.

While the results of our first study (cf. Section 4) and previous work [3, 4, 14, 18, 21, 68] uncover a wide range of root causes for misconceptions about the security of messengers and insecure behaviour, only some of them can be addressed in the design of a communication tool. For example, we identified that trusting a company or decades of positive experiences were both root causes for misconceptions. However, these can hardly be addressed in the design of a communication tool. In contrast, there are promising candidates that can directly be implemented in the user interface of a communication tool (cf. Section 3). In this section, we describe multiple online studies we conducted with the goal to investigate the impact of different encryption disclosures, icons and animations on perceived trust, security, privacy, usability, satisfaction and self-reported likeliness to use the re-designed communication service for sensitive messages.

5.1 Experiment Design

To study visualizations of encryption using text disclosures, icons, and animations, we conducted four between-groups online experiments with WhatsApp users recruited on MTurk or Prolific. Each study follows the procedure we outline below.

5.1.1 Screencasts

To study the impact of different encryption visualizations on usability, perceived trust, security, privacy, satisfaction and tool preference, we decided to show participants a screencast of a ficticious WhatsApp update [2].

Using a screencast instead of static mockup images allowed us to study both static and dynamic encryption visualizations (cf. Section 3) and include a scripted conversation to provide more context for our participants[3]. We constructed the messages this way because it mimics a realistic personal conversation and credit card information is generally perceived as confidential and worth protecting.

To mimic WhatsApp as closely as possible, we forked the Android version of the Signal mobile app and adapted the user interface respectively by changing colors, typefaces, buttons and other user interface properties.

5.1.2 Questionnaire Structure

The survey questionnaire in this study was developed through an iterative process (cf. Section 4) and included the attention checks mentioned in Section 3. Completion of the survey took 10 minutes on average and we paid participants $1.7.

Usability, Trust, Security, Privacy and Satisfaction We asked participants to answer usability, perceived trust, security and privacy and satisfaction questions. For usability, we asked participants the two items UMUX lite scale [34]. Based on prior work [44], we built a 10-item scale of perceived trust, security and privacy (cf. Appendix A.2). Finally, we asked participants to fill out the net promoter score [28] to measure how much they liked the encryption visualization.

Tool Preference We showed participants a list of the most popular communication tools from our first study (cf. Section 4) including the new ficticious WhatsApp version and asked them which tool they would prefer for both day-to-day and confidential conversations. To prevent lock-in obstacles as found in [14], we told all participants to assume that all communication partners have all tools installed. We aimed to assess whether our conditions had an effect on the participants' choice.

Demographics We asked our respondents the same demographic questions as in the questionnaire in Section 4. Table 3 provides an overview of demographic characteristics of the participants in the studies in this section.

5.2 Text Disclosures (Study 2)

Based on the disclosures in current tools that support E2E-encrypted communication (cf. Table 4) and the results of our first study (cf. Section 4), we created six different disclosures out of the terms "secret", "private", "encrypted", "secure"

[2] cf. Appendix A.6 for the video introduction
[3] cf. Appendix A.4 for the conversation

	Factor	Coef.	C.I.	*p*-value
Disclosure	"*Messages to this chat are now . . .*"			
	"*. . . private*"	0.27	[0.21, 1.07]	0.392
	"*. . . secret*"	-0.49	[-1.09, 0.11]	0.111
	"*. . . secure*"	0.06	[-0.53, 0.65]	0.845
	"*. . . encrypted*"	0.68	[0.09, 1.28]	0.030 *
	"*. . . end-to-end encrypted*"	-0.09	[-0.69, 0.51]	0.768
	"*. . . secured with end-to-end encryption*"	0.41	[-0.19, 1.01]	0.182
Icon	Icon (Baseline: Control):			
	Envelope	-0.47	[0.92, 1.69]	0.105
	Lock	-0.49	[-1.09, 0.11]	0.089
	Shield	-0.70	[-1.27, -0.14]	0.014 *
	CS Education	-0.51	[-0.97, -0.05]	0.029 *
Animation	Animation (Baseline: Control):			
	Disappearing Messages	0.08	[-0.34, 0.51]	0.707
	Encryption/Decryption	-0.01	[-0.40, 0.38]	0.969
	Progress Circle	0.25	[-0.14, 0.66]	0.210
	Age	-0.01	[-0.14, 0.65]	0.119
Validation	Animation (Baseline: Control without Disclosure):			
	Control	0.45	[0.03, 0.86]	0.034 *
	Disappearing Messages	0.40	[0.01, 0.79]	0.043 *
	Encryption/Decryption	0.43	[0.04, 0.81]	0.030 *
	Progress Circle	0.71	[0.33, 1.10]	< 0.001 *

Table 2: Results of the linear regression model examining whether different texts, icons and animations have an effect on the trust, security and privacy score in relation to a control baseline. Note the additional "Control" variable in the last study, due to "Control Without Disclosure" being the baseline. See Table 1 for further details.

and "end-to-end encrypted." In a between-groups design, we randomly assigned participants to one of the following conditions:

1. Control: "*blank*"
2. Encrypted: "*Messages to this chat are now **encrypted**.*"
3. E2E-Encrypted: "*Messages to this chat are now **end-to-end encrypted**.*"
4. Private: "*Messages to this chat are now **private**.*"
5. Secure: "*Messages to this chat are now **secure**.*"
6. Secure & E2E: "*Messages to this chat are now **secured** with **end-to-end encryption**.*"
7. Secret: "*Messages to this chat are now **secret**.*"

To make sure participants read the text of each disclosure, we showed them a screencast in fullscreen before entering the conversation for seven seconds including the respective disclosure. Overall, we recruited 196 valid participants on MTurk for whom we report findings below.

Findings We were specifically interested in the participants' opinions on usability and their perceptions of trust, security, and privacy.

