Peeking Into the Black Box: Towards Understanding User Understanding of E2EE

Leonie Schaewitz
leonie.schaewitz@rub.de
Ruhr-Universität Bochum
Bochum, NRW, Germany

David Lakotta
david.lakotta@rub.de
Ruhr-Universität Bochum
Bochum, NRW, Germany

M. Angela Sasse
martina.sasse@rub.de
Ruhr-Universität Bochum
Bochum, NRW, Germany

Nikol Rummel
nikol.rummel@rub.de
Ruhr-Universität Bochum
Bochum, NRW, Germany

ABSTRACT

End-to-end encryption (E2EE) has become available to end users, but they need to understand the nature and limitations of the protection it offers to benefit in terms of protection. Attempts to explain cryptography in general, and E2EE in particular, to non-specialists have had limited success – in part because they tried to convey detailed expert knowledge. Metaphors are a way to communicate the benefits and limitations more compactly, and support the construction of functional mental models. Previous research that attempted to do this for E2EE reported mixed results, but offered no detailed insight into how participants constructed their understanding and which aspects of particular metaphors helped or hindered their functional understanding. We repeated the previous experiment in form of a qualitative interview study with 12 participants (all users of messaging apps) and used detailed questions to better understand why the participants rated the security properties of E2EE correctly or incorrectly, and how the metaphors had been interpreted and applied. Therefore, we are able to describe to what extent, and how, the metaphors for E2EE changed participants' understanding of the security properties. We found that participants inferred the security properties of E2EE partly from the metaphors, but also from existing beliefs, for instance about the trustworthiness of providers. While the metaphors improved the assessment about confidentiality, they did not correct misconceptions about authenticity. Based on our findings we recommend the development and testing of interventions aimed at the process of changing mental models and correcting persistent misconceptions.

CCS CONCEPTS

• **Security and privacy** → **Social aspects of security and privacy**; • **Human-centered computing** → *User models*; *User studies*.

EuroUSEC '21, October 11–12, 2021, Karlsruhe, Germany

ACM ISBN 978-1-4503-8423-0/21/10.
https://doi.org/10.1145/3481357.3481521

KEYWORDS

End-to-end encryption, mental models, metaphors, messaging apps

ACM Reference Format:
Leonie Schaewitz, David Lakotta, M. Angela Sasse, and Nikol Rummel. 2021. Peeking Into the Black Box: Towards Understanding User Understanding of E2EE. In *European Symposium on Usable Security 2021 (EuroUSEC '21), October 11–12, 2021, Karlsruhe, Germany*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3481357.3481521

1 INTRODUCTION

With the rapid growth of digital communication, the need to protect personal communication from third-party access is also becoming more important for citizens. Cryptographic solutions, in particular end-to-end encryption (E2EE), have been available to the public in recent years, but adoption has lagged behind in cases where non-experts have to make a deliberate choice to use them – for instance, to use an E2E encrypted messaging service. Previous research has identified gaps in understanding and issues with usability as causes [22]. Even when a tool such as WhatsApp is used (the provider added E2EE when WhatsApp already had a significant user base), many users are not aware of the protection it offers – and switch to telephone services, SMS, or email when sending sensitive information. [1, 2]. On the other hand, if users adopt a service with E2EE without being aware of the limits of the protection it offers, it can create a false sense of security [5].

In an effort to promote secure behaviour and help users make more informed decisions, there have been many endeavours to explain how encryption works to non-experts. But these often try to convey expert knowledge in form of detailed structural mental models [13], i.e., detailed representations of how the system works, and use terminology users are not familiar with [29]. To enable non-expert users to make secure choices, functional mental models of encryption would seem a more promising starting point – they are less extensive and complex because they are rooted in the tasks users are already familiar with, and contain only the elements and properties of a system the user needs to perform those tasks [13] (Young coined term task/action mapping models [31]). Previous studies that have examined users' mental models of (E2E) encryption [1, 18, 30] have focused, implicitly or explicitly, on a more detailed understanding, i.e. structural models. One of the few attempts to 'shortcut' this deeper understanding is the study by Demjaha et al. [11] who investigated metaphors as a means to

tap existing mental models and transform them into functional understanding of E2EE by offering metaphor-based descriptions of E2EE, and testing understanding of 4 security properties before and afterwards in an online survey study. The results were inconclusive: although the metaphors led some participants to identify security properties correctly, they caused others to incorrectly assume security properties that E2EE does not provide. The conclusion was that none of the metaphors tested evoked a 'correct' mental model in participants [11]. Since the study was an online survey in which users chose pre-determined answers, it was not possible to probe the mental models the participants had developed in response to the metaphors. Our study addresses exactly that: we examine what functional understanding of E2EE participants had, and how this understanding is connected to the assessments they made. Basically, we try to peek into the 'black box' of users' reasoning processes, their understanding of the key security properties of E2EE, and where the limits of protection are. We conducted qualitative interviews with 12 participants who used instant messengers, the prevalent use case of E2EE, to answer the following two research questions:

(1) What are the underlying reasons for correct or incorrect assessments by users of the security properties of E2EE and the security properties E2EE does not provide?
(2) How are users' assessments influenced based on their interpretations of metaphors for E2EE?

Our results show that users build their functional mental models of E2EE partly on their understanding of technical aspects and the security properties of digital communications, and partly on their beliefs about the motives and capabilities of service providers and hackers as well as on their personal experiences with cybersecurity incidents. We found that the metaphors improved participants' understanding of confidentiality as a security property of E2EE; however, the security property of authenticity was still not well understood. We propose to implement interventions that attempt to address and change existing mental models and correct persistent misconceptions rather than attempting to teach new models.

2 RELATED WORK

2.1 Mental models of encryption

Since the seminal 1999 paper "Why Johnny can't encrypt" [29] by Whitten and Tygar, it has been known that the concept of public-key cryptography is difficult for people without IT security expertise to understand. Whitten herself [28] argued that a full understanding of public-key encryption was necessary to use tools such as PGP correctly, and the intervening years have shown that people are not willing or able to acquire that full understanding. Several studies have examined users' mental models of secure communications and encryption, in general [2, 30] or in the context of different applications, such as HTTPS [18], email [5, 22], and messenger services [1, 10, 15]. This research has shown that mental models of encryption are typically sparse, correspond to a model of symmetric encryption, and can essentially be reduced to a functional abstraction of access control [30]. Moreover, studies have identified a number of misconceptions and knowledge gaps about encryption and secure communication, including:

- a belief that encryption is futile, because it does not protect against skilled attackers, such as hackers or governmental actors [1, 2, 18, 30]
- a belief that application providers can still read E2E encrypted messages [10, 15]
- a belief that standard audio calls and SMS are more secure than E2E encrypted messages [1, 2, 9, 10]
- a missing concept of end point authentication in a cryptographical sense [1, 10, 18, 23] and a belief that authenticity is controlled by passwords. [2]

In addition, research has shown that users are unaware of the difference between opportunistic and authenticated E2EE [16, 23], and apply their everyday understanding to the process of verifying a person's identity (e.g., identifying a person by name, phone number, voice, or a personal question) [1, 10, 23]. These findings suggest that an improvement of users' mental models of encryption could increase their adoption and correct use of secure communication services. However, only few studies have actually attempted this or have proposed approaches to improve users' mental models of E2EE.

