Deniable Encrypted Messaging: User Understanding after Hands-on Social Experience

ANAMIKA RAJENDRAN, New Mexico State University, USA
TARUN KUMAR YADAV, Brigham Young University, USA
MALEK AL-JBOUR, New Mexico State University, USA
FRANCISCO MANUEL MARES SOLANO, New Mexico State University, USA
KENT SEAMONS, Brigham Young University, USA
JOSHUA REYNOLDS, New Mexico State University, USA

Plausible deniability in cryptography allows users to deny their participation in a particular communication or the contents of their messages, thereby ensuring privacy. Popular end-to-end encrypted messaging apps employ the Signal protocol, which incorporates message deniability. However, their current user interfaces only allow access to the blunt tool of message deletion. Denying a message requires users to claim that the counterpart in their conversation has the technical sophistication to forge a message when no usable message forgery tools are available. We evaluate a step towards bridging this gap in the form of a new transcript-editing feature implemented within the Signal app which allows each user to maintain an independent, locally-editable transcript of their conversation. We gave users hands-on experience with this app in the context of resolving a social dispute, and measured their ability to understand its implications both technically and ethically. Users find our interface intuitive and can reason about deniability, but are divided by which circumstances for which deniability is appropriate or desirable. We recommend users be given transparent access to choose when their conversations are deniable versus non-repudiable, instead of the status quo of somewhere-in-between. Our study introduces a novel approach by providing hands-on experience and evaluating its usability. This method offers insights into practical deniability implementation and lays the groundwork for future research.

CCS Concepts: • **Human-centered computing** → *Human computer interaction (HCI)*; • **Security and privacy** → **Usability in security and privacy**;

Additional Key Words and Phrases: Deniability, Message Forgery, Usability, User Experience, Privacy, Signal App

ACM Reference Format:
Anamika Rajendran, Tarun Kumar Yadav, Malek Al-Jbour, Francisco Manuel Mares Solano, Kent Seamons, and Joshua Reynolds. 2024. Deniable Encrypted Messaging: User Understanding after Hands-on Social Experience. In *The 2024 European Symposium on Usable Security (EuroUSEC 2024), September 30-October 1, 2024, Karlstad,Sweden.* ACM, New York, NY, USA, 26 pages. https://doi.org/10.1145/3688459.3688479

1 INTRODUCTION

End-to-end (E2E) encrypted messaging apps are used by billions of people, globally [20]. They are essential tools for privacy in communication that enable greater freedom for their users. Simultaneously, strong advocates push for

Authors' addresses: Anamika Rajendran, New Mexico State University, USA, arajen97@nmsu.edu; Tarun Kumar Yadav, Brigham Young University, USA, tarun14110@iiitd.ac.in; Malek Al-Jbour, New Mexico State University, USA, mjbour@nmsu.edu; Francisco Manuel Mares Solano, New Mexico State University, USA, manuelms@nmsu.edu; Kent Seamons, Brigham Young University, USA, seamons@cs.byu.edu; Joshua Reynolds, New Mexico State University, USA, jr1@nmsu.edu.

Manuscript submitted to ACM

holding individuals accountable for their words. When either party has plausible deniability about the conversation, it empowers journalists, whistleblowers, and activists to speak privately off-the-record. At the same time, it allows criminals to hide their activity, fraudsters to cover false promises, and perpetrators of abuse to avoid accountability.

Deniability is rightfully considered a controversial feature, but does in fact exist (to some degree) in the most popular encrypted messaging apps today because of their use of the Signal protocol [22]. Deniability enables individuals to communicate freely with less fear of repercussions. However, prior research has shown that the opposite property of deniability, non-repudiation, is how users perceive messaging transcripts [33]. The reality is that current messaging apps lie somewhere in between these two extremes. Lack of access to either extremes of this tradeoff has been the status quo, and only a few studies have explored this tradeoff from the users' perspective [25, 33]. In this exploratory work, we grant access to deniability in a working modification to the open-source Signal messaging application. With hands-on access to this app, we assess users' understanding of the technical, social, and ethical trade-offs of message deniability.

We designed and evaluated a modification to the open-source Signal messaging app [22] to include a local message history editing feature. The new tool allows participants to modify their message record locally, providing them with a firsthand deniability experience and introducing the possibility that conversation participants can manipulate or alter messages on their local devices. We also immersed users in a social hypothetical scenario, to demonstrate how this feature could impact an everyday interaction. We then invited them to explore the potential use cases of deniability in their own communication, and reason about when and where its use would be appropriate. This novel approach, combining social scenarios with hands-on deniability messaging app experience, sets a framework for future research on user perception of deniability. Our study addressed the following research questions (RQs):

RQ1 *After hands-on experience with an E2E encrypted messaging app that enables deniability in a social context, do participants comprehend the concept of deniability and can they extrapolate its potential applications?*

RQ2 *Is our transcript-editing interface a user-friendly method of giving users intuitive access to deniability in E2E encrypted messaging apps?*

RQ3 *Can users reason about when deniability is useful and/or ethical in the E2E encrypted messaging app context?*

RQ4 *Will users consider switching to/away from E2E encrypted messaging apps that implement deniability features, and why?*

Our study consisted of three phases. First, we designed a novel interface for local-only message alteration. Second, we conducted unstructured-interviews to refine our testing scenario. Lastly, we surveyed users (n=35) before and after their participation in a hypothetical scenario where they tried out the deniability tool in context.

The key takeaways from our research are: (1) Our tool and scenario were successful in introducing users to the concept of deniability in the context of E2E encrypted messaging apps. (2) On a standardized usability scale, they ranked the interface as user-friendly. (3) After using our new tool and considering the hypothetical scenario, users displayed a reasonable understanding of the system, and how it could be used or abused. They identified real-life scenarios where deniability could be useful. (4) Many participants articulated a stance on whether deniability is desirable for everyone, for nobody, or context-specific.

Some users appreciated deniability as it empowers freedom of speech, enabling communication without fear of repercussions. Other users expressed a desire for the opposite property, non-repudiation, which holds people accountable for every message they send. Our study has resulted in an interesting and novel finding: Users generally reported that having deniability tools at their disposal would not prompt them to change their current messaging app. Our findings suggest that deniability is understandable for average users with minimal training and a straightforward interface.

Our findings also suggest that users would appreciate the ability, based on context, to enforce either deniability or non-repudiation in their conversations.

2 BACKGROUND AND RELATED WORK

In recent years, several popular messaging apps have adopted **end-to-end encryption (E2E)**, either by default or as an optional feature [3, 14]. E2E ensures that only the sender and the intended recipient(s) can access the data/messages, keeping them unreadable to eavesdroppers. However, E2E is distinct from deniability, which allows users to plausibly deny the content or authorship of a message.

Various research efforts have explored cryptographic deniability. The Off-the-record (OTR) protocol [7] introduced deniability using Deniable Authenticated Key Exchange (DAKE) [12]. According to OTR [7], deniability means participants can deny the authenticity or authorship of specific messages, even if the content is captured or leaked. The sender can deny sending the message, and the receiver can deny receiving or acknowledging it.

Deniability has also been analyzed in contexts like Signal [22, 31] and group messaging [28]. Schliep and Hopper [28] achieved deniability by allowing participants to deny involvement in a conversation using a simulator to produce indistinguishable simulated transcripts. Unger and Goldberg [30] further strengthened deniability in secure messaging apps like instant messaging (IM) apps, emphasizing the need for secure and deniable messaging protocols due to current tools' lack of security and privacy. Additionally, research has explored deniability in file systems [18], anonymous communication [19], document recommender systems [11], and privacy-preserving data synthesis [6].

Non-expert Perspective Study - Legal Context: Reitinger et al. [25] surveyed 1200 participants to explore perceptions of deniability in encrypted communication protocols using a legal scenario. Participants imagined themselves as jury members in a bribery case involving a politician and were presented with a screenshot of a politician allegedly requesting a bribe. They used a forgery tool to edit screenshots and metadata, providing a practical understanding of deniability. The findings revealed mixed acceptance: some participants became more accepting of deniability when introduced to the concept, while others continued to trust manipulated evidence. Those informed about the forgery tool were more likely to accept deniability, suggesting practical tools can enhance its effectiveness and perception in messaging applications. The study emphasized the need for further research in non-courtroom settings to understand the benefits and risks of deniability and explore its potential to strengthen communication privacy and trust.

Multi-perspective Study: Yadav et al. [33] conducted a mixed multi-perspective study to explore user understanding, perception, and expectations of cryptographic deniability in everyday communication, unlike Reitinger et al. [25] which focused on political use cases. The study examined how factors like the relationship between the claimant and recipient and the medium of communication (e.g., chat screenshots or in-app messages) affect perceptions of deniability. Through analyzing expert email conversations, interviews, and surveying 664 participants, it explored the impact of relationships and communication channels. Legal perspectives were also considered through court case studies.

Findings showed 23.18% of participants desired deniability, with experts recommending user-friendly interfaces to raise awareness and address the perception that message forgery is difficult. Despite a high demand for non-repudiation, 12.7% of users still wanted deniability. Notably, only 0.6% of participants understood the OTR definition of deniability, highlighting the need for hands-on experiences to improve user understanding and informed decision-making.

Aspects	**Reitinger et al.** [25]	**Yadav et al.** [33]	**Collins et al. [10]**	**Our Study**
Case Study	✗	✓	✓	✗
Deniability Teaching Tools	Screenshots & desktop forgery app	verbal explanation & OTR definition	N/A	**In-app deniability feature**
Hands-on Experience Changing Messages	✗	✗	✗	✓
Interview	✓	✓	✗	✓
Survey	✓	✓	✗	✓
Context	Legal	Legal and Social	N/A	Social
Identified User Preferences	✓	✓	✗	✓

Table 1. This table highlights both similarities and differences with previous research. Notably, our study introduces novel hands-on interaction methods. A check-mark (✓) indicates that the respective aspect was assessed in the study, while a cross-mark (✗) indicates that it was not.