As a usability metric we compared the distribution of UMUX Lite answer categories between our conditions. We found no apparent differences between the conditions (Q1: Pearson's $\chi^2 = 1.56$, p-value $= 1$; Q2: Pearson's $\chi^2 = 1.22$, p-value $= 1$), suggesting no observable effect (positive or negative) on the perceived usability of the different disclosures.

To better investigate how the different conditions affect participants' perception of trust, privacy, and security, we introduced a combined score based on their answers to our set of 10 likert-item questions. For each participant, the score consists of the average of all 10 likert-item questions mapped to numerical values, e.g., between -2 (Strongly Disagree) and +2 (Strongly Agree). For these scores, we considered a set of linear regression models consisting of the conditions as required factor and all combinations of optional factors listed in Table 1 and selected the model with the lowest AIC.

The final model (see Table 2) shows that the "encrypted" condition is significant with an overall positive coefficient of about 0.7 score points compared to the control baseline. This suggests significantly higher scores for the "*Messages to this chat are now encrypted*" disclosure compared to the blank control, which is in line with previous research by Distler et al. [17]. Participants seemed to prefer the encryption text, likely due to not fully understanding the term "end-to-end," or regarding it as a subset (i.e., less secure) of being "just" encrypted.

For the net promoter score, we found that no condition dominates any other (Kruskal-Wallis $H = 7.88$, p-value $= 0.24$). Based on these results, we proceeded with the "encrypted" text for our subsequent disclosure.

> Key Insights: Disclosures.
>
> - Participants felt most secure and private within the "encrypted" disclosure condition.
> - The different disclosures did not have a significant impact on usability and satisfaction.

5.3 Icons (Study 3)

Next, we investigated three different icons. We chose a lock, a shield, and an envelope (cf. Figure 2) based on their typical usage in security and privacy contexts [7, 38], previous work in the field of security indicators [23, 54], and results from our first study (cf. Section 4). Together with the best-performing text disclosure from the previous study *"Messages to this chat are now encrypted,"* we showed participants one of the icons before the communication partners in the screencast entered the conversation.

(a) Envelope (b) Lock (c) Shield

Figure 2: Designs used in the encryption icons study.

We showed participants a screencast including the "encrypted" disclosure from the previous study and the respective

encryption icon. All screencasts lasted 94 seconds. A total of 86 WhatsApp users participated in this study. Findings reported below are limited to these valid participants.

Findings For the UMUX Lite questionnaire we found no significant differences across conditions (Q1: Pearson's $\chi^2 = 0.54$, p-value $= 1$; Q2: Pearson's $\chi^2 = 0.40$, p-value $= 1$).

Our set of linear regression models for the overall score included the icon condition as required factor and again all combinations of optional factors (cf. Table 1). To our surprise, the final model (See Table 2) shows that all three icon conditions are worse than the baseline by at least 0.47 score points. The shield condition is significantly worse by 0.7 score points. In addition, the optional computer science education factor is significant with a negative coefficient in the model. This is in line with previous work [23, 54] and additional evidence for the very limited effect of security icons on perceived trust, security and privacy.

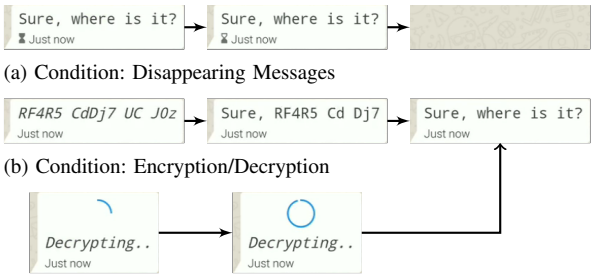
For the net promoter score, we found that no condition dominates any other (Kruskal-Wallis $H = 5.68$, p-value $= 0.128$). We chose to proceed to the next study using the control (no icon) due to the negative coefficients of all other conditions.

> Key Insights: Security Icons.
>
> - We found a negative effect of security icons on perceived trust, security and privacy by at least 0.47 score points compared to the baseline.
> - Participants with a computer science background particularly disliked the security icons we investigated, resulting in 0.51 less score points compared to participants without that background.
> - The security icons had no impact on usability and satisfaction.

5.4 Animations (Study 4)

In addition to text disclosures and encryption icons, previous work [18, 21] and the results of our first study (cf. Section 4) suggest the use of animations of the encryption process to convey that a conversation is secure. We identified three different encryption animations: (i) Distler et al. [18] used a progress circle for an e-voting app; (ii) Fahl et al. [21] studied dynamic encryption and decryption animations to protect Facebook messages; and (iii) participants in Study 1 reported feeling particularly secure with disappearing messages on apps like Snapchat. Although disappearing messages are not technically connected to E2E-encryption, we included them due to their contribution to perceived messaging security identified in previous work by Roesner et al. [45] and participants' comments in Study 1. One participant for example said it was security relevant "*Because the conversation deletes right after I read it.*" (P9) and another said "*... once you open it, it's gone forever afterward.*" (P14)

We implemented those three animations (cf. Figure 3). In contrast to Studies 2–3, we applied the dynamic encryption animations to the screencast conversation's messages, rather than as a fullscreen hint before entering a conversation. We



(a) Condition: Disappearing Messages

(b) Condition: Encryption/Decryption

(c) Condition: Progress Circle

Figure 3: Conditions in the security animations study.

pre-tested the animation duration with 20 MTurkers. Initially, we showed the animation for three seconds. Based on UMUX Lite and qualitative feedback, we gradually reduced the duration to one second. Based on the results from the previous studies, each condition included one of the encryption animations and a small text hint with the "Encrypted" disclosure. Based on our Study 3 results, we did not use an icon in this study.

Overall, we recruited 107 valid participants on MTurk for whom we report findings below.