2.2 Approaches to improve user understanding and mental models of encryption

Bai et al. [5] developed a tutorial to provide 'high-level' information about E2EE, and tested its effectiveness in a lab-based study. While the study showed that the tutorial improved users' understanding of E2EE, some misconceptions remained, and some new ones appeared. Several participants remained unconvinced that encryption cannot be broken, or that E2E encrypted messages are more secure than voice conversations. Moreover, terms like integrity and authenticity were difficult to grasp, and some participants developed a false sense of security, assuming that E2EE could also protect against malware and access to end point devices [5]. Akgul et al. [3] found that educational messages designed to improve functional mental models of E2EE could improve understanding of key security concepts (e.g., in terms of who could/could not intercept their messages) compared to a control message in an experimental questionnaire study, but not in a more realistic usage setting when embedded in a messaging app that participants actually used.

Another approach to conveying a functional understanding of E2EE builds on the use of metaphors. Metaphors can support users build a mental model for a new knowledge domain by pointing out similarities to an already known knowledge domain that can be used as a starting point for model building [7, 25].

Tong et al. [26] investigated the effectiveness of visual metaphors for public key cryptography based on a quantitative, quiz-based study design, providing preliminary results that show some improvement in users' understanding. Focussing on functional metaphors for E2EE, Demjaha et al. [11] generated metaphors based on users' descriptions of E2EE, and investigated whether they influence users' understanding of security properties that E2EE provides (confidentiality, authenticity) and those it does not provide (protection of meta data and messages on end devices). By means of a pre-post measure of comprehension questions, the authors tested and compared the influence of three metaphors as well as existing descriptions used by WhatsApp and Telegram on users' understanding

of E2EE. They found that particularly the descriptions provided by WhatsApp and Telegram slightly improved the understanding of the security properties of E2EE in some participants, but at the same time increased the number of incorrect assessments of security properties E2EE does not provide (e.g., participants also assumed that messages were kept encrypted on their phones). The authors found no statistically significant differences between the metaphors/descriptions of E2EE.

Although the interventions by Demjaha et al. [11] are plausible ways to try to improve users' understanding of encryption, the study did not provide sufficient detail on why they were partly successful, and partly not. How did users arrive at their assessments of a functional understanding of E2EE? How were the metaphors understood and processed by the participants in order to contribute to their understanding of the security properties of E2EE? Why has their understanding improved in some cases and not improved (or even worsened) in others?

The study we conducted aims to close this gap by re-using the metaphors with a set of participants, and conducting qualitative interviews asking in-depth questions about how users understand the metaphors as well as the security properties that E2EE provides and not provides, to examine how users' construct their functional understanding. Thereby this study goes beyond the quantitative approach by Demjaha et al. [11] by revealing deeper insights into users' reasoning.

3 METHODOLOGY

We conducted 12 semi-structured interviews with German instant messenger users (6 males, 6 females), to examine how they think about different security properties of E2EE and security properties E2EE does not provide in the context of instant messenger use and to investigate how they construct their functional understanding of E2EE. By randomly showing each interview participant one of three functional metaphors for E2EE and discussing them in relation to the different security properties, we explored whether they changed their functional understanding based on their interpretation of the metaphors and why or why not. All interviews were conducted in German and via the DFNconf web conferencing service [12] by the same interviewer and lasted between 20 to 35 minutes. Participants were compensated with a 10 € Amazon gift card for their participation. The interviewer conducted a pilot interview with a student assistant as interviewee to practice the interview guide, identify comprehension difficulties, and test the technical setup of the web conferencing service.

3.1 Interview procedure

The interview was divided into four parts and followed the structure of the questionnaire developed by Demjaha et al. [11]. The first part started with some general questions about participants' instant messenger use, including: "Which messengers do you use?", "Which messenger do you use most frequently?", and "How frequently do you use them?". Moreover, we asked about their general perceptions and beliefs about the security of communications, by asking "Why is sending messages with one of the following instant messengers secure or insecure: WhatsApp, iMessage, Signal, Telegram, Facebook Messenger (all offering E2E encryption)?

In the second part, we presented participants four statements, each expressing a functionality or non-functionality of E2EE. The statements were shown individually on a PowerPoint slide, and for each statement, participants were asked whether they believed the statement to be true or false about communicating with one of the instant messenger services, why they believed the statement to be true or false, and how confident they were in their answer. We explicitly did not mention the term *end-to-end encryption*.

We used the following four statements from [11]:

> **STMT1:** "Only you and the recipient can read your messages" (functionality; true)
> **STMT2:** "Other people can send a message pretending to be you" (functionality, false)
> **STMT3:** "Only you and the recipient can know the messages were sent" (non-functionality, false)
> **STMT4:** "If somebody hacks your phone, they will be able to read your messages" (non-functionality, true)

In the third part of the interview, we presented the participants with one of the following three metaphorical descriptions of E2EE [11]:

> **Special Language (SL):** "Messages and calls with this person will be translated to a special language for which only the two of you know the dictionary."
> **Treasure Hunt (TH):** "Messages and calls exchanged with this person are like a treasure hidden in a place to which only the two of you know the map."
> **Colours (C):** "Messages and calls you exchange with this person are like colours. Before sending them, you mix them with another colour, known only by you two. Nobody else can retrieve them unless they know the secret colour."

Each participant was randomly assigned one metaphor, which was presented as a textual description on a PowerPoint slide. We asked participants to explain in their own words how they imagine communicating with the messenger service based on this description. Afterwards, participants were once more asked to evaluate the four (non-)functionality statements – this time with the metaphor in mind (and visible on screen). For each statement, we asked them whether they wanted to change or stick to their previous answer and why. In the final part, we informed them about the concept of E2EE by presenting the 'technical' explanation defined in RFC 2828 [24] and provided the correct answers to the statements (as defined by Demjaha et al. [11]).

> **'Technical' description of E2EE**: "Continuous protection of data that flows between two points in a network, effected by encrypting data when it leaves its source, keeping it encrypted while it passes through any intermediate computers (such as routers), and decrypting it only when it arrives at the intended final destination" [24, p. 121].

Wherever a participant had given an incorrect answer during the second round of evaluation, we asked them what explanation might help to convey the protection provided/not provided by E2EE. Moreover, we asked participants to compare the metaphor with

the technical explanation of E2EE and asked for suggestions for improvements of metaphorical explanations of E2EE. Finally, we asked for some basic demographic information (age, occupation, IT knowledge), gave participants a chance to ask questions, and thanked them for their participation. The interview guide is listed in the Appendix A.1.

3.2 Sample and recruitment

We recruited the participants for our interviews between April and June 2020 by posting the announcement in different German Facebook groups, e.g., local networking groups, groups for students/student life, groups for searching study participants, groups related to our university. Participants were required to be 18 years or older and users of instant messengers.

We were able to recruit an equal number of male and female participants (6 each), with heterogeneous professional activities: Four participants were students or PhD students, the others worked in marketing, healthcare, tourism, social work, as computer scientist, electronics engineer, or software consultant. The age of our participants ranged from 24 to 37. Nobody had a background in IT security.

All participants used messenger services frequently. The most frequently used service by all participants was WhatsApp. Moreover, the following services were used (by # of participants): Facebook Messenger (10), Telegram (7), Instagram (5), Signal (2), Snapchat (2), Jodel (1), KiK (1), Line (1), and Viber (1).

