Omer Akgul*, Ruba Abu-Salma, Wei Bai, Michelle L. Mazurek, Elissa M. Redmiles, and Blase Ur

Secrecy, Flagging, and Paranoia Revisited: User Attitudes Toward Encrypted Messaging Apps

Abstract: With the popularity of tools like WhatsApp, end-to-end encryption (E2EE) is more widely available than ever before. Nonetheless, user perceptions lag behind. Users often do not understand E2EE's security properties or believe them sufficient. Thus, even users with access to E2EE tools turn to less-secure alternatives for sending confidential information. To better understand these issues, we conducted a 357-participant online user study analyzing how explanations of encryption impact user perceptions. We showed participants an app-store-style description of a messaging tool, varying the terminology used, whether encryption was on by default, and the prominence of encryption. We collected perceptions of the tool's security guarantees, appropriateness for privacy-focused use by whom and for what purpose, and perceptions of paranoia. Compared to "secure," describing the tool as "encrypted" or "military-grade encrypted" increased perceptions it was appropriate for privacy-sensitive tasks, whereas describing it more precisely as "end-to-end encrypted" did not. Prior work had found an association between the use of encryption and being perceived as paranoid. We found this link minimized, but still partially applicable. Nonetheless, participants perceived encrypted tools as appropriate for general tasks.

Keywords: PET adoption, Paranoia, Human factors in security and privacy

1 Introduction

The availability of encryption to non-expert users has increased dramatically in recent years, as popular messaging tools like WhatsApp and iMessage have deployed end-to-end encryption (E2EE). Other tools, including Signal and Telegram, have launched with security (particularly E2EE) as an explicit selling point [15]. These tools have overcome what was previously the most important usability challenge in encryption, key management, by leveraging centralized key-directory services. Building encryption into tools that are already popular, rather than requiring users download security-specific tools, has also mitigated some adoption challenges.

Nonetheless, this newfound encryption for the masses has not been a panacea for security. Users often do not realize that their messages are end-to-end encrypted, do not understand the security properties this implies, or do not trust that this security is sufficient [6, 14, 23]. As a result, even when users already use an E2EE communication tool, many will turn to less-secure alternatives like e-mail and SMS when they need to send confidential information [5, 23].

Toward addressing these challenges, we report on an online user study analyzing how a messaging tool's initial description of its encryption features impacts user perceptions. We presented 357 participants an app-store-style description of a messaging application. In a between-subjects protocol, we varied: (i) how encryption was described (*security term*, including "end-to-end encrypted," "secure," and "military-grade encrypted"); (ii) whether messages were encrypted by default or only upon request (*defaultness*); and (iii) whether encryption was the first feature mentioned or just included in the middle of a larger feature list (*priority*).

We specifically investigated the effects of varied descriptions on participants' perceptions (**RQ1**) of the tool's security against potential adversaries (e.g., the government, the people who made the tool), (**RQ2**) whether the tool is appropriate for people who value their privacy, and (**RQ3**) the relationship between use of the tool and paranoia. Our questionnaire was inspired both by the aforementioned shortcomings in how users perceive E2EE messaging tools, as well as by Gaw et al.'s influential 2006 study of the use of encryption in an activist organization [22]. That study found that "users saw universal, routine use of encryption as paranoid" and did not fully understand the threats encryption mitigates, inspiring us to revisit those themes in a world in which E2EE is more widely available than ever before.

We found that two of the factors we varied impacted perceptions in nuanced, important ways. Compared to "secure," describing the tool as "encrypted" or

***Corresponding Author: Omer Akgul:** Affil, E-mail: akgul@cs.umd.edu

"military-grade encrypted" increased perceptions that the tool was appropriate for privacy-sensitive tasks. In contrast, describing it more precisely as "end-to-end encrypted" did not have that effect, even though E2EE tools typically offer better security properties than those that offer more basic encryption. This finding may help explain why users turn from E2EE tools to less-secure alternatives for sending confidential information [5, 14]. Participants were more likely to think users of a "military-grade encrypted" tool are paranoid, even though they were uncertain of the (nebulous) term's meaning.

Given the negative association between encryption and paranoia documented in prior work [22, 40], we had hypothesized that a tool encrypting messages by default would make the tool seem less appropriate for general tasks. We did not find this to be the case, however. We only observed a correlation between encryption's defaultness and perceived security against adversaries.

In 2006, Gaw et al. predicted that making encryption automatic might remove some of its social stigma [22]. We found evidence that this is now the case. Even when the tool encrypted messages by default, participants still found it appropriate for general-purpose, non-confidential tasks. Participants appeared to find security features as a benefit, not annoyance [22]. Nonetheless, while some of the association that Gaw et al. observed between encryption and paranoia appears to have been mitigated, we still observed some association between and specific phrasings of encryption and paranoia.

Interestingly, we also document that users' own psychological paranoia levels have influence on how strong they think the secure communication tools are against adversarial threats, how much privacy-oriented utility they provide, and how paranoid they think using such tools are.