Findings We found no impact of the different animation conditions on the UMUX Lite questionnaire (Q1: Pearson's $\chi^2 = 0.37$, p-value $= 1$; Q2: Pearson's $\chi^2 = 0.42$, p-value $= 1$). The final linear regression model includes a somewhat increased, but not significant, coefficient (0.25) for the "Progress Circle Animation" condition compared to the baseline, and almost non-existent positive and negative effects (0.08 and -0.01) for the other two animations (cf. Table 2). For the net promoter score of the animation conditions, we found again, as in the other two studies, that no condition dominates any other (Kruskal-Wallis $H = 1.63$, p-value $= 0.654$).

> Key Insights: Security Animations.
>
> - No factors were significant. The progress circle animation had a weak positive effect on perceived trust, security and privacy.
> - Security animations did not impact usability or satisfaction.

6 Validation (Study 5)

Due to Study 4's inconclusive findings regarding the effect of animations (cf. Section 5.4), we validated our overall findings in a fifth study. Motivated by the lower data quality encountered in Study 4 compared to Studies 1–3, as well as the increasing difficulty recruiting WhatsApp users on MTurk, we switched to the Prolific recruitment platform for Study 5. Prolific provides fine-grained participant demographics, so we could directly target participants with messenger experience without requiring a qualification task. Since Prolific's pool of US workers who use WhatsApp was small (<500), we included UK participants, increasing the participant pool by 6,000. On Prolific, we recruited 145 participants from the UK and US, paying £2.70. In addition to changing recruitment

platforms, we decided to delve into the root causes of why visualizing encryption appeared to have a limited impact on user perceptions. For this, we dropped both the UMUX Lite and net promoter score as we observed no significant differences for them in our previous studies. We also added open-ended followups to each Likert question to gain deeper insights into participants' opinions. Additionally, we added Likert questions that focused on what facets suggest that messages are being sent securely. To eliminate participants' perceptions of WhatsApp as a major factor guiding their perceptions, we also changed the messenger name to "Erebus." As this study was intended as validation, we especially focused on the "encryption" text from our first survey by including a new control that displayed no text at all ("Control without Text"). We retained the previous control condition to compare with the previous studies.

Findings As for the previous surveys, we generated a linear regression model listed in Table 2. The final regression models have significant non-zero coefficients for all included variables relative to our new control (not mentioning encryption at all). Going by coefficient, the Progress animation performs best compared to the baseline (0.71), followed by Encryption/Decryption (0.43) and Disappearing (0.40). Even the control condition from the previous surveys ("Control") shows a significant coefficient compared to the newly introduced baseline "Control without Text". This suggests a significant effect of the "encryption" text.

In addition to the regression analysis, we evaluated the open-ended questions to gain insight into the limited impact of encryption visualization. We report findings below.

Observing Animations In an open-ended question, we asked participants what they observed happening (if anything) when messages were sent or received, as well as what this indicated. Almost all participants described the animations we showed them ($> 90\%$ in each condition). 60 participants (41.38%) wrote that the animation indicated an increased level of security. Hence, we can eliminate the possibility of participants ignoring the animations as the reason perceptions did not vary significantly across conditions.

Identifying Security In an open-ended question, we asked participants how they determine, in general, that a messaging app sends messages securely. The most prominent indicator for security was the *reputation* (48, 33.10%) of the service provider, followed by the mention of *encryption* (42, 28.96%).

> "*Honestly, I guess I just trust in the brand that it's safe. I do this through the popularity, good press and confidence in their service.*" - P18

The relatively similar relevance of *encryption* and *reputation* for perceived security also explains the limited impact of the presence of encryption on perceptions of security.

Identifying Encryption Given that encryption is an important security mechanism, we asked participants to detail how they identify the presence of encryption in a messaging app. Most participants report relying on textual information in the form of *disclosures* (40, 27.59%) or an app's *feature list* (11, 7.59%) mentioning encryption. However, 42 participants (28.97%) said they would not know how to recognize encryption's presence. Very few mentioned visual indicators. For example, 5 (3.44%) mentioned observing a *delay during sending*, 3 (2.07%) mentioned messages disappearing, and 2 (1.38%) mentioned seeing messages be *scrambled*. These explanations are consistent with our regression analyses (cf. Table 2), highlighting the limited effect of visualizations.

> Key Insights: Validation.
>
> - Study 5 confirmed the findings of Studies 2–4.
> - Visualizing encryption in any way, even a simple text disclosure, improves perceptions compared to not mentioning it at all.
> - Most participants saw the animations and felt they communicated "security", yet this did not change their perceptions any more than a text disclosure did.
> - An app's reputation greatly impacts perceptions.

7 Discussion

In our first of five studies, we investigated why participants use particular messaging tools, validating a prior finding [3] that many users mistakenly think SMS and e-mail are more secure than E2E-encrypted messengers. Based on these initial findings, we aimed to improve the visibility of E2E encryption in a messaging app. Across the four subsequent studies, we compared six different text disclosures, three different icons, and three different animations of the encryption process.

Impact of Encryption Visualization While investigating the impact of different visualizations of encryption, we were surprised to find that the simple "encrypt" disclosure outperformed most others (aside from the progress circle) in terms of perceived trust, security, and privacy. As expected, however, all disclosures performed better than the baseline of having no disclosure at all. We were also surprised to see that security icons had a negative effect, rather than increasing perceptions of trust, security, and privacy. This negative effect was particularly distinct for people with a CS background.

Previous work suggested that encryption visualizations might positively impact perceived trust, security, and privacy [18, 21, 48]. Those suggestions were based primarily on qualitative data. Our studies, which combined quantitative and qualitative data, reached somewhat different conclusions. Based only on the qualitative data we collected, one might have reached conclusions similar to those of prior work. For example, as reported in Section 6, nearly half of participants indicated that the animations of encryption indicated an increased level of security. In contrast, our quantitative analyses indicated that these different animations did not have a significantly different impact on perceptions of the trust, security, and privacy of E2E-encrypted messaging tools than a straightforward text disclosure that the conversation in encrypted, which is what many secure messaging apps currently display.

These findings call into question the magnitude and applicability of the effects reported in prior work.