3.3 Qualitative analysis and coding procedure

All interviews were recorded and transcribed. The transcripts were content analysed by two researchers, one with a background in educational science, one with a background in psychology and cognitive science, in a data-driven and iterative approach, using the software MAXQDA.

First, for each of the four statements presented about the security properties of E2EE and those E2EE does not provide, we coded once before and once after showing the metaphor whether the interview participant rated it as true or false. Then, an open coding procedure [8] was used to identify patterns from participants' responses describing the reasons why they believed the statements to be true or false, both at baseline and after exposure to the metaphor. In addition, also the section that focused on perceptions and understanding of the metaphors was subjected to an open coding procedure to identify common themes in respondents' answers. Both coders first independently coded all questions and then developed a coding scheme based on several rounds of discussion. The scheme was developed partly based on the research questions and interview sections (structure of the main categories) and partly inductively based on the material (generation of sub-codes), and finally applied to all interviews.

4 RESULTS

In the following sections, we first provide a general overview of participants' ratings of the four statements (Section 4.1). We then present the reasons for participants' initial assumptions about the security properties of E2EE and security properties that E2EE does not provide (Section 4.2 to Section 4.5). Finally, we present results on the perceptions of the metaphors and to what extent these, in combination with people's prior preconceptions, influence their assumptions about the security properties of E2EE and the properties E2EE does not provide (Section 4.6).

The frequencies are given to illustrate the prevalence of the themes, no quantitative analysis was carried out.

4.1 Rating of statements

Participants' initial responses indicated uncertainty and misconceptions about E2EE, both with regard to confidentiality (Statement 1: "Only you and the recipient can read your messages") and authentication (Statement 2: "Other people can send a message pretending to be you"). 75% of respondents misjudged at least one of the two statements (see left part of Table 1). In addition, we found fewer misconceptions about the security properties that E2EE does not provide, that is, metadata protection (Statement 3: "Only you and the recipient can know the messages were sent") and endpoint security (Statement 4: "If somebody hacks your phone, they will be able to read your messages"). In fact, all participants correctly identified that when the end-point device is accessed, the content of the messages is no longer protected (Statement 4).

After the presentation of the metaphors (see right part of Table 1), we saw an improvement for the first statement (confidentiality). Understanding of authenticity (Statement 2), however, yielded fewer correct responses – except for two participants who saw the *Colours* metaphor. The *Treasure Hunt* and the *Colours* metaphor each lead to an incorrect assessment of Statement 3 in one case, indicating an overestimation of the protection that E2EE provides.

In the following sections, we try to uncover the reasons for these assessments by analysing the explanations participants gave for their assessments. Participants often mentioned reasons for why a statement might be true, but also for why it might not be true. The following sections are organized by the order of the statements. Table 2 gives an overview of the codes for participants' reasons for judging the statements as true or false.

4.2 Beliefs about STMT1: "Only you and the recipient can read your messages" (functionality; true)

During the initial discussion of the statements (before the metaphor), we identified three key reasons why people believe Statement 1 to be true: **Trust in the providers** of messenger services to secure communications properly (n = 4), **trust in the supervision of providers** (i.e., that experts or laws control that the communication is secure; n = 2), or a belief that Statement 1 is 'generally' true because it would not be easy (or require **great effort**) to access the messages, which would not be worthwhile for most messages (n = 2). An example of a trust-based explanation is provided by P3SL: *"because I hope that there are enough experts in this field who would have possibly already checked this and that it would have already become public if this was not the case."*

On the other hand, participants gave several arguments why Statement 1 would have to be false: Half of the participants assumed that **the app provider can read the messages** – a misconception that has been consistently shown in previous studies [1, 2, 10, 15]. The arguments given by participants were:

Table 1: Overview of participants' ratings of the statements.

	Initial Evaluation					Evaluation after Metaphor			
P#	STMT 1	STMT 2	STMT 3	STMT 4	Metaphor	STMT 1	STMT 2	STMT 3	STMT 4
P1SL	false	true	false	true	Special Language	true	true	false	true
P2SL	true	true	false	true	Special Language	true	true	false	true
P3SL	true	false	false	true	Special Language	true	true	false	true
P4SL	true	true	true	true	Special Language	true	true	false	true
P1TH	true	true	false	true	Treasure Hunt	true	true	false	true
P2TH	true	false	true	true	Treasure Hunt	true	true	true	true
P3TH	false	false	false	true	Treasure Hunt	true	false*	false	true
P4TH	false	false	false	true	Treasure Hunt	false	false	false	true
P1C	false	true	false	true	Colours	true	false	true	true
P2C	true	true	false	true	Colours	true	false	false	true
P3C	false	true	false	true	Colours	false	true	false	true
P4C	true	false	false	true	Colours	true	true	false	true

Note. ■ ■ = correct, ■ ■ = incorrect answer; ■ ■ very sure; ■ ■ somewhat sure/with constraints; *but argues only for true; STMT1: "Only you and the recipient can read your messages"; STMT2: "Other people can send a message pretending to be you"; STMT3: "Only you and the recipient can know the messages were sent"; STMT4: "If somebody hacks your phone, they will be able to read your messages"

(1) The messages are cached on the provider's servers, and accessible to anyone with access to the server
(2) The provider has an interest in reading the message content for advertising purposes
(3) The provider must have access because authorities might demand to read messages

One participant shared the suspicion that the provider could be reading along simply because he had heard about it (hearsay), while another referred to media coverage as well as his own experiences (with personalized advertising). In addition, one participant assumed that other persons with an interest (i.e., **hackers**) could access or hack the data, arguing that *"you just hear a lot in the media that messenger services have been hacked (...) You never know if it's messages (that they hack) or really just (...) contact information, numbers, etc. I think if people can intercept the numbers, they can also intercept the messages"* (P3C).

Interestingly, three participants share the concern that people other than themselves and the recipient, could access their messages – even though they rated the statement correctly as true. One also mentioned the possibility that messages would be cached on provider servers; the other two were **generally sceptical**, either describing having no way to actually track whether the service provider encrypts the message, or expressing a general feeling that government surveillance is taking place and thus their messages are being read.

4.3 Beliefs about STMT2: "Other people can send a message pretending to be you" (functionality; false)

When evaluating Statement 2, half of the participants distinguished between services based on registration by phone number (e.g., WhatsApp) and those based on a user profile (e.g., Facebook Messenger). One reason that Statement 2 was rated as false was that participants assumed that with messengers that **link their account to their phone number**, it would not be easy for another person to send messages on their behalf.

One participant had **never heard** of the possibility of this type of abuse in their circle of friends. Several participants assumed that no other person could send messages on their behalf, because they believed that no one but themselves could **log into their account while they were logged in**. They would also notice if someone else accessed their account, as P4C states: *"My mobile number can only be used on my single device and (...) I get a message if someone wants to log in with my number."*

On the other hand, participants mentioned several reasons for rating Statement 2 as true. Four persons argued that for profile-based services, it would be easy to send messages in a person's name by creating a **fake profile** simulating the identity. P2C stated: *"It's quite easy to set up an online profile, especially on Facebook, where you simply steal a photo from another person, enter their name, and then write to people."*

However, three participants believed it also possible to **fake a phone number**, and argued that there might be some technical way to trick the servers to pretend that you are actually writing from another device. P2SL argued: *"it could also be that someone is using a program that makes the other cell phone think it has my phone number."* Moreover, several participants stated that another person could write messages on their behalf if they have **access to their login information or device**. And several participants mentioned their own experiences or the experiences of friends whose accounts had been hacked, or **media coverage** of IT security breaches as part of their reasoning. P1C stated: *"Because this has actually happened to me before (...) someone hacked my Facebook account (...) and wrote to people via my account. So, I'm 100% sure that it's possible."*

4.4 Beliefs about STMT3: "Only you and the recipient can know the messages were sent" (non-functionality; false)

Initially, only two participants believed Statement 3 was true, with low confidence. One participant discussed the differences between one-to-one and group chats, and eventually argued that in the case of private messages, only the sender and recipient could see that the messages were sent, not any other users. The other participant was not confident about this statement, made a gut decision – and came to the opposite conclusion afterwards.