Other Related Work: Collins et al. [10] explored various aspects of cryptographic deniability using the Signal Protocol [22, 31], focusing on its practical implications for users and courtroom scenarios. Their discussion highlighted the features and mechanisms of the cryptographic deniability model, emphasizing its ability to provide plausible deniability in messaging conversations. They found that judges, unaware of cryptographic deniability, dismissed the possibility of message forgery in messaging apps. Similarly, users believed message forgery was challenging, making deniability seem unattainable. Collins et al. [10] proposed adding an edit/add button to Signal App for modifying or adding messages, highlighting the need for a practical user-friendly simulator to achieve plausible deniability. This points to potential future advancements in cryptographic deniability for messaging apps.

Our study shares similarities with previous research, as shown in Table 1. For example, Reitinger et al. [25] used legal scenarios and a before-and-after study design with chat screenshots to explore changes in users' perceptions with a forgery tool. Similarly, we used scenario to assess participants' understanding before and after introducing deniability features. We also used pilot interviews and surveys, methods consistent with Reitinger et al. [25] and Yadav et al. [33].

However, our study differs significantly from prior research. While Reitinger et al. [25] focused on courtroom scenarios, we chose a more relatable social scenario, similar to Yadav et al. [33], recognizing that not everyone has legal expertise. Reitinger et al. [25] highlighted the impact of hands-on experience on user understanding. Building on Yadav et al.'s approach [33], which encouraged users to consider real-life deniability applications, we presented concrete use cases to stimulate participant brainstorming, Appendix B. Motivated by previous findings [25, 33] and Collins et al.'s observation [10] on the lack of practical deniability tools, we introduced the first in-app deniability feature. This practical interaction provided users with a clearer understanding of deniability beyond theoretical concepts. We also expanded our study to explore participants' perspectives on deniability, including when they might choose not to use it, their overall stance, and whether they would switch to another app with deniability features. Additionally, we assessed the usability of the deniability feature using the System Usability Scale (SUS) [8].

3 METHODOLOGY

We conducted an IRB approved in-lab user study to evaluate the effectiveness of an editing feature, its usability and user's willingness to adopt it (**RQ1–4**). Our aim was to determine whether hands-on interaction would improve users' understanding of deniability, as outlined in previous research (refer Section 2). The study involved participants interacting with a real application designed to provide deniability, allowing them to directly engage with the concept.

3.1 Recruitment

Recruitment was conducted through flyers posted around the university campus (libraries, cafeterias, student centers) and online announcements on university social media platforms. We targeted university students due to their familiarity with messaging apps and similar real-life scenario experiences, though this limits the generalizability of our findings (refer Section 6). Flyers included a QR code for easy registration. We recruited 40 participants, with the initial 5 in the pilot interview study and 35 in the main study. Participants provided their available dates and times before the study. We identified overlapping time slots where groups of five participants were available simultaneously. Once common availability was established, participants were assigned to these groups of five, and informed of their appointment time. All participants were university students in the USA, aged 18 or older. Each participant received a $15 Amazon gift card for completing the study, which took about 1 hour.

3.2 Demographics

In our study, the participant demographics revealed a gender distribution of 57.1% male and 42.9% female. The age distribution was predominantly in the 18-24 age group, accounting for 54.3%, followed by the 25-34 age group, shown in Table 2. Ethnically, the majority of participants identified as Asian (54.3%), followed by White (17.1%), and Hispanic (14.3%). The educational background of participants indicated that 11 out of 35 held a Bachelor's degree, while 6 out of 35 had a Master's degree. When assessing technology proficiency, a significant portion (18 out of 35) considered themselves proficient. In terms of messaging app usage, 94.3% of participants used messaging apps, with SMS/MMS, Snapchat, WhatsApp, Instagram, and Discord being the most popular choices. The frequency of messaging app use revealed that 25 out of 35 participants used it "daily", while the rest used it "frequently". Additionally, of the 35 participants, 4 were familiar with the Signal app. This familiarity could influence their perceptions and responses, as discussed in Section 6.

3.3 Deniability Interface Design

To provide the deniability properties to E2E encrypted messaging apps, the interface we created needed to satisfy two key properties. (1) Both users must understand how to deny messages by changing their local record of the conversation and (2) Third parties must not be able to detect whether any change has been made or not. Previous research [25, 33] shows that there is lack of awareness of deniability, causing (1) to be a particular challenge.

Metric	Percentage
Gender Identification	
Male	57.1%
Female	42.9%
Other	0%

(a) Gender identification

Metric	Percentage
Age	
18-24 years	54.3%
25-34 years	37.1%
35-44 years	8.6%

(b) Age distribution

Metric	Percentage
Race / Ethnicity	
Asian/Pacific Islander	54.3%
White/Caucasian	17.1%
Hispanic	14.3%
Other/Multiple ethnicity	11.4%
Decline to answer	2.9%

(c) Race / Ethnicity

Table 2. Survey participant demographics. Our study had a relatively equal distribution between participants identifying as male and female. The age distribution of the participants skewed young. The predominant ethnic identification was Asian. Limitations associated with our demographics are addressed in Section 6.

To introduce deniability to users, we used one of the methods that Collins et al. [10] recommended in their study. We created an edit function that lets the user alter the message and its meta data locally. We propose a simple editing

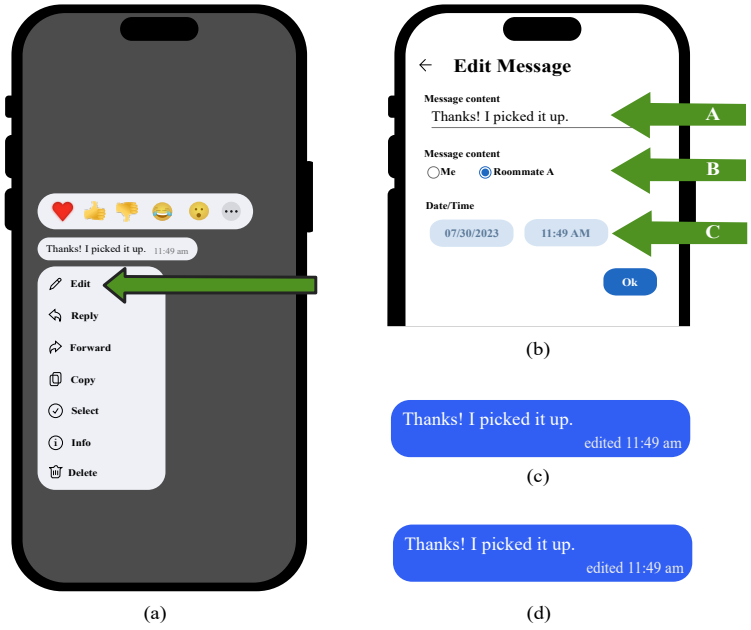


Fig. 1. Deniability feature interface implemented on Signal app. The (a) shows a new "edit" button in message context menus. After selecting the "edit" option, the user is presented with an interface with the option to modify three elements: (A) the content of the message, (B) the sender of the message, and (C) the date and/or time. Any message modifications are local, and not propagated to other participants in the conversation. (c) shows a message modified with the deniability feature enabled; (d) shows the same message without the deniability feature, displaying an "edited" flag.

feature in the messaging application that allows users to edit any sent or received message along with its metadata. This will provide the ability for the sender to deny the message as the recipient might have forged the message and introduce uncertainty in a third party's mind regarding the recipient's claim.

In our prototype, we modified the Signal Android app v6.27.7 [1] to support editing messages to convey deniability. A long-press on any sent or received message in a chat window opens a menu where the user can select the edit option, as shown in Figure 1. The *Edit message* window allows editing of (1) message content, (2) sender/receiver, and (3) sent/received date and time, as shown in Figure 1. Our prototype modifies the messages and their metadata on the user's phone locally, which means that the other party's app in that communication still has the original message details. However, it should be noted that for a third party, it would be impossible to know who modified the message details and this uncertainty is the basis of deniability claim.

Figure 1 (d) shows a normal message without the deniability feature enabled, where both the sender and the recipient can see that the message has been altered. It is important to note that since either party could theoretically fabricate an edited message with the 'edited' flag, the presence of this flag does not add to the authenticity or non-repudiation of the message. However, in scenarios where both phones show the same content marked as edited, it can be considered strong evidence that the message was indeed sent and subsequently edited. In contrast, Figure 1 (c) illustrates the deniability feature in action, where no cryptographic evidence of the original or edited message is visible, allowing both parties to deny the original content.

[1] https://github.com/signalapp/Signal-Android/releases/tag/v6.27.7

3.4 A Social Scenario To Demonstrate Deniability

Our goal was to create a natural environment for users to concentrate solely on the app and the concept of deniability. To achieve this, we implemented a casual conversation scenario featuring two roommates, A and B, who disagreed over rent payment. We chose a roommate dispute as the scenario for our study because it resonates with university students' real-life experiences, making it more relatable than the legal contexts used in previous studies, such as Reitinger et al. [25]. Using a relatable social scenario, our study offers a fresh perspective on deniability and its impact on everyday communication, highlighting its practical implications and user acceptance in familiar contexts.

The researchers replicated the scenario on two phones using the deniability feature (refer Section 3.3), designated as A and B, as depicted in Figure 3. This approach overcame the use of screenshots, which participants might distrust, as they are aware of their potential manipulation, unlike in-app messaging, where forgery seems less plausible. In our scenario, Roommate A was responsible for collecting and paying rent and utility bills. Roommate B claims to have left their portion of the rent on the table, which Roommate A denies finding. Both roommates provided their mobile chat as evidence Figure 3, forming the basis of the scenario presented to participants.