Our findings suggest that highlighting the use of encryption in basic ways (e.g., "*Messages to this chat are now encrypted*") significantly increases perceived security, privacy and trust in messaging applications. That is, having *any* visualization of encryption outperformed the control in our validation study of not calling attention to the use of encryption at all. However, richer visualizations of encryption involving icons or animations seem to have only a limited additional effect. Although we did not observe these richer visualizations of encryption to significantly impact user perceptions and satisfaction in a positive direction compared to basic text disclosures, we also did not observe a negative effect.

8 Recommendations

Given the promise of rich visualizations of encryption reported in prior work, this finding is disappointing, as it suggests that simple modifications of messaging apps' UIs are unlikely to help users better assess apps' security and privacy. Despite the use of multiple design proposals from previous work, we could not find a significant improvement (Sec. 7).

Our qualitative results imply that instead of investing more effort into studying richer visualizations of encryption, focusing on the following aspects is potentially more promising. We make recommendations for both providers of E2E encrypted communication tools and usable security researchers.

8.1 Tool Providers

Trust in Company As we have seen in the qualitative answers in the tool usage and validation study (Sec. 4.2, 6), participants report that they trust the brand and that the company would keep their data secure. Tool providers could focus on generally improving trust in the brand.

Convenience Several participants mentioned using a specific app to communicate with their peer groups that decided on that app (Sec. 4.2). Introducing an app or feature that is not compatible with their peer groups leads to them switching back to another channel. Tool providers should make sure that E2E encryption features do not lead to inconveniences for their users.

Functionality Our participants also mentioned that they switched the tools when a messenger did not support a required feature, for example with large attachments that they send via mail (Sec. 4.2). That indicates that a full feature set is required to avoid people switching to insecure channels. Making E2E encryption available in communication tools should not limit existing functionality.

8.2 Usable Security Research

Correcting Mental Models Our participants showed a number of incorrect mental models, most strikingly: Around 25% of our participants assumed that their conversations are free of eavesdroppers if the user interface shows only the names of their intended communication partner(s) (Sec. 4.2) and show a lack of understanding of man-in-the-middle attacker capabilities. Also, 14.29% of participants falsely assumed channels that are generally not encrypted by default (e.g., SMS) to be encrypted (Sec. 4.2). These misconceptions likely impact the usage of secure and private messengers significantly. Addressing them better should be a major goal for our community.

Technical Background As seen in our regression for Study 3 (Table 2), participants with a technical background tended to rate trust in security indicators lower. Investigating factors that contribute to this perception and provide improvements for these factors (e.g., increase company transparency) could help address concerns unique to that demographic.

9 Conclusion

We studied whether making a messaging app's E2E encryption more visible improves perceptions of trust, security, and privacy. To that end, we conducted five online studies with 683 total participants, including a summative validation study.

While participants felt most secure and private within the "encrypted" text disclosure condition, the different text disclosures did not have a significant impact on usability and app satisfaction. We observed a surprising negative effect of security icons on perceived trust, security, and privacy. When focusing on animations, none of the factors was statistically significant, though we identified a weak positive effect for the progress circle animation on perceived trust, security, and privacy. The animations had no impact on usability and satisfaction. We confirmed these key findings in a final summative study, validating that visualizing encryption in any way, even a simple text disclosure, improves perceptions compared to not mentioning encryption at all. Most participants saw the animations, the richest and most novel aspect of our investigation, and reported qualitatively they communicated "security." However, quantitative perceptions of trust, security, and privacy did not differ significantly compared to a text disclosure.

In our first study, we replicated the finding of prior work that a non-trivial fraction of users mistakenly believes SMS and e-mail to be more secure than E2E-encrypted messengers. While we had hypothesized that richly visualizing the process of encryption would emphasize E2E-encrypted messaging apps' security properties and combat this misconception, our results suggest that the existing practice of disclosing the use of encryption in a straightforward text disclosure may be sufficient if the text disclosure is displayed prominently.

References

[1] Noura Abdi, Kopo M. Ramokapane, and Jose M. Such. More than Smart Speakers: Security and Privacy Perceptions of Smart Home Personal Assistants. In

Proc. 15th Symposium on Usable Privacy and Security (SOUPS'19). USENIX Association, 2019.

[2] Ruba Abu-Salma, Kat Krol, Simon Parkin, Victoria Koh, Kevin Kwan, Jazib Mahboob, Zahra Traboulsi, and M Angela Sasse. The Security Blanket of the Chat World: An Analytic Evaluation and a User Study of Telegram. In *Proc. 2nd European Workshop on Usable Security (EuroUSEC'17)*. The Internet Society, 2017.

[3] Ruba Abu-Salma, Elissa M. Redmiles, Blase Ur, and Miranda Wei. Exploring user mental models of end-to-end encrypted communication tools. In *Proc. 8th USENIX Workshop on Free and Open Communications on the Internet (FOCI'18)*. USENIX Association, 2018.

[4] Ruba Abu-Salma, M. Angela Sasse, Joseph Bonneau, Anastasia Danilova, Alena Naiakshina, and Matthew Smith. Obstacles to the Adoption of Secure Communication Tools. In *Proc. 38th IEEE Symposium on Security and Privacy (SP'17)*. IEEE, 2017.

[5] Omer Akgul, Wei Bai, Shruti Das, and Michelle L. Mazurek. Evaluating In-Workflow Messages for Improving Mental Models of End-to-End Encryption. In *Proc. 30th Usenix Security Symposium (SEC'21)*. USENIX Association, 2021.

[6] Apple Inc. iMessage - Learn more about Messages. https://support.apple.com/explore/messages, 2019.

[7] Apple Inc. Human Interface Guidelines - iOS Design Themes. https://developer.apple.com/ios/human-interface-guidelines/overview/themes/, 2020.