We identified several reasons why participants did not believe the statement to be true: The most frequently mentioned one was that the **app provider must know this information** (n = 6), as this is *"simply logical"* (P1C) and necessary from a technical point of view. As P1TH puts it: *"they must be able to check their software or their service, and for this they must know whether a message has been successfully transmitted."*

Participants who had already assumed for Statement 1 that the provider can read messages repeated their conviction here (e.g., P1SL states: *"So I think, as already said, that others who have access to these servers or the messenger, can also see that, because they also have access"*), while other participants distinguished between the possibility of reading message content, and seeing that messages had been sent (e.g., P3SL stated: *"the messages are encrypted, but I think it is still visible in the back-end whether there were encrypted messages or not"*).

Two participants believed that **other people on the network**, such as the service provider could learn that messages had been sent, one mentioned **hacker**, and two mentioned the possibility that in group chats, of course, the **other users** would hear about the messages.

Interestingly, two participants based their assessment on the fact that messages are displayed with **time stamps or read receipts**. Since this information must come from somewhere or be stored somewhere, they believe that someone else knows that they have sent messages. Moreover, one participant argued that the statement was false because one cannot always know whether messages were sent, since messages sometimes may not arrive at their intended destination.

4.5 Beliefs about STMT4: "If somebody hacks your phone, they will be able to read your messages" (non-functionality; true)

All participants assessed Statement 4 correctly – nobody believed that their messages were protected if their phone was hacked. Many described that **access to messages on the phone is not additionally secured**, and that if someone gains access to their device, the messages, like everything else on their phones, would be visible. For example, P4C stated *"So, if someone does it right, he would have access to everything. For example, if he can see my screen, then he can see my messages. Everything, the whole thing,"* and P2SL stated *"then the phone is open and when you click on the messenger there is no further authentication."*

In addition, participants speculated about different **access mechanisms** hackers could use to access their phones and messages. Five participants described that someone could find out their PIN or password. Two participants thought that access to the messages is definitely possible with physical access to the hardware. One participant described the possibility that a virus could be loaded onto the smartphone, and another that the smartphone could be tricked into transferring messages to another device.

Four participants said that Statement 4 is true because it describes *"the epitome of hacking"* (P3TH). It is possible that the wording of the statement, which included **the term hacking** has triggered some participants in a way that they immediately believe the statement to be true (e.g., *"because hackers can do anything,"* P4SL).

4.6 Perception and influence of metaphors

In the following sections we describe how functional metaphors for E2EE were perceived, how they influenced participants' assessments of the security properties of E2EE and properties that E2EE does not provide, and what suggestions they made for improving metaphors for E2EE.

4.6.1 *Impression of metaphors.*

Special Language understood as E2EE, Treasure Hunt as access control. We asked participants how they understood the metaphor. All of those who saw the *Special Language* metaphor said that the text describes a form of encryption that ensures that messages are readable only on the end devices, at the sender and receiver (i.e., describing E2EE). For Example, P2SL explained: *"So it is a somewhat complicated description for encryption. Namely, the data that triggers such a message is encrypted, i.e. converted into another form of data, and with the right key, described here as the matching dictionary at the receiver, this data is converted back into the original language so that the receiver can read it. So, this is an attempt to secure the whole thing. If a third party now tries to intercept this data stream, it would not be readable because they don't have the right data key."*

Interestingly, only one participant who saw the *Treasure Hunt* metaphor mentioned the term encryption. In this group, everyone associated the metaphor with a description for access control, like P3TH who stated: *"According to the description, it would actually be that a message (…) is really well secured. That means that no one else has the ability to access that message in any way, because (…) we would be the only ones who - here it's described as a map -, who would have the access to find and read those messages."*

In the *Colours* group, perceptions were mixed: two participants described access control, one described E2EE, and one described it as *"a kind of encryption (… that) ensures that messages or calls cannot be read or intercepted by other people,"* (P3C). This description was less precise than descriptions of others characterizing E2EE because it did not explicitly state that messages were readable only at endpoints.

4.6.2 *Influences of metaphors on functional understanding of E2EE.*

Confidentiality conveyed but not authenticity. With regard to the first statement ("Only you and the recipient can read your messages"), nearly all participants found that the metaphor clearly described this functionality (n = 10). The three people who changed their evaluation of the statement from "false" to "true" also did so, because the metaphor suggested it. For example, P1SL, who initially

Table 2: Codes for participants' reasons for judging the statements as true or false

Reasons to believe that...	**n**	**Reasons to not believe that...**	**n**
"Only you and the recipient can read your messages" (true)		**"Only you and the recipient can read your messages" (true)**	
Trust in the provider	4	Provider can access	6
Trust in the supervision of provider	2	Doubts/no proof	2
Requires effort	2	Hackers can access	1
Receiver can read (no threats mentioned)	1		
"Other people can send a message pretending to be you" (false)		**"Other people can send a message pretending to be you"** (false)	
With password/login or device	4	Phone number linked to account	4
Profiles easy to clone	4	Only one login possible at a time	3
Experiences	4	No experiences	1
Technical possible to fake phone number	3		
Media reports	1		
"Only you and the recipient can know the messages were sent" (false)		**"Only you and the recipient can know the messages were sent" (false)**	
Other users cannot see	1	App provider	6
Feeling (but unsure)	1	Network	2
		Display of message status	2
		Other group chat members	2
		Hackers	1
		Not always known whether messages were sent	1
"If somebody hacks your phone, they will be able to read your messages" (true)		**"If somebody hacks your phone, they will be able to read your messages" (true)**	
Messages not protected on the phone	9	-	-
Access mechanisms	7		
Hackers	4		

rated Statement 1 as false, stated: *"Well, with the description (...) I would say that the phone of the intermediary can't decode it. So, if you can't decrypt that, then I would say that's 'true'. Then he can read that, but he can't understand it."*

However, the two participants who still rated the statement as false as well as several of the persons who actually rated the statement as true mentioned trust issues, i.e., they believed it questionable whether the messenger services provided this kind of security, a belief also described in [10]. P1C, who switched to true stated: *"It would be really hard for me to make a decision. Because there is still no proof that really only me and my chat partner know the secret colour and nobody dials in there somehow. But if I had to choose, I would switch to true because the blocking of the colour is built in."* P4TH, who still judged the statement to be false after the metaphor, said: *"I could also write 'the sun is always green.' Okay, it says so, but I don't have any reliable information that this is really the case."* Likewise, P1C, who also stayed with 'false' said: *"maybe it's really meant by the messenger services to be like what is stated in the purple box (referring to the metaphor description), but I don't believe it, if I'm completely honest."* Hence, they understood the essence of the functionality suggested by the metaphors (confidentiality), but did not trust its implementation by the services.