It's important to acknowledge that deniability inherently involves one party accusing the other. By assigning the characters as 'Roommate A' and 'Roommate B' without gender or other personal identifiers, we mitigated biases, ensuring participants engage with the scenario objectively. Our aim was to investigate deniability in a neutral, unbiased setting, allowing participants to focus solely on the scenario presented and provide responses based on communication content rather than external influences.

3.5 "Scenario + Interview" Pilot

To refine our study, we first conducted a pilot interview. Participants were introduced to the same scenario and phone evidence as in the main study and had the opportunity to interact with the deniability feature in the app, providing hands-on experience. Instead of a survey, we conducted semi-structured interviews, allowing for follow-up questions (refer Appendix A). This approach helped us refine our survey questions for the main study (refer Section 3.6) and test the effectiveness of the before-and-after survey method.

For instance, based on participant feedback, we revised survey questions like, "In which scenarios would you consider using an editing feature in an in-app messaging platform?" (refer Appendix B.3, question 12), adding options such as "Discussing information with your lawyer, legal professor, or law enforcement". We conducted the pilot study with 5 participants, each session lasting about 30 minutes on average, until no further improvements were identified.

3.6 "Scenario + Survey" Main Study

The study required participants to complete two main phases in the "Scenario + Survey" study, as shown in Figure 2. Conducted in groups of five participants each, the study presented one of the chat conversations (either A's or B's) from the scenario on a physical phone instead of a screenshot. The order was alternated to prevent bias, with four groups seeing A's chat first and three groups seeing B's chat first.

3.6.1 ***One Phone as Evidence.*** Participants were presented with the scenario (refer Appendix B) and provided with one of the physical phone chat conversations as supporting evidence, Figure 3. After viewing a conversation, they completed a survey, *Pre-Survey* (refer Appendix B.1), about their initial perspectives on the situation. Participants discussed the scenario before individually filling out a survey regarding their trust in the evidence, using lab computers to access the survey online. The initial survey aimed to understand participants' decision-making and perspectives on

deniability before any practical demonstration of editing features. These responses served as a baseline for their initial perceptions and comprehension of deniability.

The study assumed that one of the actors (Roommate A or Roommate B) involved in the chat conversation may have the ability to edit messages using the Signal app, and that the submitted in-app message evidence could potentially be forged. However, this information is not disclosed to participants until after they complete the *Pre-Survey* (refer Appendix B.1). Delayed disclosure enabled researchers to observe how participants' opinions and trust shift after learning about potential message manipulation and deniability.

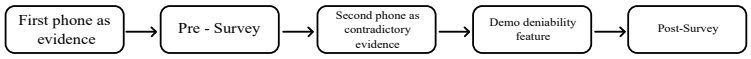


Fig. 2. Participant procedure flow within the scenario phase: Participants were surveyed before and after their interaction with the new deniability tool in our scenario. Questions and our script are available in Appendix B.

3.6.2 ***Both Phones as Evidence with Edited Messages.*** The next step involved demonstrating the editing feature in the Signal app. Participants were shown the second chat conversation as evidence. To introduce the concept of deniability, researchers demonstrated that there were two different records of the same conversation, explaining the deniability functionalities of the feature, Figure 3.

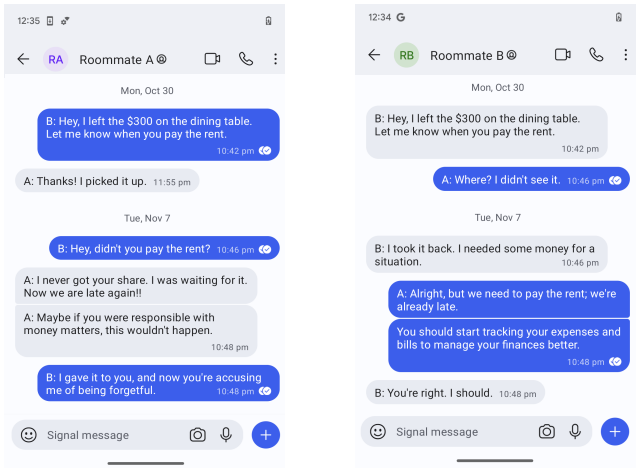
Each participant then tested the feature by sending a message and using each editable content option, allowing them to see firsthand how messages can be altered in the local apps, which leads to potential deniability. Participants were encouraged to discuss their views and experiment with the feature to enhance their understanding. After the demonstration, participants took *Post-Survey* (refer Appendix B.3), designed to identify any shifts in opinions on the scenario and changes in trust levels. The survey also delved deeper into deniability in everyday communication and included 10 questions from the System Usability Scale (SUS) [8] to measure the user-friendliness of the interface.

The survey aimed to address several questions, including: (1) Whether hands-on interaction with deniability led to a clearer understanding of the concept, (2) Whether participants could envision using deniability in real-life social scenarios after the hands-on experience, (3) Whether participants' trust in messaging apps changed after realizing the possibility of deniability, and (4) Whether users preferred an app that offered the editing feature and how likely they were to use such a feature. At the end, the demographics are collected using another survey. The participants were debriefed and compensated at the end. The full survey tool could be seen in Appendix B.1 and B.3.

3.7 Measuring Before And After Survey

To assess the impact of hands-on experience on participants' understanding of deniability and its practical usage, we measured their understanding before and after introducing them to deniability. Two surveys were employed for this purpose. *Pre-Survey* was administered after participants were introduced to the scenario and presented with one of the phones as evidence, aiming to gather initial perspectives on the scenario and participants' reasoning (refer Appendix B.1). To gauge changes in perspective after introducing deniability, *Post-Survey* was conducted (refer Appendix B.3). This survey mirrored *Pre-Survey* but included additional questions focused on deniability, aiming to determine if participants' opinions shifted after learning about the possibility of forgery. It also explored participants' preferences regarding deniability or repudiation. Both surveys included multiple-choice questions, Likert item questions[2], and open-ended questions. A 4-point Likert item was used instead of the standard 5-point item which helps in reducing the tendency of

[2]A Likert item question is a type of survey question that asks respondents to rate their level of agreement or disagreement with a statement on a scale, typically ranging from "strongly agree" to "strongly disagree."



(a) Rommate B's phone record. (b) Rommate A's phone record.

Fig. 3. Message conversation records on Roommate A's and B's phone: These message records were shown to the participants interchangibly to illustrate the concept of deniability. This demonstrates how deniability allows for two different records of the same conversation to exist simultaneously. The social scenario is used to help users understand the idea of deniability.

participants to select a neutral position when they are unsure or indifferent. This encourages them to make a more definitive choice, which can provide clearer insights into their attitudes and perceptions [9, 16]. Since each response category in a 4-point item carries more weight, small shifts in opinion can become more noticeable.

3.7.1 ***Quantitative Analysis.*** We asked participants in both the *Pre-Survey* and *Post-Survey* whether they believe roommate B's story, and how strongly, using a Likert item, Section 4.1. Because of deniability, there is no provable "right answer" to this question. Rather, if participants understood deniability, we expected fewer strong opinions on either extreme (Strongly Agree or Strongly Disagree) and more weaker opinions (Agree or Disagree). To analyze this, we tested whether the proportion of participants significantly changed between the binary categories [15, 17] of "Strong Opinion" or "Weaker Opinion" using a two-tailed, paired, McNemar's Exact (non-parametric) Test (this test is also known as a non-parametric, binomial sign test) [2, 24]. We used a significance level of α=0.05 and the hypotheses were defined before data analysis.

H_0: The proportion of participants with a strong opinion versus a weaker opinion is the same before and after their interaction with our deniability interface. H_a: The proportion of participants with a strong opinion versus a weaker opinion differs significantly. In applying McNemar's test, we considered the following assumptions: Participants provided paired data in both the *Pre-Survey* and *Post-Survey*, allowing us to compare opinions before and after interacting with our deniability interface. Opinions were categorized as "Strong" or "Weaker" (Dichotomous variables). The test assumed symmetry in the probability of opinion shifts and independence between participant pairs. Additionally, a visualization graph was used to aid in understanding the shift in opinions, Figure 4. This strategy was also employed to determine whether participants would opt to continue using their current messaging app or consider switching to an alternative platform, Section 4.4 .

3.7.2 ***Qualitative Analysis.*** For the analysis of open-ended question responses, we employed a deductive qualitative analysis method [1, 13]. Predefined codes were used for the questions, and the two researchers collaborated to develop a codebook (refer to Appendix C, Table 5), which was subsequently used to categorize the responses. This coding process

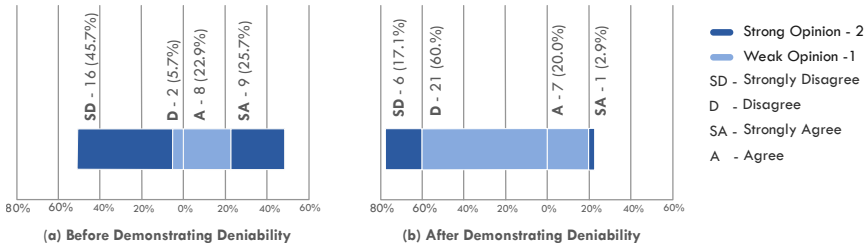


Fig. 4. Comparison of the extent to which users believe in our scenario before and after introducing deniability. Strong agreement indicates that roommate B is in the wrong, while strong disagreement suggests that roommate A is in the wrong. After demonstrating the deniability tool, we notice a shift towards the more weaker opinion in the center. The graph also highlights the distinction between 'Strong opinion' and 'Weak opinion' responses.

was followed by thematic analysis [23], which helped in identifying themes within the responses. Any disagreements between researchers regarding categorization were resolved through discussion to ensure consistency. Our aim was not to draw generalizable conclusions but to determine participants' preferences for deniability and their understanding of the concept.