2 Background and Related Work

In this section we discuss related work on the usability of secure communication tools, the importance of mental models to establishing trust in these tools, and the importance of social factors in their adoption. Further, because we explore the connection between individuals' own levels of paranoid thoughts and their perceptions of secure messaging, we provide a brief overview of psychological definitions of paranoia.

Usability of secure communication Originally, studies of secure communication focused heavily on usability of encrypted email. In their seminal 1999 paper, Whitten and Tygar demonstrated usability problems with PGP 5.0 and argued that visual metaphors were needed to help users develop valid mental models of encryption tools [37]. In a similar study, Garfinkel and Miller found that automating key management and creating a more usable interface could improve email encryption outcomes [21]. More recently, Ruoti et al. compared the user experience of multiple mail systems: one optimized for maximum usability, one for making security transparent and one as a hybrid [31, 32]. They argued that the hybrid approach is a reasonable tradeoff. Similarly, Bai et al. found that users see key-directory systems as "good enough" when compared to manual key-exchange systems [8]. When key-directory services are used, users can verify they have the correct key for another user in authentication ceremonies. Unfortunately, these can be difficult [25] and slow [25, 36]. Furthermore, users may not understand these ceremonies' role in providing stronger security guarantees [36]

Over the last decade, end-to-end encryption has been widely adopted in instant messaging systems, leading researchers to investigate the usability of these systems. In particular, researchers have focused on the difficulty of understanding and performing authentication ceremonies [25, 34, 36]. Abu-Salma et al. also note that UI inconsistencies and technical jargon make it difficult use these tools correctly and securely [4].

In this work, we explore factors related to perception and adoption of these tools, rather than their usability explicitly.

Mental models and trust Reachers have found that some users do not trust secure communication tools, in part because of mental models that may be misaligned with the underlying technologies. Wu and Zappala identify concerns such as perceiving encryption for personal use as paranoid and doubting its strength [40]. Other researchers have found that users overestimate the strength of adversaries [14], find SMS or landline phone calls more secure than end-to-end-encrypted communications [5, 6], or simply do not trust that chat apps can be secure[23]. On the other hand, strong design choices can contribute to well-aligned mental models, e.g. about deletion of messages [33]. Preliminary attempts to clarify misaligned mental models show promising results [7] but need further research. While misaligned mental models are not the focus of our study,

we do further confirm prior findings, such as ways in which users describe ciphertext [40].

Adoption and social factors Much research suggests that social factors are critical to secure-communication tool adoption. In early work, Gaw et al. established that members of an activist group saw encryption as useful only for very secret, highly important communications (*secrecy*) [22]. Overuse of encryption was seen as suspicious or paranoid (*paranoia*), as well as potentially annoying when misleading recipients about urgency (*flagging*). The authors argued that automated key generation and distribution systems would help to improve these social factors. Fourteen year later, we revisit this work to examine the current state of secrecy, flagging, and paranoia with respect to adoption of encrypted communication.

Social factors have been shown to be influential in secure behavior adoption broadlky [10–12]. More specifically, De Luca et al. found that although privacy protections might have a small role, peer influence is the top factor in secure messaging adoption, even for tools primarily marketed for privacy [13]. Through qualitative means, Abu-Salma et al. similarly found small user bases, lack of interoperability, and low quality of service contribute to lack of adoption [6].

We add to this work by focusing on how the security description of a messaging tool affects user perceptions. We also expand the discussion of paranoia to include not only whether use of encryption is perceived as a paranoid [22, 40] , but also how an individual's own level of psychological paranoia contributes to their perceptions of secure messaging.

Defining and measuring paranoia

Psychological research has established a loose hierarchy of paranoia. At lower levels, individuals exhibit social concerns and *thoughts of reference*, or believing that other people's actions or conversations focus on you. At higher levels, individuals experience thoughts of mild, moderate, and severe threats directed at themselves [19]. These are sometimes operationalized as two factors: *thoughts of social reference* (generally milder paranoia) and *thoughts of persecution* (generally more severe) [39]. Elevated levels of thoughts of reference usually build up to thoughts of persecution [19]; however, the two can also exist independently [35].

Startup and Startup argue that beliefs of surveillance are significantly associated with persecutory thoughts but not with thoughts of reference; further, a vast majority of individuals with persecutory delusions cope with such thoughts by adopting "security behaviors" such as avoiding social gatherings and trying to anonymize themselves [18, 35]. These findings suggest that individuals' susceptibility to these thoughts may relate to their perceptions of secure communications.

Psychologists have developed many metrics for measuring paranoia [16, 20, 24]. We administer a self-report questionnaire, the Revised Green et al. Paranoid Thoughts Scale (R-GPTS) [20]. This scale concisely and separately measures thoughts of reference and thoughts of persecution.

3 Methods

To investigate our research questions, we designed a survey study (n=357) using mock app-market description pages we created for a fictional secure messaging app called *Inara*. In the survey, we investigated how differences in the description of the app's security features affected participants' impressions of the app's security, as well as suitability for both general-purpose and especially privacy-relevant tasks. The study was approved by our organization's ethics review board.

In the following subsections we describe our experimental conditions, the questionnaire, our recruitment process, our data analysis approach, and limitations of the study.