Assessing the second statement was difficult for the participants – even after the metaphors. However, for two people who saw the *Colours* metaphor, the metaphor seemed to have conveyed authenticity as they switched from an incorrect to a correct judgement – as P2C stated: *"So if I had to specify a key colour there, it would obviously make it much less likely that it's the wrong person."*

The remaining participants kept their incorrect assessment (some repeated their previously stated reasons for assessing the statement as true, such as that someone could obtain their password, create a clone profile, or have the technical capability and hacking skills) or switched from a correct to an incorrect assessment. The three participants (P3SL, P2TH, P4C) who switched to an incorrect assessment argued that the metaphor description did not exclude the possibility that someone could figure out the secret (colour, special code, or location where the treasure is hidden). P3SL argued: *"So just from the description, actually any other person, as long as they manage to get hold of the dictionary, could very well send messages and also pretend to be the person."* P3SL further states that the metaphor description lack information on this security property (referring to Statement 2): *"Because in this statement there is actually nothing about how the identification works, but only that it needs a dictionary."* P2TH thought based on the *Treasure Hunt* metaphor that another person could get to the place where the messages are hidden by observing the communication partners – as he explained: *"So, with the map, I know it well and the friend also knows about it, but the thing is, it was not specified if there is another person around us here. That's why I had changed my statement."* And P4C reasoned that *"If by some great coincidence the person guesses the colour, then of course they can pretend to be me. Therefore, I would change my answer here and guess 'true'."*

Conceptual baggage. While most participants assessed the limitations of E2EE correctly, and did not assume additional security properties that E2EE does not provide (security of meta data or end point devices, as addressed in Statement 3 and 4), we note two exceptions where the metaphor did lead participants to assume that E2EE protects meta data information (i.e., knowledge about the transmission status of messages, Statement 3). P1C and P2TH evaluated Statement 3 incorrectly after the metaphor was presented. Features implied by the metaphor but not supported by the system are called *conceptual baggage* [4], which increases the amount of processing required to adapt a metaphor to a functional model.

With P1C, the metaphor seemed to have conveyed too much security – as P1C stated: *"Because I add this other colour before I send the message, i.e. now based on the new description with the colours, I would then also go with true"*. P2TH interpreted into the metaphor that the way to the treasure must also be encrypted, by arguing: *"In the description it says that only me and the other person know the map, so on the server the messages should also be encrypted somehow, so also the way (to the treasure) should be encrypted somehow, so I assume that nobody gets the notification that the message was sent."*

4.6.3 Suggestions of participants.

No metaphor is a perfect match. Participants identified several weaknesses of metaphors. The *Special Language* metaphor was criticized for being too unspecific, as P4SL described: *"If it said 'code', that would be clearer. Language can also be a programming language or something else"*. In addition, another participant mentioned that use of a dictionary implies a manual procedure of decryption, which does not fit, because decryption is done automatically by the chat program. For the *Treasure Hunt* metaphor, one participant described that the map was less important to know than the actual location of the treasure. To improve the understanding of Statement 3, the participant suggested changing the description so that only the exact location is secret, but that the map or the route to the location can be known. With regard to the *Colour* metaphor, two participants also mentioned that the description does not adequately reflect Statement 2, i.e. that no one can send a message pretending to be you. P4C explicitly describes that the *Colour* metaphor is misleading, because it suggests that the colour palette is limited and that therefore, someone could relatively easily guess the secret colour.

Provide more 'technical' and concrete information. Several participants found the 'technical' description of E2EE suitable (n = 3), or suggested to provide a mixture of metaphorical and technical description (n = 3). Two participants (one saw the *Treasure Hunt*, one saw the *Colours* metaphor) said they needed more concrete information about the implementation of secret communication, e.g., how the secret colour is communicated, or how the two communication partners get to the treasure. Moreover, two participants (*Special Language*, *Colours*) stated that the metaphor could be improved by including information about message transmission and the process of routing. Two participants suggested to include graphical examples, one mentioned to use easy wording.

Other metaphor suggestions. Participants came up with several ideas for metaphors describing E2EE functionality: Three participants referred to a "sealed letter" or "postal package" metaphor. Two from the *Treasure Hunt* group described a portable storage box, which resembles the key/lock metaphor described in [26]. One mentioned the key metaphor, and one suggested a description of "direct communication", such as slipping someone a note (probably to signal privacy of the communication and that no one else could intercept).

5 DISCUSSION

Our results provide a number of insights about how users construct functional understanding of E2EE in the context of messaging services, and how metaphors can influence their assumptions.

Understanding users' reasoning processes is the necessary first step to help users to develop a functional mental model of E2EE that conveys the security properties of E2EE, and the limits of the protection it can offer. By better understanding how users connect what they know about encryption and digital communication to their beliefs and experiences, we can develop better measures to help them make informed security decisions.

We derive several implications for the research on user perceptions of security and privacy. In addition, our findings contribute to research on the design possibilities of pedagogical approaches to teach and promote functional mental models. Finally, we also discuss several methodological implications.

5.1 What's in the black box?

Understanding vs. believing. Demjaha et al. [11] found that user-generated functional metaphors for E2EE did not significantly improve their participants' understanding of the security properties E2EE provides and those it does not provide. Our qualitative analysis of the interviews provides a number of insights that explain why. Several participants in our study understood the security properties of E2EE correctly (at least confidentiality) – but did not trust them to be implemented correctly. Not *believing* in the motivation or ability of service providers is a different issue than not *understanding* the security properties of E2EE.

We think that it is important to distinguish between different components that influence the formation of a mental model. It is necessary, but not sufficient to have a functional understanding of the security properties of E2EE to develop a correct functional mental model. Beliefs about the behaviour of other actors or technologies involved – e.g., about the motives and capabilities of service providers – can be stronger and "override" the functional understanding. We saw that these beliefs have a major impact on users' security decisions. People tend to hold on to beliefs or find arguments that favour conclusions they want to believe in (a phenomenon called motivated reasoning [19]), e.g., that they do not need to bother with encryption because vendors (can) do what they want anyway. Another example in IT security is that incorrect threat models lead people not to take precautions they would be capable of doing, but do not think warrant the effort. *"I'm not a target because I'm not rich or famous"* first reported 20 years ago [27], and still heard in conversations with users today.

Authenticity is difficult to grasp. Participants mainly had difficulties evaluating the statement about the security property of authenticity (STMT2: "Other people can send a message pretending to be you"). The thought processes of the participants when evaluating the statement clearly show that no one attributed this

security property to E2EE or associated it with E2EE. Many participants shared the view that authenticity is determined solely by the protection of the end device, the credentials, and the password, and not by any security vulnerabilities in the transmission, confirming the results of previous studies [2, 5]. Bai et al. [5] concluded that conveying the property of authenticity remains challenging, as their participants did not regard it as important, or absorbed it into their models of confidentiality. Hence, they recommended to not try to explain authenticity independently from confidentiality. Our participants suggested more information about authenticity should be included in the metaphors, which mainly conveyed confidentiality.