4 RESULTS

In this section, we present our analysis of qualitative and quantitative results from the experiential study and surveys, organized to address each of our four research questions.

4.1 RQ1 - Deniability Comprehension

After hands-on experience with an E2E encrypted messaging app that enables deniability in a social context, do participants comprehend the concept of deniability and can they extrapolate its potential applications?

To begin to gauge participants' understanding of the purpose of deniability, we explored whether participants, post-introduction and hands-on experience, could conceptualize real-life scenarios or applications for deniability. Through our analysis conducted before and after introducing deniability, we observed shifts in participants' trust beliefs regarding their roommates. Additionally, a significant number of participants demonstrated an ability to identify scenarios where deniability could be applied. We also analysed the open ended response from the participants wherein they were asked to take a stance on deniability (refer Appendix B.3).

In the initial survey, participants were asked about their beliefs regarding whether roommate B paid the rent in a provided chat conversation scenario. The survey design included alternating versions of the conversation, utilizing either A's or B's phone, to explore potential variations in trust levels and decision-making based on the conversation's flow. Analysis revealed that groups presented with version A's conversation tended to favor responses rated 1 or 2 (strongly disagree and disagree, respectively), while those provided with version B's conversation showed a higher tendency to strongly agree that B paid the rent on a scale of 1–4. This resulted in roughly equal shifts in both directions. After introducing the deniability concept in the subsequent survey, participants were asked the same question to assess if their opinions or verdicts had changed, as shown in Figure 4 (b).

The visualization of the Likert item results in Figure 4 clearly shows how responses cluster around 'Agree' and 'Disagree', with 'Strongly Agree' and 'Strongly Disagree' responses pulling the consensus further from the center. Upon examining Figure 4, we observe that the group of 'Strong Opinion' values is predominant in Figure 4 (a), with 71.4% of the total responses. In contrast, in Figure 4 (b), the 'Weak Opinion' values now make up the majority, comprising 80% of the total responses. To analyze the data, as discussed in Section 3.7.1, we conducted an Exact McNemar's test, which

yielded an $p < .001$. This result leads us to reject the null hypothesis that there is no difference in the proportions of participants with strong opinions and weaker opinions. Therefore, we conclude that the proportion of participants with strong opinions differs significantly from those with weaker opinions.

The survey results also reveal a notable shift in participants' perspectives regarding the factors influencing their opinions on the verdict after being introduced to the deniability feature. Initially, 28 out of 35 participants considered "The available evidence in the messaging app" is the most significant factor that influenced their opinion on the verdict, indicating a high level of trust in the in-app messages as evidence. Additionally, 10 participants assigned importance to their trust in the messaging app. However, after learning about the deniability feature, the factor "Fact that text messages are editable" gained prominence, with 29 out of 35 participants acknowledging its significance. This suggests that the hands-on experience with the deniability feature led participants to reevaluate the reliability of chat evidence, recognizing the potential for messages to be edited. Interestingly, one participant expressed skepticism early on indicating a pre-existing mistrust in the reliability of in-app messages.

> *P503: "There is no evidence that Roommate A sent those messages. As we saw messages in Roommate B mobile. Here, those mobiles might be compromised."*

To assess participants' comprehension of the deniability feature's purpose, a question regarding when they would use the editing feature was posed in the survey. The majority (20 out of 35) selected 'Privacy' as the primary reason, followed by 'Accountability' and 'Legality' (both 14 out of 35) as shown in Figure 5 (a).

Furthermore, we conducted an analysis of participants' understanding of deniability using deductive open-coding, as detailed in Section 3.7.2. However, the responses were imprecise, making it difficult to determine whether participants fully comprehended deniability. Therefore, we do not present this categorization. Future research may benefit from more detailed questions to better understand participants' grasp of deniability.

Participants provided insightful opinions on deniability in addition to the information gathered by the research coordinators. These perspectives could be broadly classified into three categories: participants who recognized the relevance and benefits of deniability, participants who had strong concerns about its potential for misuse, and participants who identified potential limitations that could be put in place to avoid misuse. In this section, we have highlighted a few user responses that demonstrates their opinions on deniability.

4.1.1 ***Participants recognizing the relevance and benefits of deniability.***

> *P806: "... I however, also see the potential to do good especially in the form of allowing for freedom of speech that is not prosecutable by the governments trying to prevent it. ... Therefore, for the sake of journalists, other countries, and freedom of thought, I think it the pros of deniability outweigh their draw backs in my opinion."*

> *P1003: "I think that this feature can be a great tool for whistle blowers and other people who really need the deniability for their and society's benefit."*

4.1.2 ***Participants concerned about potential misuse.***

> *P202: " ... It seems like something criminals and people in power would exploit on the regular in comparison to everyday people just trying to have conversations."*

> *P802: " ... It will definitely super power the people/groups with criminal mentality and even planned well organized money heist or any such crime could take place at micro or macro level. This feature will also create rife disharmony among individuals and give boost to misunderstandings thus promote negativity in society in general."*

4.1.3 *Participants offering potential limitations to avoid misuse.*

> *P804: " ... This app should not be free of charge and in this way not the majority who do not need it won't be using it."*
>
> *P901: " ... May be only authorized personnel could use this feature rather than general public "*

Furthermore, participants recommended exploring the possibility of implementing this feature on a distinct platform. This suggests that users requiring deniability may have to utilize a separate messaging apps, limiting access to a more specific user base rather than making it publicly available.

> *P303: " ... if someone were undercover and needed to share info, they could have a dedicated app that is disguised as something else, rather than a common messaging app that any normal person can use."*

Overall, participants demonstrated diverse perspectives of deniability, acknowledging its possible advantages while actively supporting limitations to prevent future abuse. Further exploration revealed that participants associated the deniability feature with scenarios involving the communication of sensitive information across different legal or cultural norms (14 out of 35), as well as situations where individuals act as journalists (16 out of 35) or whistleblowers (16 out of 35), Figure 5 (b). This indicates a nuanced understanding of the feature's practical applications, demonstrating participants' ability to relate it to real-life situations and align it with concepts such as privacy, security, and confidentiality. This assisted us in addressing our **RQ1**, where hands-on experience played a crucial role in enhancing users' understading of deniability, evident in their shifting opinions post introduction to the concept. Furthermore, the connections participants established regarding the specific scenarios for utilizing the deniability feature provided insights into their grasp of the concept's purpose and functionality.

4.2 RQ2 - Interface Usability

> *Is our transcript-editing interface a user-friendly method of giving users intuitive access to deniability in E2E encrypted messaging apps?*

Deniability features have been implemented in other services before. However, our study is focused on introducing this capability in the context of an E2E encrypted messaging app by providing the ability to alter messages on local devices. During a dedicated session, participants engaged in hands-on interactions with the deniability feature, allowing them to send text messages and sequentially utilize deniability to edit content, sender/recipient, date, and time. This practical experience aimed to assess the user-friendliness and interactivity of the deniability feature, contributing to our exploration of **RQ2**.

In the main survey study, each participant was requested to alter the message's text as well as its meta data, by interacting with the deniability feature. This was crucial in determining whether the interface was engaging and user-friendly, which helped us answer our **RQ2**. To measure the user experience, we incorporated the System Usability Scale (SUS) at the end of *Post-Survey*. Participants, having actively engaged with the deniability feature, provided ratings on its user-friendliness. For a more detailed interpretation of each SUS score, we relied on the insights of various researchers who extensively analyzed SUS scores in multiple studies [4, 5, 27]. Following Sauro et al.'s analysis of SUS scores from studies by Bangor et al. [5], Tullis et al. [29], and their own data [27], calculated percentile values and assigned letter grades based on percentile ranges. Bangor et al. [4] introduced adjective-based ratings, aligning scores with the closest adjective in terms of standard deviations.

Notably, in our findings, the mean SUS score obtained was 72.4, with a median of 72 and standard deviation of 18.9, categorizing it as having "Good" usability according to the adjective-based rating [4]. The user interface usability is

above the 50th percentile (graded as "B"). This outcome suggests that the user interface of the deniability feature was well-received, meeting the criteria of user-friendliness. This positive outcome addressed our research question **RQ2**.

4.3 RQ3 - Logical Reasoning

Can users reason about when deniability is useful and/or ethical in the E2E encrypted messaging app context?

As we transitioned from exploring various aspects of deniability and assessing user understanding, it became crucial to examine whether users could articulate logical reasons for when deniability is useful or not, along with incorporating ethical considerations (**RQ3**). To achieve this, we posed an open-ended question, allowing users to express their stance on deniability and suggest any limitations they found necessary for enhancing its utility.

We analyzed the qualitative data using deductive open coding, as discussed in Section 3.7.2, to identify overarching themes. Our aim was to determine if participants favored deniability and understood the concept, providing justification for their stance. Our analysis revealed that 6 out of 35 participants preferred deniability, while 20 opposed it, primarily due to misuse concerns. Additionally, 8 participants favored a balance of accountability and deniability, suggesting about 25% supported a mixed approach (refer Table 3).

We attempted to categorize participants' ethical reasoning about deniability into various ethical frameworks. However, the responses lacked precision, consistency, and clarity regarding the specific ethical framework applied by users. Therefore, we opt not to present this results here. Future researchers interested in exploring this aspect further may need to incorporate more extensive questioning to effectively capture users' ethical reasoning styles.

4.4 RQ4 - User App Preferences

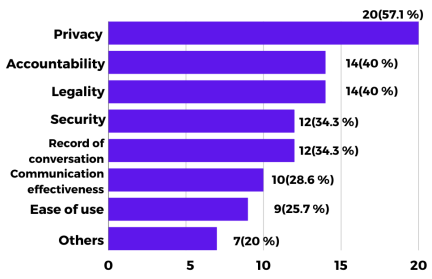
Will users consider switching to/away from E2E encrypted messaging apps that implement deniability features, and why?