3.1 Experimental conditions

Descriptions of Inara vary across three key variables, as follows.

Security term — the high-level security mechanism mentioned in the description — had four possible options: "secure communications"(SEC), "encrypted communications"(ENC), "end-to-end encrypted communications"(E2EE), and "military-grade encrypted communications"(MGE). We varied the security term to explore whether these terms have different connotations for users.

We also varied *defaultness*: whether the security term was described as "always" on "by default" (*default*) or whether the description mentioned that users could "Turn on [security term] by just one click" (*manual*). This variable was designed to evaluate whether on-by-default security suggests to users that an app is primarily designed for special circumstances rather than general-purpose communications.

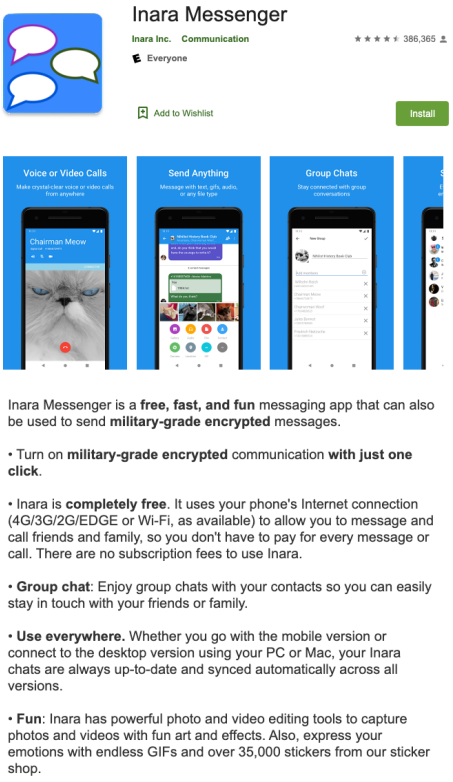


Fig. 1. App description for the military-grade encrypted, manual, high-priority description version of *Inara*

Finally, we varied the *priority* with which the security mechanism was emphasized in the app description. For *high* priority, we mention security term in the first sentence of the description, and defaultness is the top feature listed among many app features. For *low* priority, we do not include security term in the first sentence, and the defaultness statement appears toward the end of the feature list.

We tested all eight combinations of security term and defaultness. To keep the number of conditions manageable, we varied priority only for E2EE, our default security term. All other security terms were tested using the high-priority version only. As with defaultness, this variable investigates perceptions about whether prioritizing security makes an app less palatable for general-purpose use.

In order to make the application look realistic, we mimic the layout used in the Google Play Store application-description interface. We based our design on a common pattern seen in popular messaging applications (such as Whatsapp, Signal, Viber, Slack, and Facebook Messenger, among others): summarizing the focus of the app with one sentence and then listing (usually with bullet points) many relevant features of the app. Thus, we include (in all versions) mainstream features such as being free, multi-platform, supporting calls, supporting group chat, and supporting various multimedia options. An MGE, manual, high-priority version of the description (as presented to participants) is shown in Figure 1.

All the examined conditions were inspired by real-world privacy tool descriptions. For instance, As of writing this paper, Signal's google play store description page ([1]) closely mirrors E2EE, on-by-default, priority. Viber Messenger would be SEC/E2EE, on-by-default, not priority [3]. Telegram has an mix of multiple description versions: it could be considered E2EE, not default, not priority or secure/encrypted, default, priority [2]. Although not used frequently in popular messaging applications, military-grade encryption is commonly used by market leaders to describe other privacy tools such as commercial VPNs.[1] It is important to note that while ENC and E2EE have fairly precise meanings, both SEC and MGE are vague and could mean many things.

3.2 Questionnaire

After providing consent, participants were shown one, randomly assigned, Inara description. On the same page, we asked three comprehension questions, designed to ensure the participant paid attention to the description.

Next, we address RQ1 by asking participants to select, using multiple-choice, multiple-answer responses, who they think would like to use Inara, and for which purposes. Answer choices related to both general-purpose communication (e.g., "People who need to keep in touch with a large group of friends," "Making plans") and more privacy-critical communication (e.g., "People who have something to hide," "Sharing health information/diagnoses/medications"). Participants were allowed to select as many answers as they wished.

In the next section, we asked a series of Likert-scale questions assessing whether Inara is suitable for people who need privacy, whether it seems secure, and (relating to RQ3) whether people who might use it are paranoid. These were followed by free-response questions about

1 NordVPN, which has the biggest market share of the privacy-focused commercial VPN market (https://www.pcmag.com/news/nordvpn-dominates-vpn-market-share-and-that-will-likely-continue), has a webpage dedicated to military-grade encryption (https://nordvpn.com/features/military-grade-encryption/)

the upsides and downsides participants perceived regarding Inara.

In the next section, designed to address RQ2, we ask a series of questions about how likely it is that different possible adversaries could intercept or otherwise interfere with Inara communications. These adversaries — selected based on prior work investigating attitudes toward end-to-end encryption [5, 6, 8] — include "someone with a strong computer science background" (CS), "people who work at Inara" (EMP), "the United States government" (GOV), and "your Internet Service Provider" (ISP). For each adversary, we ask six questions about different capabilities.