[8] Erinn Atwater, Cecylia Bocovich, Urs Hengartner, Ed Lank, and Ian Goldberg. Leading Johnny to Water: Designing for Usability and Trust. In *Proc. 11th Symposium on Usable Privacy and Security (SOUPS'15)*. USENIX Association, 2015.

[9] Wei Bai, Michael Pearson, Patrick Gage Kelley, and Michelle L. Mazurek. Improving Non-Experts' Understanding of End-to-End Encryption: An Exploratory Study. In *Proc. 2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 2020.

[10] K. P. Burnham. Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, 33(2):261–304, 2004.

[11] Kathy Charmaz. *Constructing Grounded Theory*. SAGE Publications, 2014.

[12] Sandy Clark, Travis Goodspeed, Perry Metzger, Zachary Wasserman, Kevin Xu, and Matt Blaze. Why (Special Agent) Johnny (Still) Can't Encrypt: A Security Analysis of the APCO Project 25 Two-Way Radio System. In *Proc. 20th Usenix Security Symposium (SEC'11)*. USENIX Association, 2011.

[13] Juliet Corbin and Anselm Strauss. Grounded theory research: Procedures, canons and evaluative criteria. *Zeitschrift für Soziologie*, 19(6):418–427, 1990.

[14] Alexander De Luca, Sauvik Das, Martin Ortlieb, Iulia Ion, and Ben Laurie. Expert and Non-Expert Attitudes towards (Secure) Instant Messaging. In *Proc. 12th Symposium on Usable Privacy and Security (SOUPS'16)*. USENIX Association, 2016.

[15] Sergej Dechand, Dominik Schürmann, Karoline Busse, Yasemin Acar, Sascha Fahl, and Matthew Smith. An Empirical Study of Textual Key-Fingerprint Representations. In *Proc. 25th Usenix Security Symposium (SEC'16)*. USENIX Association, 2016.

[16] Albesë Demjaha, Jonathan Spring, Ingolf Becker, Simon Parkin, and M. Angela Sasse. Metaphors considered harmful? An exploratory study of the effectiveness of functional metaphors for end-to-end encryption. In *Proc. Workshop on Usable Security (USEC'18)*. The Internet Society, 2018.

[17] Verena Distler, Carine Lallemand, and Vincent Koenig. Making Encryption Feel Secure: Investigating how Descriptions of Encryption Impact Perceived Security. In *Proc. 5th European Workshop on Usable Security (EuroUSEC'20)*. IEEE, 2020.

[18] Verena Distler, Marie-Laure Zollinger, Carine Lallemand, Peter B. Rønne, Peter Y. A. Ryan, and Vincent Koenig. Security - Visible, Yet Unseen? In *Proc. CHI Conference on Human Factors in Computing Systems (CHI'19)*. ACM, 2019.

[19] Serge Egelman and Eyal Péer. Scaling the Security Wall: Developing a Security Behavior Intentions Scale (SeBIS). In *Proc. 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI'15)*. ACM, 2015.

[20] Ksenia Ermoshina, Francesca Musiani, and Harry Halpin. End-to-End Encrypted Messaging Protocols: An Overview. In *Proc. 6th International Conference on Internet Science (INSCI'19)*. Springer, 2016.

[21] Sascha Fahl, Marian Harbach, Thomas Muders, Matthew Smith, and Uwe Sander. Helping Johnny 2.0 to Encrypt His Facebook Conversations. In *Proc. 8th Symposium on Usable Privacy and Security (SOUPS'12)*. ACM, 2012.

[22] Adrienne Porter Felt, Alex Ainslie, Robert W. Reeder, Sunny Consolvo, Somas Thyagaraja, Alan Bettes, Helen Harris, and Jeff Grimes. Improving SSL Warnings: Comprehension and Adherence. In *Proc. 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI'15)*. ACM, 2015.

[23] Adrienne Porter Felt, Robert W. Reeder, Alex Ainslie, Helen Harris, Max Walker, Christopher Thompson, Mustafa Embre Acer, Elisabeth Morant, and Sunny Consolvo. Rethinking Connection Security Indicators. In *Proc. 12th Symposium on Usable Privacy and Security (SOUPS'16)*. USENIX Association, 2016.

[24] Ronald P Fisher and R Edward Geiselman. *Memory-Enhancing Techniques for Investigative Interviewing: The Cognitive Interview*. Charles C Thomas Publisher, 1992.

[25] Simson L. Garfinkel, David Margrave, Jeffrey I. Schiller, Erik Nordlander, and Robert C. Miller. How to Make Secure Email Easier To Use. In *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI'05)*. ACM, 2005.

[26] Simson L. Garfinkel and Robert C. Miller. Johnny 2: A User Test of Key Continuity Management with S/MIME and Outlook Express. In *Proc. 1st Symposium on Usable Privacy and Security (SOUPS'05)*. ACM, 2005.

[27] Nina Gerber, Verena Zimmermann, Birgit Henhapl, Sinem Emeröz, and Melanie Volkamer. Finally Johnny Can Encrypt: But Does This Make Him Feel More Secure? In *Proc. 13th International Conference on Availability, Reliability and Security (ARES'18)*. ACM, 2018.

[28] D. F. Hamilton, J. V. Lane, P. Gaston, J. T. Patton, D. J. MacDonald, A. H. R. W. Simpson, and C. R. Howie. Assessing treatment outcomes using a single question: the net promoter score. *The Bone & Joint Journal*, 96(5):622–628, 2014.

[29] Julia Hanson, Miranda Wei, Sophie Veys, Matthew Kugler, Lior Strahilevitz, and Blase Ur. Taking Data Out of Context to Hyper-Personalize Ads: Crowdworkers' Privacy Perceptions and Decisions to Disclose Private Information. In *Proc. CHI Conference on Human Factors in Computing Systems (CHI'20)*. ACM, 2020.

[30] David J. Hauser and Norbert Schwarz. Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, 48(1):400–407, 2016.