Our research into user behavior regarding deniability extended beyond understanding purpose and interface to explore the practical implications of incorporating this feature. The focus was on whether users, when presented with the opportunity to use deniability, would opt to continue using their current messaging app or consider migrating to an alternative platform, **RQ4**. This aspect holds particular importance for messaging app companies contemplating the introduction of deniability, as it aids in evaluating the potential impact on user engagement and attraction.

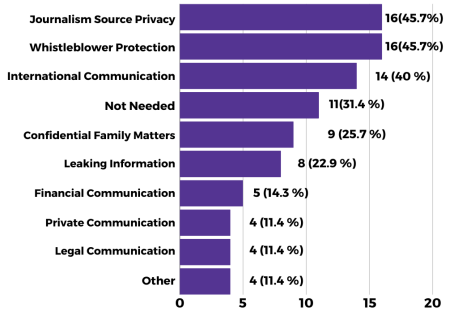
Stance on Deniability	Frequency	Comprehensive Understanding
Preferred Deniability	6	2
Opposed Deniability	20	3
Favored Combination	8	6

Table 3. This table illustrates the different stances participants took on deniability, along with how many had a comprehensive understanding of the concept and provided logical justifications for their positions. The data indicates that a balanced approach was favored by about 25% of the participants.

To assess user preferences, we designed two questions (see Appendix B.1 question 13 and 14) and administered both before and after the introduction of deniability. To analyze the data, as discussed in Section 3.7.1, we conducted an Exact McNemar's test and obtained an $p = .77$ which was not less than 0.05 (refer Table 4). As a result, we cannot conclude that the deniability feature alone impacts users' decisions to stay with their current app or switch to another platform. This suggests that users' choices may be influenced by more subtle features or overall user experience. Hence, deeper human-centered research is necessary to assess its true impact.



(a) Relative popularity of reasons participants gave for why deniability might be desirable.



(b) Relative popularity of contexts in which participants said deniability would be desirable.

Fig. 5. The figure shows the relative popularity of reasons and contexts for deniability. (a) shows 'Privacy' was the most frequently chosen reason, followed by 'Accountability', with 'Legality' responses nearly equivalent to accountability. This suggests that participants correctly relate deniability features with privacy concerns. (b) shows most chose scenarios involving the communication of sensitive information, demonstrating logical reasoning extending this tool's capabilities to new contexts.

4.5 Additional Analysis: Exploring User Perspectives Beyond RQs

The data analysis highlights a notable shift in participants' perception of the trustworthiness of messaging apps following the introduction of the deniability feature. Initially, when asked to rate their trust in messaging apps on a scale of 1–4, with 4 being highly trustworthy, a significant majority leaned towards scores of 3 or 4. However, after being exposed to the deniability concept, 51.14% (18 out of 35 participants) indicated a diminished trust in in-app messages, Figure 6. This shift suggests that participants now recognize the presence of plausible deniability within text conversations and increased awareness that the content may be unreliable.

An interesting observation emerged as some participants expressed a desire for non-repudiation, indicating a preference for freezing their messages, in contrast to the concept of deniability. This question was posed both before and after the introduction of deniability, aiming to investigate whether the introduction of deniability influenced participants' inclination towards non-repudiation. Notably, 24 out of 35 participants indicated a desire for some form of non-repudiation, expressing a wish to freeze text messages at a certain point. This suggests that even with the implementation of deniability, users might still seek a level of accountability over their messages. This finding opens avenues for further research, suggesting the potential need for features that allow the detection of any modifications i.e non-repudiation, while still allowing for the selective application of deniability to specific messages.

Research Questions	Test	p-value	Findings
RQ1	Exact McNemar's Test (Opinion Shift)	$p < .001$	Significant difference in strong vs. weaker opinions.
RQ4	Exact McNemar's Test (App Switch)	$p = .77$	No significant difference.

Table 4. This table illustrates key findings for RQ1 and RQ4, showing Exact McNemar's Test results and p-values.

4.5.1 ***Ties to Users' Privacy and Security Posture.*** Our surveys included standardized scales for measuring both the privacy and security posture of our participants. We performed statistical tests to measure whether participants'

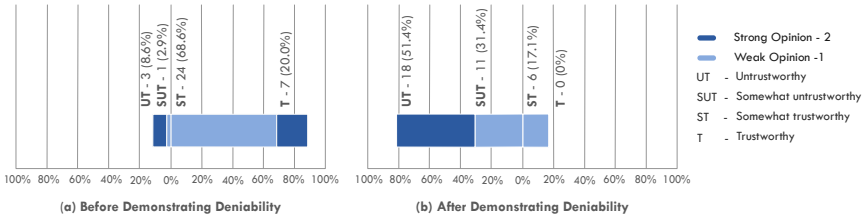


Fig. 6. Comparison of participants' trust levels for messaging app transcripts before and after demonstrating our deniability feature, rated from 1 (*Untrustworthy*) to 4 (*Trustworthy*). The noticeable shift in the graphs (a) and (b) indicates a clear decrease in trust of message transcripts. This indicates that participants can reason about deniability, as applied here.

desire for deniability was correlated with their security posture, but the results lacked statistical significance in both cases.

5 DISCUSSION

Contrary to the narrative that deniability in messaging would be too complex for non-cryptographers to effectively employ, we show that users can quickly start to reason about deniability. This is true both in the technical sense, as well as the ethical sense. Experiencing these functionalities firsthand swiftly enabled users to form opinions in this space. They were able to recognize the record-ambiguity this feature caused, and suggest use cases for the technology.

A majority wanted at least some limits on this feature, whether entirely banning it or making it available only in specific contexts. Contrary to our initial assumptions, we found that a high level of privacy and security awareness does not necessarily correlate with a preference for deniability features. This indicates a more complex relationship between user awareness, digital communication privacy, and app functionality. Users' preferences for deniability features appear to be influenced by a variety of factors, not solely their consciousness of privacy and security issues.

To address the variety of user opinions in this area, we suggest that E2E encrypted messaging applications which want to make this feature available to users introduce *both* strong deniability *and* strong non-repudiation together. Let users decide the contexts in which they want the privacy of deniability, versus the accountability of non-repudiation. Our findings suggest that they will be able to manage this choice reasonably, if given the chance.

There is some indication that app developers are considering such a feature. WhatsApp recently created a limited messaging editing feature, which allows users to retract and correct recently sent messages [21]. If the recipient goes offline before receiving the edit, there will be briefly inconsistency between transcripts. However, this feature stops far from providing deniability through local-only message editing. There is an encrypted messaging protocol under development by Wang et al. (NOTRY) [32] that aims to combine the properties of deniability with the retroactive ability to "avow" past messages, certifying their contents. This demonstrates one way that the choice between deniability and non-repudiation could be presented to users – supported by a strong cryptographic protocol.

6 LIMITATIONS

Geography and culture: Our research context involved participants from the same institution, potentially influencing their perspectives due to the specific cultural, societal, and legal environment. Privacy concerns vary widely based on cultural, political, and socio-economic factors, meaning our findings may not be universally applicable across different regions.

Age: The majority of study participants were between the ages of 25 and 34. It is important to recognize that results could vary for older age groups, which calls for taking into account broader demographic viewpoints.

Sampling bias: The sample size (n=40) and specific background of our participants limit the generalizability of our findings. As the participants may not represent the broader population, this study should be seen as exploratory, with the intent of using these results to inform a larger-scale project in the future.
Participant fatigue: Although the one-hour session included study, setup, and feedback, and was designed to be interactive to maintain engagement, we acknowledge the potential for participant fatigue. However, participants did not report any fatigue, suggesting that the session length and structure were appropriate for the tasks involved.
Technology proficiency: Given the nature of our sample, which is comprised primarily of non-expert individuals from diverse backgrounds, the generalizability of our results to other populations, particularly those who are more technologically adept or privacy-conscious, may be limited.
Hypothetical scenario: Our study relied on hypothetical scenarios within a controlled experimental setup. While these scenarios aimed to mimic real-life situations, participants' responses may have been influenced by the artificiality of the environment.
Educational background: All participants were students, and the study was carried out in a university controlled environment. The predominant educational level among participants was high school. It is essential to note that results may exhibit variations in different settings and with participants from diverse backgrounds.
Self-reported data: The study relied on self-reported data from participants, which can introduce biases. Participants' responses to survey questions may be influenced by their perceptions, memory, and social desirability bias, where they may respond in a manner they believe is favorable or acceptable rather than being entirely candid.
App usage and understanding of end-to-end encryption (E2E): The study did not directly assess participants' understanding of E2E encryption or their long-term usage of the Signal app. Despite being informed about the difference between edited messages in WhatsApp and deniability, participants might still have been confused. These factors could influence their perception of the deniability feature, as familiarity with E2E encryption and experience with the app may affect their attitudes and expectations. Furthermore, the study found that more than 90% of participants utilized messaging apps on a daily basis. In situations where individuals don't use messaging applications frequently, the results could be different. Understanding the study's scope and generalizability requires taking these factors into account. The limitations of our research, including the specific demographic and cultural context, as well as the controlled experimental setup, highlight the need for future studies by including a wider range of populations and real-world scenarios.

7 FUTURE WORK

Usable deniability in messaging applications is still in its early stages, but more research in this area can greatly enhance the field. To raise user knowledge and provide more light on the complexities of deniability, more human-centered research is necessary.