Next, we ask the participant to explain, in their own words, their understanding of the security term they saw in the app description and rate how comfortable they were explaining the term.

Finally, we administer both sections (referred to as *paranoia.reference* and *paranoia.persecution*) of the R-GPTS paranoia scale (more in section 2) and ask about general demographics. As a proxy for tech-savviness, we ask how frequently the participant is asked by family or friends for computer or technology advice.

Suitability for privacy tasks In order to investigate RQ1 we analyze Likert-scale responses to "People who care about their privacy would use Inara." We also measure how many privacy-sensitive options the participant selected when answering "Who do you think would be interested in using Inara", and "Which of the following can Inara be used for?" (referred to as WHO and WHAT-FOR from this point on).

Security against adversaries For RQ2 we analyze Likert-scale responses to "Inara seems secure," as well as Likert-scale responses to the adversary-capability questions. We sum all six capability questions for each adversary into a single total,[2] leaving us with four adversary scores per participant.

Perception that using Inara is paranoid For, RQ3 we analyze Likert-scale responses to "People who might use Inara are paranoid."[3] We also measure how many general-purpose (not specifically privacy-sensitive) options the participant selected for WHO and WHAT-FOR.

3.3 Piloting and expert reviews

To refine and validate our survey instrument, we conducted five cognitive interviews with demographically-diverse lay users and five expert reviews with computer security and privacy researchers with survey expertise, as well as our institution's Research Ethics consultant. Cognitive interviews are conducted to pre-test questionnaires and glean insights into how survey respondents might interpret and answer questions [38].

Finally, we piloted the survey on 20 Prolific participants to validate survey flow and randomizations, check for floor and ceiling effects, and identify any other possible abnormalities. The pilot indicated no major issues; therefore, we continued with the deployment of the survey.

3.4 Recruitment

Participants were recruited through an online crowdsourcing platform, Prolific [4]. We used Prolific's built-in prescreening tool to select participants who live in the United States, are 18 or older, and have 95% approval rate on the platform. The study was titled "Messaging App Study" in order to minimize selection bias.

Participants who completed the study received $2, for an average hourly wage of $7.60. Although we eliminated 18 responses (discussed in section 4.1), participants with at least two incorrect comprehension answers and/or non-sensical free-response answers were not paid (n=7).

3.5 Analysis

We analyze quantitative responses using regression models. For Likert scores, we use ordinal logistic regression, appropriate for ordinal data. For the counts of WHO and WHAT-FOR options, as well as adversary scores, we use linear regression.

For each regression, we consider several input variables: security term, defaultness, security priority, paranoia.reference, paranoia.persecution, how often the par-

2 Before summing Likert questions, we validated that they could be combined reliably using Cronbach's α, a measure of inter-item correlation. We found $\alpha > 0.9$ in all cases, indicating good reliability.

3 We use "paranoid" here in the colloquial sense we expect participants to understand, rather than the clinical sense described in Section 2.

4 https://www.prolific.co

ticipant gives tech advice, and participant age. We include tech advice as a proxy for tech savviness, to understand whether more tech knowledge affects perceptions of encryption and secure messaging. We include age because we hypothesize that perceptions may have changed over time, which may be reflected in different age cohorts.

To avoid overfitting, we construct models with different subsets of these covariates and select the final model with minimum Akaike Information Criterion (AIC), a measure of fit. AIC is recommended when searching for a model that is explanatory of the data without including unnecessary variables [9]. We only consider models that include both security term and defaultness, as these are our main variables of interest.

To compare participants' confidence in their own definitions of their assigned security term, we consider the response options as an ordinal scale: "Yes, I have heard of the term [security term] and I feel confident explaining what it means.", "Yes, I have heard of the term [security term] However, I do not feel confident explaining what it means," and not "No, I have not heard of the term [security term]"). After observing a significant omnibus test (Kruskal-Wallis χ^2, $p = 0.0005$), we ran pairwise tests (two-tailed Mann-Whitney U), with Bonferroni correction for multiple testing, comparing SEC to all other security terms.

In order to analyze free-text response questions, we employed exploratory, inductive qualitative coding. For each question, two researchers worked together to create codebooks using the a random 10% of responses, then independently coded the rest in random random batches of 10%. In between each batch, we calculated inter-rater reliability using Cohen's κ. If agreement was not yet sufficient, the researchers iteratively updated the codebook and previously coded responses, then moved to the next batch. Once an acceptable level of inter-rater reliability was achieved, one researcher coded the remaining responses for that question. Our κ values of 0.85, 0.80, and .82 represent "excellent" agreement [17].

3.6 Limitations

Our work has several limitations common to human subjects research in general. We use self-report data, which can suffer from biases related to satisficing [26], social desirability [27], and demand effects [28]. We mitigate this by doing extensive testing (section 3.3) to validate the questionnaire, using comprehension and free-response questions to identify and exclude low-quality data, and by focusing on comparisons among conditions rather than absolute values. Prior work suggests that self-report security data can be useful for establishing directional and comparative effects [30].