[31] Ryan Kennedy, Scott Clifford, Tyler Burleigh, Ryan Jewell, and Philip Waggoner. The Shape of and Solutions to the MTurk Quality Crisis. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3272468, 2018.

[32] Klaus Krippendorff. *Content Analysis: An Introduction to Its Methodology (2nd ed.)*. SAGE Publications, 2004.

[33] Jon A. Krosnick and Stanley Presser. *Question and Questionnaire Design*, pages 263–314. Emerald Publishing, 2010.

[34] James R. Lewis, Brian Utesch, and Deborah E. Maher. UMUX-LITE: when there's no time for the SUS. In *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI'13)*. ACM, 2013.

[35] Max-Emanuel Maurer, Alexander De Luca, and Tobias Stockinger. Shining Chrome: Using Web Browser Personas to Enhance SSL Certificate Visualization. In *Proc. 13th IFIP TC 13 International Conference on Human-Computer Interaction (INTERACT '11)*. Springer, 2011.

[36] Micah Lee. Edward snowden explains how to reclaim your privacy. https://theintercept.com/2015/11/12/edward-snowden-explains-how-to-reclaim-your-privacy/, 2015.

[37] Mainack Mondal, Günce Su Yilmaz, Noah Hirsch, Mohammad Taha Khan, Michael Tang, Christopher Tran, Chris Kanich, Blase Ur, and Elena Zheleva. Moving Beyond Set-It-And-Forget-It Privacy Settings on Social Media. In *Proc. 26th ACM Conference on Computer and Communication Security (CCS'19)*. ACM, 2019.

[38] Jakob Nielsen. Enhancing the Explanatory Power of Usability Heuristics. In *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI'94)*. ACM, 1994.

[39] Stefan Palan and Christian Schitter. Prolific.ac — A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17:22–27, 2018.

[40] Eyal Peer, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti. Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70:153 – 163, 2017.

[41] Eyal Peer, Joachim Vosgerau, and Alessandro Acquisti. Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods*, 46(4):1023–1031, 2014.

[42] Prolific. Prolific | Online participant recruitment for surveys and market research. https://prolific.co/, 2020.

[43] Elissa M Redmiles, Sean Kross, and Michelle L Mazurek. How Well Do My Results Generalize? Comparing Security and Privacy Survey Results from MTurk, Web, and Telephone Samples. In *Proc. 40th IEEE Symposium on Security and Privacy (SP'19)*. IEEE, 2019.

[44] Juan Carlos Roca, Juan José García, and Juan José de la Vega. The importance of perceived trust, security and privacy in online trading systems. *Information Management & Computer Security*, 17(2):96–113, 2009.

[45] Franziska Roesner, Brian T Gill, and Tadayoshi Kohno. Sex, Lies, or Kittens? Investigating the Use of Snapchat's Self-Destructing Messages. In *Proc. 18th International Conference on Financial Cryptography and Data Security (FC'14)*. Springer, 2014.

[46] Christoph Rottermanner, Peter Kieseberg, Markus Huber, Martin Schmiedecker, and Sebastian Schrittwieser. Privacy and Data Protection in Smartphone Messengers. In *Proc. 17th International Conference on Information Integration and Web-based Applications & Services (iiWAS'15)*. ACM, 2015.

[47] 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 *Proc. CHI Conference on Human Factors in Computing Systems (CHI'16)*. ACM, 2016.

[48] Scott Ruoti, Jeff Andersen, Travis Hendershot, Daniel Zappala, and Kent E. Seamons. Private Webmail 2.0: Simple and Easy-to-Use Secure Email. In *Proc. 29th Annual Symposium on User Interface Software and Technology (UIST'16)*. ACM, 2016.

[49] Scott Ruoti, Jeff Andersen, Daniel Zappala, and Kent E. Seamons. Why Johnny Still, Still Can't Encrypt: Evaluating the Usability of a Modern PGP Client. *ArXiv e-prints*, 2016.

[50] Scott Ruoti, Nathan Kim, Ben Burgon, Timothy Van Der Horst, and Kent Seamons. Confused Johnny: When Automatic Encryption Leads to Confusion and Mistakes. In *Proc. 9th Symposium on Usable Privacy and Security (SOUPS'13)*. ACM, 2013.

[51] M. A. Sasse, S. Brostoff, and D. Weirich. Transforming the 'Weakest Link' — a Human/Computer Interaction Approach to Usable and Effective Security. *BT Technology Journal*, 19(3):122–131, 2001.

[52] J. Schaad, B. Ramsdell, and S. Turner. Secure/Multipurpose Internet Mail Extensions (S/MIME) Version 4.0 Message Specification. https://tools.ietf.org/html/rfc8551, April 2019.

[53] Nora Cate Schaeffer and Stanley Presser. The Science of Asking Questions. *Annual Review of Sociology*, 29(1):65–88, 2003.

[54] Stuart E Schechter, Rachna Dhamija, Andy Ozment, and Ian Fischer. The Emperor's New Security Indicators. In *Proc. 28th IEEE Symposium on Security and Privacy (SP'07)*. IEEE, 2007.

[55] 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 *Proc. 1st European Workshop on Usable Security (EuroUSEC'16)*. The Internet Society, 2016.

[56] Signal, a 501c3 nonprofit. Signal Messenger. https://www.signal.org, 2019.

[57] Jennifer Sobey, Robert Biddle, Paul C. van Oorschot, and Andrew S. Patrick. Exploring User Reactions to New Browser Cues for Extended Validation Certificates. In *Proc. 13th European Symposium on Research in Computer Security (ESORICS'08)*. Springer, 2008.

[58] Andreas Sotirakopoulos, Kirstie Hawkey, and Konstantin Beznosov. On the challenges in usable security lab studies: Lessons learned from replicating a study on ssl warnings. In *Proc. 7th Symposium on Usable Privacy and Security (SOUPS'11)*. ACM, 2011.

[59] Statista Inc. Most popular mobile messaging apps in the United States as of September 2019, by monthly active users. https://www.statista.com/statistics/350461/mobile-messenger-app-usage-usa, 2019.