The question, however, is how providing information about authenticity can successfully contribute to improving a functional mental model of E2EE. Literature from cognitive psychology and the learning sciences suggests that it is not always effective to provide more knowledge to fill knowledge gaps – sometimes, learning requires correcting existing misconceptions – which is called conceptual change [6]. Psychologist Jean Piaget has distinguished between two processes in cognitive development: assimilation and accommodation [21]. Assimilation involves inserting information into pre-existing knowledge structures, while accommodation involves restructuring and changing knowledge structures to integrate the information. If users believe that controlling access to their device and keeping their credentials private is sufficient to achieve authentication, this is a misconception. The challenge will be to find ways to accommodate existing knowledge structures and modify them as we integrate new information, if we want to convey correct assumptions about how authentication is established.

Absence of threat models. When evaluating security properties of the communication with messenger services, participants distinguished between messengers that are based on a phone-number registration and those based on user profiles. In line with the findings by Abu-Salma et al. [2], linking a phone number to an account was perceived by some participants as a security guarantee; they believed that it ensured that another person could not simply impersonate them and communicate as them (i.e., via their account). This focus on access control by messenger services shows that participants do not have clear or readily available mental models of threats, such as a machine-in-the-middle attack (see also [23]). Research suggests that without awareness and understanding of the security and privacy threats that E2EE can protect against, users will not consider the tools available to them for E2E encrypted communications [1, 2]. Therefore, it is not only a challenge to correct previously misconceived knowledge, but also to assimilate missing knowledge about cybersecurity threats sufficient for a functional understanding of E2EE.

Factors that influence users' security assessments. Our interviews show that users base their assessments of the security properties of E2EE in the context of messenger services on a variety of factors. In addition to their knowledge and assumptions about technical aspects of digital communication, they include their beliefs about the motives and capabilities of communications providers, their own experience or experiences of family, friends, and acquaintances with cybersecurity threats, news reports, design features of apps (e.g., read receipt), and their stereotypes about hackers. Hence, when we think about ways to improve users' functional mental models, it is necessary to consider all of these potentially influencing factors when designing interventions.

5.2 Recommendations

This study contributes to the larger body of research looking for effective ways of communicating security knowledge to non-expert users, and can make some recommendations in this regard.

Correcting misconceptions. Users' heads are not empty vessels, waiting to be filled with relevant security knowledge. When new knowledge is learned, it is assimilated into existing knowledge structures [21]. These existing structures can also include misconceptions. Hence, to integrate 'correct' security information, users need to initiate accommodation processes, i.e., a change or even replacement of old assumptions based on new information. We propose that we need to debunk and correct those false beliefs and misconceptions first. Approaches from research on debunking misinformation could be helpful with this endeavour [20].

Metaphors are powerful, but not all-powerful. On the one hand, we believe that functional metaphors have the great advantage of starting with users' existing working models, and tasks they are familiar with. Unlike expert knowledge often presented in training materials, they are not overwhelming, but quick and easy to access. This is an advantage, as understanding benefits and limitations of encryption is not users' primary goal, and, building on users' functional mental models – in contrast to teaching a structural model [13, 31] – seems the promising way forward. Moreover, regarding the risk that metaphors might suggest more security than actually offered as a result of their 'conceptual baggage' [4], our study tends to indicate this is a rather low risk, as only two participants rated a security property that E2EE does not provide as true based on the metaphor. However, this would need to be tested in a larger study.

On the other hand, our results confirm what previous research has suggested: finding an appropriate metaphor to evoke a functional mental model of E2EE is not straightforward [11]. At the very least, it seems difficult to find a helpful metaphor when it is sought with the goal of integrating new knowledge. It seems plausible that metaphors, like any other form of intervention or training, can only work if they take the approach of correcting misconceptions and accompany the process of changing mental models, as it is much more difficult to remove and correct false beliefs than to teach new ones. Therefore, we need new forms of interventions that accompany mental model changes.

Trust-building measures. In addition to addressing users' understanding, communications must address trust. It could be helpful if users had a way to better assess whether a service provider is competent and trustworthy. For example, information about why open source code is more trustworthy (not less, as some users currently assume [2]), assessments by independent security experts, or a history of known security/privacy breaches could be ways forward. As with other trust signals, even though many users do not follow up on this information, they are reassured by the fact that they could [17].

Concrete but simple. In line with Distler et al. [14] who showed that app descriptions using the word "secure" or "encrypt" made users feel more secure than the metaphorical description "translating to secret code", we also find that some participants prefer to call encryption by its name, rather than a "flowery euphemism". Moreover, and in line with Bai et al [5], our participants value accessible wording in educational material.

5.3 Methodological implications

Our research highlights some methodological issues that should be considered within the field of usable security when aiming to test the effectiveness of communication strategies on user understanding quantitatively. It is important to not only consider users' understanding of the communication strategy, but also of how the outcome measures or test items are understood. Users might misinterpret a survey question trying to measure their understanding in a way not intended by the researchers. In our interviews, several participants stated that the statement about the functionality of authentication was not precise enough, as the wording did not explicitly exclude the possibility of physical access to the end device. In addition, 'trigger words', such as the term *hacking* in Statement 4 (see Section 4.5), can influence response behaviour.

6 LIMITATIONS AND FUTURE WORK

Our study was conducted with a small sample that is not representative of the larger population. Our participants are younger and more educated than average. However, we think that if we find misconceptions even in this group, we can also learn something for understanding an older and probably less technology-savvy population. We believe that our results provide important insights into users' understanding processes and have shown something relevant: If we want to achieve that users develop correct functional mental models of E2EE, we need to focus on correcting misconceptions and finding ways to guide the transformation process of mental models.

As any interview study, we cannot exclude the possibility of demand characteristics of the interview situation. Participants might have attempted to answer questions in a way they believed the researcher wanted them to answer. However, we believe that this concern is rather low, because we experienced quite frank conversations, as some participants actually speculated about the correct answers to the quiz questions but argued to stick to their own opinion as to what they believed was the correct answer we wanted to achieve with the metaphors.

Although we have already asked in-depth qualitative questions to identify the reasons behind users' security assessments, we believe it is worth drilling deeper to identify the underlying beliefs even more clearly. Future research could investigate why it is believed that the app provider can read its users' messages by asking even more targeted questions and follow-up questions about emerging justifications.

Another limitation is that we explicitly did not mention E2EE, but asked users to imagine the communication with a messenger that offers E2EE. It is plausible that users have prior attitudes towards different services. An alternative would have been to make it more explicit that we were interested in their understanding of the security properties of E2EE, but then we would not have been able to obtain their unbiased reaction toward the metaphors and could not have explored whether users associate the metaphors with E2EE.

A further limitation is the choice of the security properties covered in the four statements. In addition to confidentiality and authenticity, we have not covered integrity separately.

7 CONCLUSIONS

The goal of this study was to examine in more detail how users arrive at their judgements about the security properties of E2EE and the limitations of its protection in the context of instant messenger communication, and how functional metaphors influence their reasoning. Behind this is the ultimate goal of finding ways to build on users' existing mental models to achieve a functional understanding of E2EE. To this end, we conducted twelve semi-structured interviews, exposing participants to different functional metaphors for E2EE.

Our results show that understanding about the core functionality of E2EE, ensuring confidentiality of messages between two communication partners, can be improved by functional metaphors of E2EE. However, some persistent beliefs concerning trust in communication providers remain. The authenticity of communication as another functionality of E2EE remains more difficult to convey by metaphors. Many participants share the view that authenticity is determined solely by the protection of the end device, the credentials, and the password, and not by any security vulnerabilities in the transmission. The metaphors studied so far have not been able to close this gap in understanding. We suggest that instead of searching for more and better metaphors, it might be more purposeful to find interventions that target the process of changing mental models and correct persistent misconceptions.