Privacy and security are difficult to universally define; differences between participants' perceptions of these concepts could affect our results. To mitigate this, we deploy multiple questions to measure these concepts from different angles, and rely primarily on comparisons among conditions. Our analysis suggests consistency in responses across questions associated with the same concepts.

We use only an app store description page (based on the Google Play store), excluding other ways someone might learn about an app's features, and we make small modifications (such as increasing font size for readability) to real-world description designs. Further, many realistic descriptions use more than one security term to describe an app. We believe our approach effectively balances realism (to maximize generalizability) with ease of comparison and improved participant attentiveness.

Our participants are sampled from a crowd-sourcing platform. As expected, they are younger and more educated than the overall U.S. population, somewhat limiting generalizability. On the other hand, prior work has observed that crowd-worker samples can be quite representative of the U.S. population when it comes to privacy and security related topics [29].

4 Results

In this section, we first provide context on our participants and their demographics. We then present quantitative results for each of our three research questions, followed by qualitative results drawn from free-response questions.

A high-level summary of our quantitative results is shown in Table 1, which shows how the different input factors are correlated with outcomes corresponding to our research questions.

4.1 Participants

In total, 375 participants completed the study. Of these, 18 responses were discarded as invalid due to an incorrect answer to the security-related comprehension question, incorrect answers to two other comprehension questions, or non-sensical answers to free-response ques-

	Utility for Privacy			Strength Against Adversaries				Perceived as Paranoid
	Likert	Who	What for	Likert	CS	GOV	ISP	Likert
Description: Encryption	↑	↑	↑	–	–	–	–	–
Description: E2EE	–	–	–	–	–.	–	–	–
Description: Military-Grade	↑	↑	↑	–	–	–	–	↑
On by Default	–	–	–	↓	–	↓	↑	–
Paranoia: Reference	↓	↓	↓	↑	↑	↑	–	–
Paranoia: Persecution	↑	–	↑	–	–	–	–	↑
Give tech advice often	–	–	–	↓	–	–	–	
Age	–	↓	–	–	–	–	–	↓

Table 1. Positive (↑) or Negative (↓) correlations between the output variables and the input variables. Only correlations that are significant ($p < 0.05$) are listed. Note that the strength Likert describes Inara's strength; inversely, CS, GOV, and ISP scores describe adversary strength.

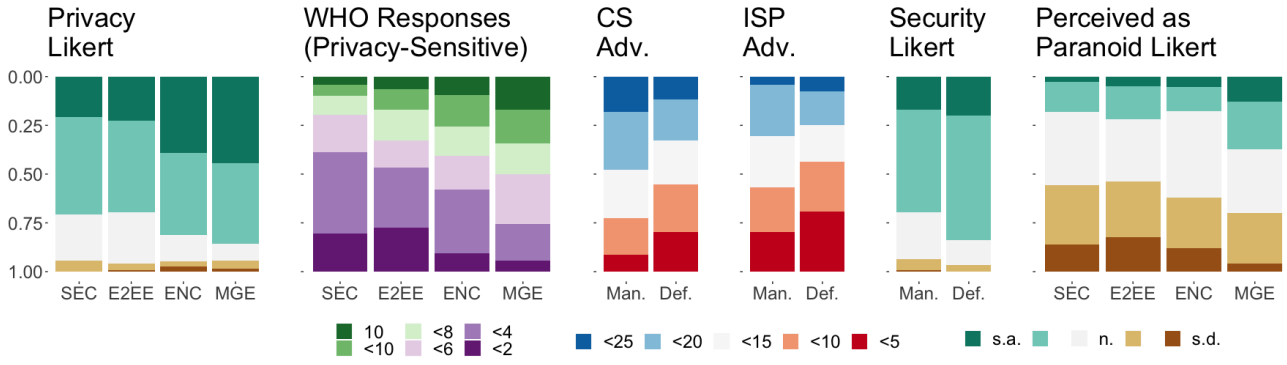


Fig. 2. Utility for Privacy graphs on the left (Privacy Likert and privacy-sensitive WHO responses), graphs for Strengths Against Adversaries in the middle (ISP and GOV adversary-capability scores, Security Likert), Perceived as Paranoid Likert at the right. For privacy sensitive-responses to WHO, counts are binned in ranges of two (except 10). Adversary-capability scores are binned in ranges of five. Likert scales are: strongly agree (s.a.), agree, neither agree nor disagree (n.), disagree, strongly disagree (s.d.). Darker colors indicate extremes.

tions. We use the remaining 357 responses in our analysis. The distribution of these participants over the 10 conditions (Section 3.1) is shown in Table 3; each condition contains between 33 and 38 responses.

As is typical for a crowdworker sample, compared to 2018 American Community Survey data[5], our population is much younger, significantly more educated, less Hispanic, and slightly more Asian. Participant demographics are shown in Table 2. As expected, our population generally aligns with the non-clinical population from the R-GPTS paranoia study [20].