[60] Anselm Strauss and Juliet M Corbin. *Grounded theory in practice*. SAGE Publications, 1997.

[61] Joshua Sunshine, Serge Egelman, Hazim Almuhimedi, Neha Atri, and Lorrie Faith Cranor. Crying Wolf: An Empirical Study of SSL Warning Effectiveness. In *Proc. 18th Usenix Security Symposium (SEC'09)*. USENIX Association, 2009.

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

[63] Anthony Vance, David Eargle, Jeffrey L. Jenkins, C. Brock Kirwan, and Bonnie Brinton Anderson. The Fog of Warnings: How Non-essential Notifications Blur with Security Warnings. In *Proc. 15th Symposium on Usable Privacy and Security (SOUPS'19)*. USENIX Association, 2019.

[64] WeAreDynamo.org. Guidelines for Academic Requesters. http://wiki.wearedynamo.org/index.php/Guidelines_for_Academic_Requesters, 2017.

[65] Tara Whalen and Kori M. Inkpen. Gathering Evidence: Use of Visual Security Cues in Web Browsers. In *Proc. Graphics Interface 2005 Conference (GI'05)*. Canadian Human-Computer Communications Society, 2005.

[66] WhatsApp LLC. WhatsApp - Features. https://www.whatsapp.com/features/, 2019.

[67] Alma Whitten and J. D. Tygar. Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0. In *Proc. 8th Usenix Security Symposium (SEC'99)*. USENIX Association, 1999.

[68] Justin Wu and Daniel Zappala. When is a Tree Really a Truck? Exploring Mental Models of Encryption. In *Proc. 14th Symposium on Usable Privacy and Security (SOUPS'18)*. USENIX Association, 2018.

[69] Yuxi Wu, Panya Gupta, Miranda Wei, Yasemin Acar, Sascha Fahl, and Blase Ur. Your Secrets Are Safe: How Browsers' Explanations Impact Misconceptions About Private Browsing Mode. In *Proc. 27th International Conference on World Wide Web (WWW'18)*. International World Wide Web Conferences Steering Committee, 2018.

[70] Philip Zimmermann. PGP Version 2.6.2 User's Guide. ftp://ftp.pgpi.org/pub/pgp/2.x/doc/pgpdoc1.txt, October 1994.

A Appendix

A.1 Demographics

Table 3 shows the demographics of participants in all studies.

	Study: Tool Usage	Study: Disclosures	Study: Icons	Study: Animations	Study: Validation
Participants					
Started	173	234	100	253	159
Finished	160	210	90	159	150
Valid ($n=$)	149	196	86	107	145
Gender					
Male	60.7%	54.1%	47.7%	67.3%	40.7%
Female	37.9%	44.9%	51.2%	29.0%	57.9%
Not M/F	1.4%	1.0%	1.2%	2.8%	1.4%
Ethnicity†					
White	78.5%	68.9%	68.6%	71.0%	89.0%
Asian or Pacific Islander	6.0%	14.3%	16.3%	7.5%	6.9%
Black or African American	13.4%	7.7%	11.6%	14.0%	0.0%
Hispanic or Latino	4.7%	12.2%	11.6%	10.3%	2.8%
Native American	0.7%	1.5%	0.0%	0.0%	0.0%
Other & Prefer not to say	0.7%	0.5%	1.2%	0.9%	4.1%
Smartphone OS†					
Android	67.1%	57.7%	41.9%	60.7%	59.3%
iOS	37.6%	50.5%	66.3%	39.3%	40.7%
Other	0.0%	1.0%	3.5%	0.0%	0.0%
No smartphone	1.3%	0.0%	0.0%	0.0%	0.0%
Prefer not to say	0.7%	0.0%	1.2%	0.0%	0.0%
Computer Science					
CS Education	28.9%	24.5%	24.4%	31.8%	26.9%
CS Job	22.1%	26.0%	29.1%	32.7%	19.3%
Age in years					
Mean	37.5	33.9	35.6	32.5	37.7
Std. dev. (σ)	10.7	9.0	10.9	8.4	10.5
Median	35.0	33.0	33.0	31.0	35.0

† Multiple answers allowed, may not sum to 100%

Table 3: Participant demographics.

A.2 Scale of Perceived Trust, Security and Privacy

Ten item scale of perceived trust, security and privacy. Participants choose from a 5-point likert scale on each question.

1. I think the new WhatsApp version is trustworthy.
2. I do not doubt the honesty of the new WhatsApp version.
3. I think the new WhatsApp version is secure.
4. I think only me and the recipient(s) can read our messages.
5. I think other people cannot send a message pretending to be me.
6. I think no one can unnoticeable modify messages sent between me and the recipient(s).
7. I think that if somebody hacks my phone, they will not be able to read my messages.
8. I think only me and the recipient(s) can know the messages were sent.
9. I think the new WhatsApp version does not collect more personal information than strictly needed.
10. I think the new WhatsApp version will not use my personal information for other purposes without my authorization.

A.3 Messenger Usage

The following figures show the messenger usage among our participants in the first survey. Figure 4 shows the preferred messenger for day-to-day conversations, Figure 5 shows the preferred messenger for sensitive or confidential conversations. Figure 6 shows all messengers used in the last 6 months.

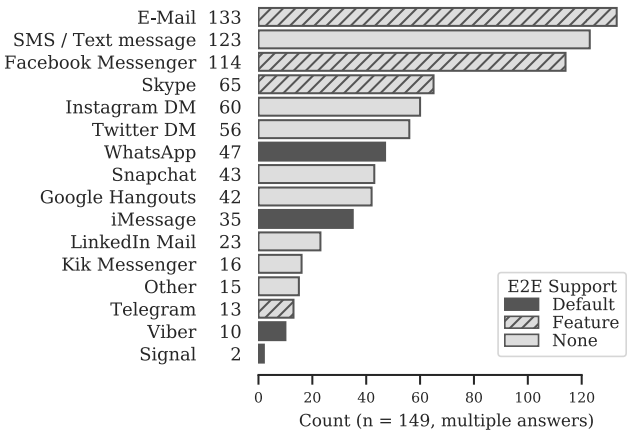


Figure 6: Study: Tool Usage - "Which online communication tools have you used in the last 6 months?"