One possible direction for further research entails testing user ability to employ deniability to protect sensitive information – utilizing paired interactions as was done with the secure email study by Routi et al. [26]. This study can utilize paired interactions by bringing novice participants and instructing them to exchange messages and use the deniability feature with each other. To ensure a more natural interaction, participants could be encouraged to bring a friend or a family member along, ensuring pre-existing familiarity. This approach would overcome some of the stiltedness of a fixed scenario and demonstrate whether users can collaboratively reason about how to employ deniability. Our study was conducted on a limited population, enabling us to analyze user perspectives and gain a deeper understanding within a controlled environment. We discovered important insights by concentrating on the real-world applications of deniability in messaging applications, which offer a solid foundation for more extensive

generalization. Moving forward, it is important for future studies to expand upon our findings by applying them to a more diverse general population and exploring potential cultural differences. The message-transcript editing feature we built into Signal has not been battle-hardened to resist forensic efforts to detect whether message modifications have been made. This evidence could be leaked using, for example, remnants on the hard drive, traffic analysis at the time of the original conversation, or evidence from app error logs. Future work could improve the robustness of the deniability feature against stronger adversaries. Our study yielded a score of 72 in SUS usability scale, which indicates decent usability with room for improvement. Future work should focus on improving usability of the tool. Incorporating user feedback and addressing usability challenges will be crucial to enhancing the overall user experience and effectiveness of the deniability. A final direction we encourage for future research and development is to use message signing to accomplish non-repudiation. It could be achieved by allowing users to sign particular communications, so freezing them in their original form. Simultaneously, granting users the choice to enable deniability for additional communications would present a complete and adaptable communication experience. Examining the viability and acceptability of such a dual approach by users may provide insightful information and influence the future development of deniability in messaging apps.

8 CONCLUSION

Deniability allows individuals to communicate freely. However, users often perceive messaging transcripts as non-repudiable. Currently, people do not have practical access to tools that provide both deniability and non-repudiation. The primary goal of our study was to provide users a hands-on access to an application, which allows altering of message records, enabling deniability. We measured users' understanding of the significance and the trade-offs associated with deniability. We discovered that users were able to comprehend the idea of plausible deniability when presented with a chance to experience deniability firsthand in a social setting. Moreover, even the users who disagreed with the concept of deniability have identified real-life situations in which they thought deniability could be useful. This indicates that raising user awareness about deniability is not challenging; instead, more user-centric research with increased real-world application and an effective teaching method is needed. Additionally, it was found that while some users desired deniability, there existed a subset of users who wanted non-repudiation. We suggest that if deniability is to be provided, then it should be in conjunction with non-repudiation. Users deserve to understand and control where their conversations exist on the spectrum of deniability to non-repudiation.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation, Grants No. 1757207, 2348181, and 1816929. We deeply appreciate Kaylia Reynolds for her invaluable statistical expertise, which was instrumental in refining our analysis. Special thanks to Dr. Gaurav Panwar and Dr. Bill Hamilton for their guidance, Kartick Kolachala and the Ciberseguros lab team, including Jack Nolen, Bryan White, Xindi Zheng, and Cheyenne Salem, for their contributions. Malek Al-Jbour thanks Al-Zaytoonah University of Jordan for their sponsorship. We also thank our families and friends for their support, and the reviewers for their valuable feedback.

REFERENCES

[1] Hiteshkumar Aacharya. 2022. CONTENT ANALYSIS. 07 (04 2022), 04.

[2] Omolola A Adedokun and Wilella D Burgess. 2012. Analysis of paired dichotomous data: A gentle introduction to the McNemar test in SPSS. *Journal of MultiDisciplinary Evaluation* 8, 17 (2012), 125–131.

[3] Wei Bai, Michael Pearson, Patrick Gage 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)*. 210–219. https://doi.org/10.1109/EuroSPW51379.2020.00036

[4] Aaron Bangor, Phil Kortum, and James Miller. 2009. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *J. Usability Stud.* 4 (04 2009), 114–123.

[5] Aaron Bangor, Philip T. Kortum, and James T. Miller. 2008. The System Usability Scale (SUS): an Empirical evaluation. *International Journal of Human-Computer Interaction* 24 (08 2008), 574. https://doi.org/10.1080/10447310802205776

[6] Vincent Bindschaedler, Reza Shokri, and Carl A Gunter. 2017. Plausible deniability for privacy-preserving data synthesis. *arXiv preprint arXiv:1708.07975* (2017).

[7] Nikita Borisov, Ian Goldberg, and Eric Brewer. 2004. Off-the-Record Communication, or, Why Not to Use PGP. In *Proceedings of the 2004 ACM Workshop on Privacy in the Electronic Society* (Washington DC, USA) *(WPES '04)*. Association for Computing Machinery, New York, NY, USA, 77–84. https://doi.org/10.1145/1029179.1029200

[8] John Brooke. 1996. Sus: a "quick and dirty'usability. *Usability evaluation in industry* 189, 3 (1996), 189–194.

[9] Seung Chyung, Katherine Roberts, Ieva Swanson, and Andrea Hankinson. 2017. Evidence-Based Survey Design: The Use of a Midpoint on the Likert Scale. *Performance Improvement* 56 (11 2017), 15–23. https://doi.org/10.1002/pfi.21727

[10] Daniel Collins, Simone Colombo, and Loïs Huguenin-Dumittan. 2023. Real World Deniability in Messaging. Cryptology ePrint Archive, Paper 2023/403. https://eprint.iacr.org/2023/403

[11] Juan Vera del Campo, Josep Pegueroles, Juan Hernández-Serrano, and Miguel Soriano. 2014. DocCloud: A document recommender system on cloud computing with plausible deniability. *Information Sciences* 258 (2014), 387–402. https://doi.org/10.1016/j.ins.2013.04.007

[12] Mario Di Raimondo, Rosario Gennaro, and Hugo Krawczyk. 2006. Deniable Authentication and Key Exchange. In *Proceedings of the 13th ACM Conference on Computer and Communications Security* (Alexandria, Virginia, USA) *(CCS '06)*. Association for Computing Machinery, New York, NY, USA, 400–409. https://doi.org/10.1145/1180405.1180454

[13] Satu Elo and Helvi Kyngäs. 2008. The qualitative content analysis. *Journal of advanced nursing* 62 (05 2008), 107–15. https://doi.org/10.1111/j.1365-2648.2007.04569.x

[14] Kseniia Ermoshina, Francesca Musiani, and Harry Halpin. 2016. End-to-End Encrypted Messaging Protocols: An Overview. In *Third International Conference, INSCI 2016 - Internet Science (Lecture Notes in Computer Science (LNCS), Vol. 9934)*, Franco Bagnoli, Anna Satsiou, Ioannis Stavrakakis, Paolo Nesi, Giovanna Pacini, Yanina Welp, Thanassis Tiropanis, and Dominic DiFranzo (Eds.). Springer, Florence, Italy, 244 – 254. https://doi.org/10.1007/978-3-319-45982-0_22

[15] Norman G. 2010. Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education* 15 (2010), 625, 32.

[16] Ron Garland. 1991. The Mid-Point on a Rating Scale: Is it Desirable? https://api.semanticscholar.org/CorpusID:146702037

[17] Spencer E. Harpe. 2015. How to analyze Likert and other rating scale data. *Currents in Pharmacy Teaching and Learning* 7 (2015), 836–850. https://api.semanticscholar.org/CorpusID:54775132

[18] Michal Kedziora, Yang-Wai Chow, and Willy Susilo. 2017. Threat models for analyzing plausible deniability of deniable file systems. *Software Networking* 1 (2017), 241–264. https://doi.org/10.13052/jsn2445-9739.2017.012

[19] Christiane Kuhn, Maximilian Noppel, Christian Wressnegger, and Thorsten Strufe. 2021. Plausible Deniability for Anonymous Communication. In *Proceedings of the 20th Workshop on Workshop on Privacy in the Electronic Society* (Virtual Event, Republic of Korea) *(WPES '21)*. Association for Computing Machinery, New York, NY, USA, 17–32. https://doi.org/10.1145/3463676.3485605

[20] Julia Len, Esha Ghosh, Paul Grubbs, and Paul Rösler. 2023. Interoperability in End-to-End Encrypted Messaging. Cryptology ePrint Archive, Paper 2023/386. https://eprint.iacr.org/2023/386 https://eprint.iacr.org/2023/386.

[21] WhatsApp LLC. 2023. Now You Can Edit Your WhatsApp Messages. https://blog.whatsapp.com/now-you-can-edit-your-whatsapp-messages.

[22] Moxie Marlinspike. 2016. Signal Protocol Specifications. https://signal.org/docs/.

[23] Muhammad Naeem, Wilson Ozuem, Kerry Howell, and Silvia Ranfagni. 2023. A step-by-step process of thematic analysis to develop a conceptual model in qualitative research. *International Journal of Qualitative Methods* 22 (2023), 16094069231205789.

[24] Matilda QR Pembury Smith and Graeme D Ruxton. 2020. Effective use of the McNemar test. *Behavioral Ecology and Sociobiology* 74 (2020), 1–9.

[25] N. Reitinger, N. Malkin, O. Akgul, M. L. Mazurek, and I. Miers. 2023. Is Cryptographic Deniability Sufficient? Non-Expert Perceptions of Deniability in Secure Messaging. In *2023 2023 IEEE Symposium on Security and Privacy (SP) (SP)*. IEEE Computer Society, Los Alamitos, CA, USA, 274–292. https://doi.org/10.1109/SP46215.2023.00095

[26] Scott Ruoti, Jeff Andersen, Luke Dickinson, Scott Heidbrink, Tyler Monson, Mark O'neill, Ken Reese, Brad Spendlove, Elham Vaziripour, Justin Wu, Daniel Zappala, and Kent Seamons. 2019. A Usability Study of Four Secure Email Tools Using Paired Participants. *ACM Trans. Priv. Secur.* 22, 2, Article 13 (apr 2019), 33 pages. https://doi.org/10.1145/3313761

[27] J. Sauro. 2011. *A Practical Guide to the System Usability Scale: Background, Benchmarks and Best Practices.* Measuring Usability LLC. https://books.google.com/books?id=BL0kKQEACAAJ

[28] Michael Schliep and Nicholas Hopper. 2018. End-to-End Secure Mobile Group Messaging with Conversation Integrity and Minimal Metadata Leakage. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security* (Toronto, Canada) *(CCS '18)*. Association for Computing Machinery, New York, NY, USA, 2282–2284. https://doi.org/10.1145/3243734.3278506

[29] Jacqueline N Stetson and Thomas S Tullis. 2004. A comparison of questionnaires for assessing website usability. *UPA Presentation* (2004).