4.2 Using Inara with privacy in mind

Three questions in our survey specifically target participants' perceptions of whether Inara is appropriate for privacy-sensitive tasks: multiple-answer questions about who the participant thinks would use Inara and for what, and a Likert-scale question directly asking if privacy-sensitive people would use Inara. As shown in Table 1, all three questions yield parallel results.

Summary of results In general, "encrypted" and "military-grade encryption" are more likely to be seen as appropriate for privacy than "secure"; "end-to-end encrypted" doesn't differ from "secure." Further, participants with higher scores in paranoia.persecution are more likely to find Inara appropriate for privacy tasks. Somewhat surprisingly, participants with higher paranoia.reference scores have the opposite reaction: they are less likely to associate Inara with privacy.

5 https://www.census.gov/acs/www/data/data-tables-and-tools/data-profiles/2018/

Gender	**Female**	**48.6%**
	Male	**51.3%**
	Other	**0.0%**
Age	18-24	16.5%
	25-29	19.3%
	30-39	35.3%
	40-49	15.1%
	50+	13.7%
Hispanic Origin	No	89.9%
	Yes	10.1%
Ethnicity	White	75.9 %
	Black or African American	12.6 %
	Asian	10.9 %
	American Indian or Alaska Native	2.0 %
	Native Hawaiian or Other Pacific Is.	0.1 %
Education	Completed H.S. or below	11.2 %
	Some college, no degree	26.6 %
	Associate's degree	9.2 %
	Bachelor's degree	37.5 %
	Master's degree or higher	14.8 %
IT background	Yes	22.7%
	No	75.1%

Table 2. Participant demographics. Percentages might not add to 100% due to "other" categories and multiple options selected (ethnicity).

Security term	Priority	Defaultness	Count
Secure	High	Default	37
		Manual	35
Encrypted	High	Default	37
		Manual	37
End-to-end	High	Default	37
		Manual	37
	Low	Default	38
		Manual	33
Military-grade	High	Default	37
		Manual	33

Table 3. The number of participants who saw each description.

	β	95% CI		T-value	p-value
end-to-end	0.969	[−0.554	0.490]	−0.119	0.905
encrypted	1.992	[0.078	1.305]	2.202	0.028*
military-grade	2.412	[0.252	1.516]	2.735	0.006*
on-by-default	1.082	[−0.308	0.467]	0.400	0.689
reference	0.949	[−0.094	−0.011]	−2.460	0.014*
persecution	1.052	[0.014	0.087]	2.737	0.006*

Table 4. Regression table for final selected model for "People who care about their privacy would use Inara" (privacy Likert) regression output. Pseudo-$R^2 = 0.11$. Statistically significant covariates are indicated with *

Privacy Likert For the Likert-scale question "People who care about their privacy would use Inara," the median response for each security term was "somewhat agree," as illustrated in Figure 2. In the final regression model (Table 4), MGE and ENC were associated with significantly stronger privacy responses than the baseline SEC category ($p = 0.006$, $p = 0.028$): on average, participants in the MGE category were 2.4× as likely than SEC participants to increase one step on the Likert scale; ENC participants were 2.0× as likely.

Also according to the regression model, paranoia.reference is associated with a weaker privacy response, while paranoia.persecution is associated with a stronger one ($p = 0.014$, $p = 0.006$). In particular, an increase of 10 points on the paranoia.persecution scale (corresponding to an increase of 1-2 levels in paranoia severity [20]) is associated with a 1.7× increased likelihood to move up a Likert step in agreement that Inara would be used by privacy-sensitive people. In contrast, an increase of 10 in the paranoia.reference scale (again a change of 1-2 levels) yields an estimate of 0.6×.

Defaultness appears in the final model, but is not significant. No other covariates were selected.

Who would use Inara? The analysis of WHO closely mimics the privacy Likert. The mean number of privacy-sensitive options selected was 3.7, 4.2, 5.2, and 6.1 (out of a possible maximum of 10) for SEC, E2EE, ENC, and MGE respectively. Our fitted model aligns with this trend (Table 7 in Appendix B). On average compared to SEC, ENC results in 1.4 more privacy-sensitive options selected ($p = 0.003$), while MGE is associated with 2.2 additional selected options ($p < 0.001$). E2EE is not significantly different from SEC.

Further, similar to privacy Likert, paranoia.reference and paranoia.persecution are significantly correlated with the number of privacy-sensitive options selected ($p = 0.005$, $p = 0.001$ respectively). A 10 point jump in paranoia.reference scores yields 0.9 fewer privacy-sensitive selected and conversely, the same increase in paranoia.persecution scores results in 0.9 more options selected.

The final model also includes defaultness, but it is not significant.

What purposes would Inara be used for? Although generally in line with WHO, analysis of WHAT-FOR (Table 8 in Appendix B) generally shows smaller effect sizes. For instance, although paranoia.reference is signif-

icantly associated with selecting privacy-critical options ($p = 0.019$), a 10 point increase in paranoia.reference decreases the number of privacy-critical options selected by 0.5 as compared to 0.9 in WHO. Similarly, compared to SEC, ENC is associated with 0.8 more selections ($p = 0.040$), MGE with 1.1 ($p = .008$). These results are similar in direction as WHO but with smaller effect size.