ACKNOWLEDGMENTS

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2092 CaSa – 390781972.

REFERENCES

[1] Ruba Abu-Salma, Elissa M. Redmiles, Blase Ur, and Miranda Wei. 2018. Exploring user mental models of end-to-end encrypted communication tools. In *8th USENIX Workshop on Free and Open Communications on the Internet (FOCI 18)*. USENIX Association, Baltimore, MD. https://www.usenix.org/conference/foci18/presentation/abu-salma

[2] Ruba Abu-Salma, M. Angela Sasse, Joseph Bonneau, Anastasia Danilova, Alena Naiakshina, and Matthew Smith. 2017. Obstacles to the adoption of secure communication tools. In *2017 IEEE Symposium on Security and Privacy (SP)*. IEEE, 137–153. https://doi.org/10.1109/SP.2017.65

[3] Omer Akgul, Wei Bai, Shruti Das, and Michelle L. Mazurek. 2021. Evaluating in-workflow messages for improving mental models of end-to-end encryption. In *30th USENIX Security Symposium (USENIX Security 21)*. USENIX Association, 447–464. https://www.usenix.org/conference/usenixsecurity21/presentation/akgul

[4] Ben Anderson, Michael Smyth, Roger P. Knott, Marius Bergan, Julie Bergan, and James L. Alty. 1994. Minimising conceptual baggage: making choices about metaphor. In *BCS HCI '94: British Computer Society Conference on Human-Computer Interaction*. Cambridge University Press, New York, NY, 179–194.

[5] Wei Bai, Michael Pearson, Patrick G. Kelley, and Michelle L. Mazurek. 2020. Improving non-experts' understanding of end-to-end encryption: An exploratory study. In *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 210–219. https://doi.org/10.1109/EuroSPW51379.2020.00036

[6] Michelene T. H. Chi. 2008. Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In *International Handbook of*

Research on Conceptual Change, Stella Vosniadou (Ed.). Routledge, New York, NY, 61–82.
[7] Allan Collins and Dedre Gentner. 1987. How people construct mental models. In *Cultural Models in Language and Thought*, Dorothy Holland Herring (Ed.). Vol. 243. Cambridge University Press, 243–265.
[8] Juliet M. Corbin and Anselm Strauss. 1990. Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology* 13, 1 (1990), 3–21. https://doi.org/10.1007/BF00988593
[9] Alexander de Luca, Sauvik Das, Martin Ortlieb, Iulia Ion, and Ben Laurie. 2016. Expert and non-expert attitudes towards (secure) instant messaging. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*. USENIX Association, Denver, CO, 147–157.
[10] Sergej Dechand, Alena Naiakshina, Anastasia Danilova, and Matthew Smith. 2019. In encryption we don't trust: The effect of end-to-end encryption to the masses on user perception. In *2019 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE, 401–415. https://doi.org/10.1109/EuroSP.2019.00037
[11] Albese Demjaha, Jonathan Spring, Ingolf Becker, Simon Parkin, and Angela Sasse. 2018. Metaphors considered harmful? An exploratory study of the effectiveness of functional metaphors for end-to-end encryption. In *2018 Workshop on Usable Security*, Yasemin Acar and Sameer Patil (Eds.). https://doi.org/10.14722/usec.2018.23015
[12] Deutsches Forschungsnetz. 2020. DFNconf – Der Konferenzdienst im Deutschen Forschungsnetz. https://www.conf.dfn.de/
[13] Andrea diSessa. 1986. Models of computation. In *User Centered System Design New Perspectives on Human-Computer Interaction*, A. Norman Donald and W. Draper Stephen (Eds.). Lawrence Erlbaum, Hillsdale, NJ, 201–218.
[14] Verena Distler, Carine Lallemand, and Vincent Koenig. 2020. Making encryption feel secure: Investigating how descriptions of encryption impact perceived security. In *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 220–229.
[15] Nina Gerber, Verena Zimmermann, Birgit Henhapl, Sinem Emeröz, and Melanie Volkamer. 2018. Finally Johnny Can Encrypt. In *Proceedings of the 13th International Conference on Availability, Reliability and Security*. ACM, New York, NY, 1–10. https://doi.org/10.1145/3230833.3230859
[16] Amir Herzberg and Hemi Leibowitz. 2016. Can Johnny finally encrypt?. In *Proceedings of the 6th Workshop on Socio-Technical Aspects in Security and Trust*, Gabriele Lenzini, Giampaolo Bella, Zinaida Benenson, and Carrie Gates (Eds.). ACM, New York, NY, 17–28. https://doi.org/10.1145/3046055.3046059
[17] Iacovos Kirlappos, M. Angela Sasse, and Nigel Harvey. 2012. Why trust seals don't work: A study of user perceptions and behavior. In *Trust and Trustworthy Computing (Lecture notes in computer science)*, Stefan Katzenbeisser (Ed.). Springer, Heidelberg, 308–324. https://doi.org/10.1007/978-3-642-30921-2_18
[18] Katharina Krombholz, Karoline Busse, Katharina Pfeffer, Matthew Smith, and Emanuel von Zezschwitz. 2019. "If HTTPS were secure, I wouldn't need 2FA" - End user and administrator mental models of HTTPS. In *2019 IEEE Symposium on Security and Privacy (SP)*. IEEE, 246–263. https://doi.org/10.1109/SP.2019.00060
[19] Ziva Kunda. 1990. The case for motivated reasoning. *Psychological bulletin* 108, 3 (1990), 480–498. https://doi.org/10.1037/0033-2909.108.3.480
[20] Stephan Lewandowsky, John Cook, and Doug Lombardi. 2020. Debunking Handbook 2020. https://doi.org/10.17910/B7.1182
[21] Saul Mcleod. December 07, 2020. Jean Piaget's Theory and Stages of Cognitive Development: Background and Key Concepts of Piaget's Theory. https://www.simplypsychology.org/piaget.html
[22] Karen Renaud, Melanie Volkamer, and Arne Renkema-Padmos. 2014. Why doesn't Jane protect her privacy? In *Privacy Enhancing Technologies*, Emiliano de Cristofaro and Steven J. Murdoch (Eds.). Springer International Publishing, Cham, 244–262. https://doi.org/10.1007/978-3-319-08506-7_13
[23] Svenja Schröder, Markus Huber, David Wind, and Christoph Rottermanner. 2016. When SIGNAL hits the fan: On the usability and security of state-of-the-art secure mobile messaging. In *Proceedings 1st European Workshop on Usable Security*, Karen Renaud and Melanie Volkamer (Eds.). Internet Society, Reston, VA. https://doi.org/10.14722/eurousec.2016.23012
[24] Robert. W. Shirey. 2007. Internet Security Glossary, Version 2: RFC: 4949. https://datatracker.ietf.org/doc/html/rfc4949
[25] Nancy Staggers and Anthony F. Norcio. 1993. Mental models: concepts for human-computer interaction research. *International Journal of Man-machine studies* 38, 4 (1993), 587–605. https://doi.org/10.1006/imms.1993.1028
[26] Wenley Tong, Sebastian Gold, Samuel Gichohi, Mihai Roman, and Jonathan Frankle. 2014. Why King George III can encrypt. https://freedom-to-tinker.com/2014/06/06/why-king-george-iii-can-encrypt/
[27] Dirk Weirich and Martina Angela Sasse. 2001. Pretty good persuasion: a first step towards effective password security in the real world. In *Proceedings of the 2001 workshop on New security paradigms*, Victor Raskin, Steven J. Greenwald, Brenda Timmerman, and Darrell Kienzle (Eds.). ACM, New York, NY, 137–143. https://doi.org/10.1145/508171.508195
[28] Alma Whitten. 2004. *Making Security Usable*. Ph.D. Dissertation. Carnegie Mellon University, Pittsburgh, PA. http://reports-archive.adm.cs.cmu.edu/anon/anon/usr/ftp/usr0/ftp/2004/CMU-CS-04-135.pdf
[29] Alma Whitten and J. Doug Tygar. 1999. Why Johnny can't encrypt: A usability evaluation of PGP 5.0. In *8th USENIX Security Symposium (USENIX Security '99)*, USENIX Association (Ed.). USENIX Association, Washington, D.C., 169–184.
[30] Justin Wu and Daniel Zappala. 2018. When is a tree really a truck? exploring mental models of encryption. In *Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018)*. USENIX Association, Baltimore, MD, 395–409.
[31] Richard M. Young. 1983. Surrogates and mappings: Two kinds of conceptual models for interactive devices. In *Mental Models*, Dedre Gentner and Stevens A.L. (Eds.). Lawrence Erlbaum Associates, Hillsdale, NJ, 35–52.