[30] Nik Unger and Ian Goldberg. 2018. Improved Strongly Deniable Authenticated Key Exchanges for Secure Messaging. *Proc. Priv. Enhancing Technol.* 2018, 1 (2018), 21–66.

[31] Nihal Vatandas, Rosario Gennaro, Bertrand Ithurburn, and Hugo Krawczyk. 2020. On the Cryptographic Deniability of the Signal Protocol. In *Applied Cryptography and Network Security: 18th International Conference, ACNS 2020, Rome, Italy, October 19–22, 2020, Proceedings, Part II* (Rome, Italy). Springer-Verlag, Berlin, Heidelberg, 188–209.

[32] Faxing Wang, Shaanan Cohney, Riad Wahby, and Joseph Bonneau. 2023. NOTRY: deniable messaging with retroactive avowal. *Cryptology ePrint Archive* (2023).

[33] Tarun Kumar Yadav, Devashish Gosain, and Kent Seamons. 2023. Cryptographic Deniability: A Multi-perspective Study of User Perceptions and Expectations. In *32nd USENIX Security Symposium (USENIX Security 23)*. USENIX Association, Anaheim, CA, 3637–3654. https://www.usenix.org/conference/usenixsecurity23/presentation/yadav

A SEMI-STRUCTURED INTERVIEW

This script and these questions guided the pilot semi-structure interviews:

Thank you for participating in our interview. Your input is crucial for our research on Internet communication and messaging app features. Your feedback will help improve messaging platforms to better suit user preferences and needs.Please take a moment to read and sign the provided consent form before we begin. Before we start the interview, while answering our questions, remember, there are no right or wrong answers. Your experiences and opinions are what matter most, so please answer the questions based on your knowledge and personal experiences. Let's begin the session.

(1) Do you use texting apps?
(2) How often do you use messaging apps? Do you use it daily or weekly?
(3) Could you please tell me, which messaging apps you commonly use?

Now, let's imagine a situation where two of your friends (Friend A and Friend B) have had a disagreement, and both of them come to you, sharing their side of the story. But the stories are different. Because they are different you know they can't both be true at the same time.

(1) How do you determine whom to believe?
(2) What factors influence your decision-making process?
(3) Can you please explain in more detail the thought process you went through to reach your decision in this situation?
(4) Does your relationship with each friend influence your decision-making process in this situation?
(5) What proof or evidence do you need to believe one over the other?

In the same situation, suppose the friend you didn't believe from the previous scenario showed you the chat conversation of their fight as evidence.(Now, based on this):

(1) Which friend do you believe?
(2) What factors contribute to your decision?
(3) Does that mean the chat evidence made you come up with a decision?
(4) What about the chat made it more trustworthy? Was it the technology or just the personal relationship?
(5) Did you change your mind? Why or why not?

Consider this chat conversation between Roommate A and Roommate B, two roommates who are in a dispute over the utility bill. One of the friends brought in her chat conversation as evidence. You can go over the chat conversation, Figure 3 .

(1) Based on the chat conversation that you just read, whom do you trust?
(2) Do you believe Friend 2 has paid the rent?

(3) Why do you think so?
(4) What about the chat made you believe that?
(5) Do you believe that Friend 2 made comments about Friend 1's disorganized nature?
(6) Why do you believe that?
(7) Do you believe it because it is over a chat/messaging app?

Are you aware that some of the messaging apps have an edit feature that allows users to go back and correct their messages after sending them?

Yes: Can you tell me what you know about them?
No: It's a feature that allows users to edit the content of a text message after it has been sent. When a message is edited, it is displayed in the chat with the label 'Edited' to indicate that changes have been made.

(1) What do you think about this editing feature?
(2) What about the feature you like?
(3) Do you think the other person knows about the changes you did?
(4) How do they know that?
(5) Do you think there can be a feature where you can edit a message without the other person being notified about the changes?
(6) What do you think about such a feature?
(7) Why do you believe it can exist or not exist?
(8) What is your opinion on such a feature?
(9) Do you believe it is possible for such a feature to exist in messaging apps?
(10) Why or why not?

I will demonstrate one of the features that we have implemented in the messaging application. When you click on a message, it displays an edit option, allowing you to modify the message content, change the recipient or sender information, and even edit the metadata.

The interesting part of this feature is that any changes you make on your local phone are not visible to the other person. They remain unaware that you have edited your version of the chat. [Do they have a different version on their phone - yes, that's true.] Allowing you to deny that you ever said any of those things. This aspect here is known as deniability. This means that the chat conversation over the messaging app resembles verbal communication in that it is unreliable. All parties involved in the conversation have the ability to deny their part or even their presence in the discussion. It gives an extra layer of privacy and confidentiality to the communication.

(1) Now since you know there is such a deniability feature, from the previous chat conversation. Whom do you trust now in the dispute?
(2) Why do you believe that?
(3) What about the chat or feature that made you change your view?
(4) Do you believe Friend 2 has paid the rent?
(5) Why? Why not?
(6) Do you think the deniability feature that was demonstrated made you change your opinion?

(7) Can you please explain what you have understood about this deniability feature?
(8) Can you give an example and elaborate?
(9) Did the hands-on experience with the app, helped you?
(10) In what situations do you think you might use this feature, if at all?
(11) Can you provide some real-life examples where you find this feature useful?
(12) Do you use it for sharing personal information?
(13) What do you believe other people might use this feature for?

Now that you understand the concept of deniability. Deniability, as a feature, has its pros and cons. On one hand, it helps protect privacy and gives individuals control over their conversations. On the other hand, there's a worry that it might be misused by criminals to hide their malicious activities.

(1) What do you think about deniability?
(2) Do you believe the benefits of privacy outweigh the risks of potential misuse?
(3) Why do you believe that?
(4) Do you think there should be limitations on this feature to prevent abuse?
(5) What kind of limitations?

We have completed the interview process. Thank you for participating and sharing your valuable opinion. We hope you had as much fun as we did having you here. Before we conclude, could you please take a moment to review the debrief form? And if you have any questions, feel free to ask.

B DENIABILITY SCENARIO SURVEYS AND SCRIPT

This is the script and the survey questionnaire used for focus group study phase.

Welcome to our research study. We appreciate your participation today. In this study, we have a scenario where two roommates have come to us for assistance in resolving a financial dispute. Your role is crucial in helping us make an informed decision. So let us walk you through the scenario. Roommate A and Roommate B have been living together for several months, and, for the most part, things have been going well. However, recently, a dispute arose about paying rent and utility bills. Roommate A is responsible for collecting all the money and paying the bill. Roommate B has to give their share of the bills. Roommate B insists that they've left their portion of the rent and utility expenses on the table, ready for Roommate A to collect. However, Roommate A, says they never found any money left on the table.
It's our job to decide who might be in the right, and for that, we need your help. Here is the Roommate <A/B> phone that they provided as part of the evidence. You can go over the chat conversation that Roommate<A/B> had with Roommate<A/B>, Figure 3. Now on the computer in the lab, you will find a short survey related to the chat conversation you just read. Please take your time to complete the survey thoughtfully. You will be asked to share your opinions and reactions to the conversation and provide input on who you think may have a more valid point in this dispute.

B.1 Pre-Survey

(1) Participant ID
(2) On a scale of 1 to 4, to what extent do you believe that Roommate B paid the bill in this scenario? 1 - Strongly disagree, 2 - Disagree, 3 - Agree , 4 - Strongly agree
(3) On a scale of 1 to 4, to what extent do you believe the statement (Roommate A criticized Roommate B of being irresponsible) in this scenario? 1 - Strongly disagree, 2 - Disagree, 3 - Agree , 4 - Strongly agree

(4) What was the most significant factor that influenced your opinion on the verdict?
   a. The available evidence in the messaging app
   b. The credibility of the parties involved in the conversation
   c. The context and tone of the messages
   d. My personal beliefs about the criminal justice system
   e. My level of trust in text message
   f. The level of emotional tension or conflict in the conversation
   g. Other
(5) According to you, how trustworthy are in-app messages? 1 - Untrustworthy, 2 - Somewhat untrustworthy, 3 - Somewhat trustworthy, 4 - Trustworthy
(6) On a scale of 1-4, to what extent do you believe that users should be able to change their in-app message records for privacy reasons? 1 - Strongly disagree, 2 - Somewhat disagree, 3 - Somewhat agree, 4 - Strongly agree
(7) What possible reasons do you think someone might have for wanting an editing feature in in-app messaging platform? Please select all that apply.
   a. Privacy
   b. Accountability
   c. Security
   d. Legality
   e. Record of conversation
   f. Communication effectiveness
   g. Ease of use
   h. Other
(8) How useful do you think the editing feature would be in real-life situations? Not at all, Somewhat, Very, Extremely
(9) How much would YOU (not the other person) like to have the ability to edit messages in an in-app messaging platform? Not at all, Somewhat, Very, Extremely
(10) How much would you like EVERYONE to have the ability to edit messages in an in-app messaging platform? Not at all, Somewhat, Very, Extremely
(11) How likely are you to use an app where EVERYONE has the ability to edit messaging records? Not at all, Somewhat, Very, Extremely
(12) On a scale of 1 to 4, how often would you use the message editing feature, if it was available in your messaging app? 1 - Never, 2 - Rarely, 3 - Occasionally, 4 - All the tim
(13) On a scale of 1-4, if the messaging app you are currently using provided a feature to edit message records, to what extent would this influence your decision to continue using the app? 1 - Definitely stop using the app, 2 - Probably stop using the app, 3 - Probably continue using the app, 4 - Definitely continue using the app
(14) On a scale of 1-4, if another messaging app provided a feature to edit message records, to what extent would this influence your decision to switch to that messaging app? 1 - Not switching to the other app , 2 - Unlikely to switch to the other app, 3 - Very likely to switch to the other app, 4 - Definitely would switch to other app
(15) On a scale of 1-4, how useful do you find the option to freeze certain in-app messages and prevent them from being edited? 1 - Not useful at all, 2 - Somewhat not useful, 3 - Somewhat useful, 4 - Extremely useful