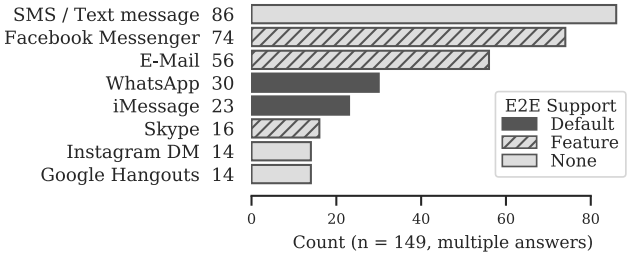


Figure 4: Study: Tool Usage - "Which tools do you prefer for day-to-day conversations?" (Top 8)

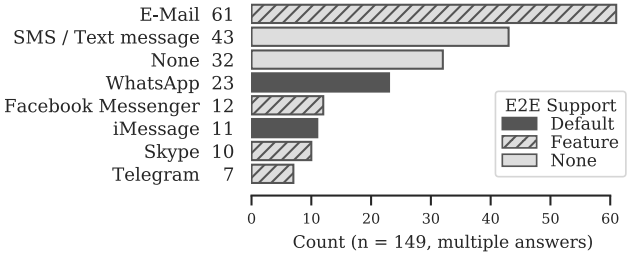


Figure 5: Study: Tool Usage - "Which tools do you prefer for sensitive or confidential conversations?" (Top 8)

A.4 Scripted Conversation

The text in Figure 7 was used in the scripted conversations that were shown to the participants in the videos. The overall screencast took on average 95 seconds.

A.5 Data Quality

Participant diversity and data quality on MTurk and Prolific is generally perceived as satisfactory [39–41]. We followed best practices [31, 41, 64] and required workers to be U.S. residents who have already completed 100+ HITs with a 95% approval rate.

During piloting, we experienced similar data quality issues as reported in recent work [31]. Therefore, we implemented a set of countermeasures, including blocking participants whose IP address came from outside the U.S. or belonged to a VPN or proxy service provider even within the U.S.[4]

Following best practices [30, 31], we added three attention checks to all questionnaires.

To remove a potential confound for Studies 2–5, we wanted to include only participants familiar with secure messaging apps. Therefore, we added an MTurk qualification task in which we asked participants which messaging apps they currently used and invited only WhatsApp users to the actual study itself. Workers who had participated in one study were ineligible for all subsequent ones. We paid each participant $0.15 for the short qualification task.

We took advantage of the pre-screening provided by Prolific, where participants had to report the regular use of WhatsApp in a pre-screening questionnaire[5], and required participants to be located in the U.S. or UK, have a 95% or higher approval rate and at least 100 previous submissions.

Me: Hi, darling. I forgot my wallet at home. Could you please look up my credit card number? I need to place an order before I forget.

<Wait 10s>

Remote: Sure, where is it?

Me: It should be on my desk.

<Wait 5s>

Remote: One second.

<Wait 25s>

Remote: It's 1234-5678-9012-3456, valid until 12/21, and the security code is 456.

Me: Thanks

Figure 7: Chat messages displayed in the application.

A.6 Study Video Introduction

We introduced the video to our participants in the following way:

[4] We used the https://iphub.info service to filter VPNs and proxies.

[5] The exact question we asked in the pre-screening questionnaire was: Which of the following chat apps do you use regularly? [multiple-choice]

Name (Alphabetical order)	E2E-Encrypted (Protocol Name)	E2E Indicator Color	Icon	Text	Platforms Android	iOS	Downloads (On Android)
E-Mail	○	-	-	-	●	●	-
E-Mail with PGP or S/MIME	● (PGP or S/MIME)	d	d	d	●	●	-
Facebook Messenger	◐ (Signal)	●	Lock	●[1]	●	●	5.000M+
FaceTime	● (SRTP)	○	○	○	○	●	-
Google Hangouts	○	-	-	-	●	●	5.000M+
iMessage	● (unknown)	●	○	○	○	●	-
Instagram DM	○	-	-	-	●	●	1.000M+
Kik Messenger	○	-	-	-	●	●	100M+
LinkedIn InMail	○	-	-	-	●	●	500M+
Signal	● (Signal)	○	Lock	○	●	●	50M+
Skype	◐ (Signal)	○	○	●[2]	●	●	1.000M+
SMS	○	-	-	-	●	●	-
Snapchat	○	-	-	-	●	●	1.000M+
Telegram	◐ (MTProto2.0)	●	Lock	●[3]	●	●	500M+
Twitter DM	○	-	-	-	●	●	1.000M+
Viber	● (unknown)	●	Shield[4]	●[5]	●	●	500M+
WhatsApp	● (Signal)	○	Lock	●[6]	●	●	5.000M+

● Yes (for E2E-Encrypted: Yes, by default) ○ No ◐ Has a "secret mode" which uses E2E encryption, but is not active by default [d] Depends on the client used [1] *Secret conversation* [2] *Private conversation* [3] *Secret chat* [4] *Has an additional Secret Chat, uses a lock icon* [5] *Messages sent in this conversation are encrypted* [6] *Messages to this chat and calls are now secured with end-to-end encryption*

Table 4: List of popular communication tools.

"*Imagine that WhatsApp will soon release a new version. This new version would have a different user interface in some places but have the same features as the version you are used to. Below we show you a brief video of what this new user interface for WhatsApp might look like.*"

A.7 Replication Material

The videos and questions used within this paper are available on our webpage at https://publications.teamusec.de/2021-soups-e2e/.