A APPENDIX

A.1 Interview Guide

The following sections include the questions and instructions used in our interviews.

Introduction

[Welcome, introduction of researcher and study aim, asking for consent of recording, giving opportunity to ask questions.]

Before starting the Interview, I would like to emphasize, that I am interested in your personal understanding, your thoughts, and opinions. Therefore, I would like to ask you to answer the questions as openly and honestly as possible. Participation in this interview is voluntarily and you are free to end the interview any time and without giving any reason. Thanks a lot for your support. If you have any further questions, please feel free to ask.

Warm-up

To exchange messages with another person, there are now a variety of online messenger services.

(1) Which messenger services do you have installed on your smartphone?
(2) Which messenger service do you use most frequently? How frequently do you use it? [Only if a non-E2EE tool is mentioned: Do you use WhatsApp, iMessage, Facebook Messenger, Signal or Telegram?]
(3) Please imagine that you are sending a private message to a person using a messenger app like WhatsApp, iMessage, Signal, Telegram, Facebook Messenger. Why is sending messages with one of these messengers secure or insecure? [Why is it secure? What about the communication is secure? Why is it insecure or what about the communication is insecure?]

Quiz questions about the (non)security properties of E2EE (based on Demjaha et al., 2018)

In the following, I will show you some statements that you should rate as 'true' or 'false'. Concerning the following statements, please imagine that you are using one of the following messenger services: WhatsApp, Signal, Telegram, iMessage or Facebook Messenger.

[Go through the following 4 statements one by one with the participant and note their choice of 'true' or 'false'.]

> **STMT1:** "Only you and the recipient can read your messages"
> **STMT2:** "Other people can send a message pretending to be you"
> **STMT3:** "Only you and the recipient can know the messages were sent"
> **STMT4:** "If somebody hacks your phone, they will be able to read your messages"

[Ask the following 3 questions for each statement:]

(1) Please choose: Is the statement true or false.
(2) Why do you think the statement is true or false?
(3) How sure are you about your evaluation, and why?

Perception of metaphors & reflection of quiz questions

Please further imagine that you are using one of the following messenger services: WhatsApp, Signal, Telegram, iMessage or Facebook Messenger. Please read the following description about the communication with these messengers. Afterwards, I will ask you some questions about it. [1 of the following 3 metaphors is shown on screen:]

> **Special Language (SL):** "Messages and calls with this person will be translated to a special language for which only the two of you know the dictionary."
> **Treasure Hunt (TH):** "Messages and calls exchanged with this person are like a treasure hidden in a place to which only the two of you know the map."
> **Colours (C):** "Messages and calls you exchange with this person are like colours. Before sending them, you mix them with another colour, known only by you two. Nobody else can retrieve them unless they know the secret colour."

(1) How do you understand the description? Please state in your own words what this description says about the communication.
(2) Is there anything unclear about the description? If so, what?

I would like to ask you, to take another look at the 4 statements to see whether you would confirm or change your evaluation based on the description. [Go through each statement in turn; allow time for reflection.]

For each statement:

(1) Would you change or confirm your initial evaluation? Why?

Solution to the quiz

In the following, I would like to show you the solutions to the true/false statements you evaluated with regard to the communication with messenger services such as WhatsApp, Signal, Telegram, iMessage or Facebook Messenger. These messenger services secure their messages and calls with end-to-end-encryption. The description you have just read about the communication of these services intends to explain the principle of end-to-end-encryption. To explain end-to-end-encryption to you once more, I would like to show you another explanation of end-to-end-encryption:

> **'Technical' description of E2EE**: "Continuous protection of data that flows between two points in a network, effected by encrypting data when it leaves its source, keeping it encrypted while it passes through any intermediate computers (such as routers), and decrypting it only when it arrives at the intended final destination" (Shirey, 2007, p. 121).

(1) Do you understand the description, or do you have any questions?

Now let us go back to the four statements that I showed you previously and your answers.

For each statement: For statement 1 [2, 3, 4], after the second round, you said that it is true/false. This is correct/incorrect [depending on the participant's answer].

Answer scripts and questions about incorrectly answered statements:

(1) STMT1: "Only you and the recipient can read your messages" (functionality; true).
E2EE encrypts the content of a message on the sender's device before it is sent so that only the sender and recipient can decrypt and read it. The encryption means that no one can read the message in transit.
IF the statement was answered incorrectly: What would an explanation have to look like to make this (non-)functionality understandable?
(2) STMT2: "Other people can send a message pretending to be you" (functionality, false).
Without full access to your smartphone, E2EE prevents others from pretending to be you and sending messages on your behalf.
IF the statement was answered incorrectly: What would an explanation have to look like to make this (non-) functionality understandable?
(3) STMT3: "Only you and the recipient can know the messages were sent" (non-functionality, false).
The provider forwards the messages and must know to whom to forward the message. Each encrypted message is therefore provided with readable metadata, such as the desired recipient of the message. However, the contents of the messages remain secret through E2EE.
IF the statement was answered incorrectly: What would an explanation have to look like to make this (non-) functionality understandable?
(4) STMT4: "If somebody hacks your phone, they will be able to read your messages" (non-functionality, true).
If someone gains full access to your smartphone, e.g., steals it and hacks it, they can read your messages.
IF the statement was answered incorrectly: What would an explanation have to look like to make this (non-) functionality understandable?

Closing question to compare the metaphor and technical description

(1) Now, if you look again at the explanation of E2EE and the metaphorical description of E2EE that you read beforehand, do you have a suggestion on how the concept of end-to-end encryption could be described even more simply in a pictorial way?

Demographics

(1) What do you do professionally?
(2) What is your level of education?
(3) How old are you?
(4) What gender do you identify with?
(5) Do you have prior knowledge/experience in IT or IT-Security? How would you rate your knowledge?