B.2 Hands-on Experience Script

We also have Roommate <A/B> phone with us. Please have a look at their end of the conversation, Figure 3. It's different, right? Why is it? We have a special feature in our mobile app. It allows you to change your chat messages after you've sent them, but here's the twist: these changes are only visible on your device, and the person you're chatting with won't know you've made any edits. This feature is called "cryptographic deniability". In simple terms, it means that in a secure chat system, if someone sends you a message, they can later say they never sent it or deny the content of the message. And there's no way to prove that they did, just like verbal communication. It's like having a secret conversation that leaves no trace, so even if you show the message to others, the sender can say they never actually sent it or that they didn't write the text content that way, and there's no evidence to prove otherwise. This feature is often used to protect the privacy and confidentiality of messages in secure communication systems. We'd like to show you how it works. Our feature could edit/forge messages in three ways: content, date/time, and recipient.
Deniability Functionality:
Content: You can edit the content of a message. This means you can modify the actual text of the message.
Date/Time: You can also adjust the date and time of the message, which can be useful for reordering messages in a chat history.
Recipient: You have the option to change the recipient of the message, allowing for a different interpretation of who received it.
Feel free to experiment with the feature, and don't hesitate to ask us if you have any questions or need assistance. An important thing to remember is that these changes/modifications occur only on your own device. It's like having a little secret toolkit on your phone. You can tweak your messages and adjust the conversation, all in the privacy of your own device. And the other person in the chat conversation does not have the modification. Now that you have played around with the app and understand deniability. Let's go back to our scenario, the task for which you were here decide whether Roommate A or Roommate B is right. You noticed that there are two phones over here that have two different versions of the same conversation. So, that means either any one of them could have used our deniability feature and is lying to us or maybe both of them could have used it together. We don't know!
Now that you've had the opportunity to explore the deniability feature and functionality, we would like you to take a new survey. Please access the survey on the computer systems once again and provide your feedback based on your observation. We have a System Usability scale at the end of this survey, where you could give feedback about the deniability feature and your experience with the interface.

B.3 Post-Survey

(1) Participant ID
(2) On a scale of 1 to 4, to what extent do you believe that Roommate B paid the bill in this scenario? 1 - Strongly disagree, 2 - Disagree, 3 - Agree , 4 - Strongly agree
(3) On a scale of 1 to 4, to what extent do you believe the statement (Roommate A criticized Roommate B of being irresponsible) in this scenario? 1 - Strongly disagree, 2 - Disagree, 3 - Agree , 4 - Strongly agree
(4) What was the most significant factor that influenced your opinion on the verdict?
    a. The available evidence in the messaging app
    b. The credibility of the parties involved in the conversation
    c. The context and tone of the messages
    d. My personal beliefs about the criminal justice system

e. Fact that the text messages are editable
f. The level of emotional tension or conflict in the conversation
g. Other

(5) According to you, how trustworthy are in-app messages? 1 - Untrustworthy, 2 - Somewhat untrustworthy, 3 - Somewhat trustworthy, 4 - Trustworthy
(6) On a scale of 1-4, to what extent do you believe that users should be able to change their in-app message records for privacy reasons? 1 - Strongly disagree, 2 - Somewhat disagree, 3 - Somewhat agree, 4 - Strongly agree
(7) What possible reasons do you think someone might have for wanting an editing feature in their own local mobile within an in-app messaging platform? Note that the edits are not visible to the other person involved in the conversation. Please select all that apply.
a. Privacy
b. Accountability
c. Security
d. Legality
e. Record of conversation
f. Communication effectiveness
g. Ease of use
h. Other
(8) How useful do you think this feature (editing message records)be in real-life situations? Not at all, Somewhat, Very, Extremely
(9) How much would YOU (not the other person) like to have the ability to edit messages (i.e., have deniability) in your local in-app messaging platform? Not at all, Somewhat, Very, Extremely
(10) How much would you like EVERYONE to have the ability to edit messages (i.e., have deniability) in their local in-app messaging platform? Not at all, Somewhat, Very, Extremely
(11) How likely are you to use an app where EVERYONE has the ability to edit messaging records? Not at all, Somewhat, Very, Extremely
(12) Imagine you are in the following scenarios. In which of the scenarios, if you wanted, would you consider using an editing feature in an in-app messaging platform?
a. Discussing personal or sensitive matters with friends, family and spouse
b. Discussing confidential work-related matters with unauthorized personals (colleagues, friends, family etc.)
c. Discussing confidential work-related matters with authorized personals (manager, clients, etc.)
d. Discussing financial information with financial institution or bank
e. Discussing information with your lawyer, legal professor or law enforcement
f. Communicating sensitive information with individuals or groups who are located in a different country with different legal or cultural norms
g. Communicating sensitive information as a journalist to protect sources
h. Communicating sensitive information as a whistleblower to report misconduct or illegal activity
i. Not needed
j. Other
(13) On a scale of 1 to 4, how often would you use the message editing feature, if it was available in your messaging app? 1 - Never, 2 - Rarely, 3 - Occasionally, 4 - All the time

(14) On a scale of 1-4, if the messaging app you are currently using provided a feature to edit message records in you local messaging app, to what extent would this influence your decision to continue using the app? 1 - Definitely stop using the app, 2 - Probably stop using the app, 3 - Probably continue using the app, 4 - Definitely continue using the app
(15) On a scale of 1-4, if another messaging app provided a feature to edit message records, to what extent would this influence your decision to switch to that messaging app? 1 - Not switching to the other app , 2 - Unlikely to switch to the other app, 3 - Very likely to switch to the other app, 4 - Definitely would switch to other app
(16) On a scale of 1-4, how useful do you find the option to freeze certain in-app messages and prevent them from being edited? 1 - Not useful at all, 2 - Somewhat not useful, 3 - Somewhat useful, 4 - Extremely useful
(17) On one hand, deniability empowers individuals to protect their privacy and maintain control over their personal conversations. Yet, on the other hand, it raises concerns about its potential misuse by criminals or wrongdoers to conceal malicious activities.

Now, it's your turn to take a stance! Share your perspective on deniability and its implications.

(1) Do you believe that the benefits of privacy outweigh the risks of potential misuse? Or do you think that the potential for criminal exploitation is too significant, warranting limitations on deniability features?

There is a final survey that is just to get some demographic details from you and general questions. You also have the option in each question to decline to answer.

(1) Participant ID
(2) What is your level of technology proficiency? 1 - Not Proficient at All, 2 - Slightly Proficient, 3 - Moderately Proficient, 4 - Very Proficient, 5 - Extremely Proficient
(3) To what degree are you knowledgeable about computer security? Not at all, Somewhat, Very, Extremely
(4) How much do you care about privacy? Not at all, Somewhat, Very, Extremely
(5) Do you use messaging app? (Eg: WhatsApp, Signal) Yes, No
(6) Which of the following messaging apps do you use?
WhatsApp, Signal, SMS/MMS (standard text messaging), Facebook Messenger, Instagram Direct, iMessage (Apple), SnapChat, WeChat, Viber, Skype, Telegram, Discord, Hangouts (Google), Slack, BlackBerry Messenger (BBM), GroupMe, Line, Other

(7) How long have you been using messaging apps?
a. Less than 1 Year
b. 1-2 years
c. 2-5 years
d. More than 5 years
e. Other
(8) On scale of 1 to 5, how frequently do you use messaging apps? 1 - Never, 2 - Rarely, 3 - Sometimes, 4 - Frequently, 5 - Always

We appreciate your valuable input and your willingness to participate in this part of the study. If you have any questions or need assistance, please feel free to ask our research team.

C CODE BOOK FOR OPEN-CODING

Code	Definition	Key Points	Example
Deniability Stance - Yes/No			
Yes (Privacy)	Measures and practices that ensure personal information and communications are protected from unauthorized access.	Participants in favor of deniability, cited the need to protect sensitive personal information, avoid surveillance, and maintain control over their digital records.	*"I however, also see the potential to do good especially in the form of allowing for freedom of speech that is not prosecutable by the governments trying to prevent it."*
No (Accountability)	Mechanisms in place to ensure that actions and decisions can be attributed to individuals or entities and that they can be held responsible for their actions.	Participants who did not seek deniability, emphasized the importance of transparency, responsibility, and trustworthiness, need for clear records to ensure accountability.	*"I believe words have power. Once we say something, we would need to stick to it."*
Understand Deniability (Yes or No)	Participants explicitly acknowledge and understand the concept and implications of deniability in communications.	Scenarios of deniability usage, facts about deniability, clear idea and/or examples	*"I believe that whether benefits outweigh the risks of potential misuse depends on specific situation. For example, if people are sharing some sensitive information and don't want other people to see it, the benefits do outweigh the risks. However, if people edit the message because they want to hide the evidence which prove their inappropriate or even illegal actions, the risks of potential misuse definitely overweigh the benefits."*

Table 5. Code book used for the open-coding for the qualitative data.