Upon further inspection, we see that the relatively low effect size of WHAT-FORmight be due to to ceiling effects. Participants on average chose 9.3 of the 12 total choices ($\sigma = 3.3$) and 46.7% of participants selected all options (for reference, with "who" 7.6% selected all options).

Unlike in the other models, age is also significantly associated with WHAT-FOR. The final model estimates that 10 additional years of age corresponds to 0.2 additional privacy-relevant selections. No other input variables were selected for the final model.

4.3 Perceptions of security against adversaries

We examine perceptions of security using the Likert-scale question "Inara seems secure" as well as the adversary scores generated from the adversary-capability questions. The results were roughly consistent across these metrics.

Summary of results For these metrics, security term did not show any significant effects. On the other hand, participants in on-by-default conditions were more likely to agree Inara is secure and to attribute less power to possible adversaries. In addition, higher levels of reference paranoia correlated with weaker perceptions of security.

Security Likert In general, participants agreed that "Inara seems secure," with median responses of "somewhat agree" for both manual and on-by-default (see Figure 2). Our final logistic regression model (Table 5) shows that defaultness is a significant factor in security perception ($p = 0.008$). Participants in the on-by-default condition were 1.8× more likely than manual participants to increase one point on the Likert scale.

The final model does not find any effect of security term on security perception. Although statistically not significant, the final model also includes paranoia.persecution as a possible factor. No other covariates were included.

	β	95% CI		T-value	p-value
end-to-end	0.862	[−0.709	0.408]	−0.522	0.602
encrypted	1.435	[−0.276	1.000]	1.109	0.267
military-grade	1.508	[−0.246	1.071]	1.225	0.221
on-by-default	1.754	[0.151	0.978]	2.665	0.008*
persecution	1.024	[−0.001	0.048]	1.899	0.058

Table 5. Regression table for final selected model for "Inara seems secure" (security Likert) regression output. Pseudo-$R^2 = 0.072$ Statistically significant covariates are indicated with *

Adversary scores We calculate adversary scores for "someone with a strong computer science background" (CS), "people who work at Inara" (EMP), "the United States government" (GOV), and "your Internet Service Provider" (ISP), finding some variance among adversaries. The mean scores (range 0-24, with 24 being most powerful) were 12.2, 14.4, 13.3, and 9.7 respectively. The final regression model for EMP was poorly fit (adjusted $R^2 < 0.02$), so we do not discuss it here.

Across all adversaries, capability scores are slightly lower on average for on-by-default than for manual. This is reflected in two of the three regression models: for CS, on-by-default is associated with an estimated 2.8-point drop in adversary score ($p < 0.001$); for ISP, it's 1.5 ($p = 0.025$). Defaultness is not significant in the GOV model. (Details are given in Tables 9, 11, 10 in Appendix B.)

The paranoia.reference score also appears in all four final adversary models. For CS, GOV, and ISP, it is a small but significant factor (all $p \leq 0.001$). An increase of 10 points in paranoia.reference score is associated with 1.8 additional points of adversary capability in each case.

In the CS model only, participants who rate themselves as giving computer advice "often" or "always" were associated with a 1.6-point drop in adversary capability relative to those who chose "sometimes," "rarely," or "never" ($p = 0.045$). This makes some intuitive sense: people with more computing experience may realize that a strong CS background by itself is likely insufficient to enable an adversary to break strong protections. This covariate also appeared in the final models for GOV and ISP, but was not significant.

The security term was not significantly correlated with any adversary score. No other covariates (not mentioned above) appeared in any adversary model.

4.4 Perception that using Inara is paranoid

Prior work has suggested that many people view use of encrypted communications tools as paranoid, or only appropriate for illicit or secretive communications [22, 40]. This might manifest as reluctance to use encrypted communications for fear of appearing odd to others. We intended to measure this factor in two ways: with a Likert-scale question ("People who might use Inara are paranoid") and by measuring how many general-communications (not privacy-specific) tasks participants selected in the WHO and WHAT-FOR questions.

Summary of results On average, participants were neutral as to whether or not use of Inara is paranoid. However, participants in the MGE condition were most likely to view Inara users as paranoid. Smaller effects were seen for paranoia.persecution and age: participants with higher paranoia.persecution scores and younger participants were more likely to view Inara users as paranoid.

Unfortunately, the models of general-purpose WHO and WHAT-FOR proved to have poor fit (Adjusted $R^2 < 0.02$), so we do not discuss them further. This is likely due to strong ceiling effects — for WHO, the mean was 4.7 of 6 options, with 47.3% of participants selecting all six. The trend was even stronger for WHAT-FOR: mean 4.5 of 5 options, with 77.3% selecting all five. This suggests that none of our possible covariates are important factors in whether Inara appears appropriate for general-purpose communications.

Paranoid Likert Overall (and for each security term), the median response to the Likert question was "Neither agree nor disagree." However, the final regression model (Table 6) suggests that MGE participants were 2.5× as likely as SEC participants to move up one point on the Likert scale ($p = 0.003$). (ENC and E2EE were not significantly different from SEC). This is by far the largest effect in the model.

In contrast, a 10 point jump in paranoia.persecution increases the likelihood of a higher rating by 1.4× ($p = 0.008$). The other small but significant factor in the final model is age: a 10-year increase in age corresponds to 0.8× the likelihood of increasing agreement ($p = 0.26$). That is, older participants are less likely than younger participants to view Inara use as paranoid. This contradicts our initial hypothesis that older users, with more experience prior to routine encryption of communications, would find secure messaging less socially palatable.

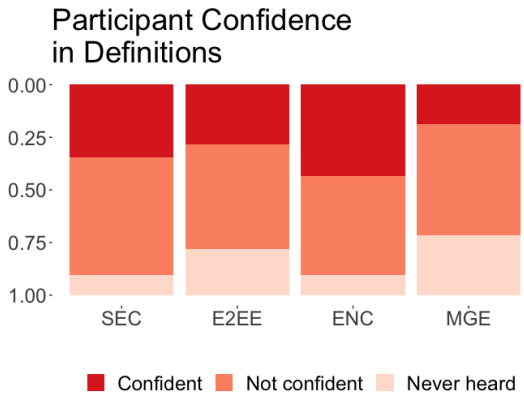


Fig. 3. Participant confidence in explaining the security term assigned to them. Darker colors indicate more confidence.

	β	95% CI		T-value	p-value
end-to-end	0.977	[−0.535	0.488]	−0.090	0.928
encrypted	1.187	[−0.404	0.748]	0.585	0.558
military-grade	2.492	[0.310	1.522]	2.955	0.003*
on-by-default	0.934	[−0.446	0.309]	−0.353	0.724
persecution	1.033	[0.009	0.057]	2.660	0.008*
oft. advice	0.669	[−0.837	0.032]	−1.810	0.070
age	0.981	[−0.035	−0.002]	−2.219	0.026*

Table 6. Regression table for final selected model for "People who might use Inara are paranoid" (paranoia Likert) regression output. Pseudo-$R^2 = 0.13$. Statistically significant covariates are indicated with *

Defaultness and tech advice frequency were also included in the final models but were not found to be significant factors.

4.5 Qualitative Responses

In this section, we describe qualitative responses related to the definitions of the security terms as well as benefits and drawbacks of Inara. We provide percentages as a rough indicator of prevalence. We note that participants can (and often do) report multiple responses. Further, as in any qualitative responses, failure to mention a particular item does not necessarily imply that the participant disagrees with that item; it may simply not have been top of mind when answering.

4.5.1 Participant definitions of security terms

We asked participant to rate their understanding of their assigned security term among confident they understood it, had heard of it but were not confident they understood it, and had not heard of it. We then asked them to define the term in their own words.

Confidence ratings are shown in Figure 3. Notably, compared to SEC participants were significantly (Bonferroni corrected $p = 0.007$) less comfortable defining MGE. (Other security terms were not significantly different from secure.)

We highlight below several key themes from participants' free responses, which broadly align with prior work in mental models of encryption.

Technical details When defining the security terms about a third of our participants (30.3%) mentioned specific technical details consistent with prior work.

Participants described transformation from plain to ciphertext in a variety of ways (12.9%), including "scrambled," "special coding language," "special encoding process," and "they'd get nothing but random letters, symbols, and numbers." Mentions of transformation to ciphertext were most common when defining ENC (37.8%), followed by E2EE(8.5%), MGE (5.7%), and SEC (2.8%). The terms listed align well with prior work [40].

Other participants focused on the need for a secret (8.4%), again aligning with prior work suggesting mental models that parallel symmetric-key encryption (8.4%) [40]. Many referred to the secret as a key (e.g., "cannot be viewed by anyone who does not have the key"), but some seemed to imply a secret algorithm instead ("know how to decrypt it," "need to have a decryption algorithm"). A small fraction (0.8%) mentioned or described asymmetric encryption. Need for a secret was mentioned most frequently for ENC (24.3%), followed by E2EE (7.1%), and MGE (5.7%). It was not mentioned at all for SEC.

Among SEC participants, 30.1% implied or hoped that SEC involved encryption. Examples include "the communications are encrypted in some way" and "I believe communication should be encrypted."

Small fractions of participants listed specific encryption algorithms (0.6%, e.g., AES256), described account protection (e.g., "Only administered users can access it.", 0.6%), and mentioned protection (or non-protection) of metadata (0.9%).

Protection from whom? Again consistent with prior work [5], one-third of our participants (34.5%) mentioned general or specific adversaries when describing what security terms would protect against.

Many participants (14.3%) specifically noted that only the sender and receiver could see messages. As one participant noted, "any third party that acquires the data at some other point in the transmission process has no means of interpreting it." A similar number of participants (12.6%) named adversaries more specific than any possible third party but still fairly vague overall, such as "someone peeping on the network." Further, 2.5% specifically mentioned "hackers.".

Smaller numbers named adversaries similar to those we asked about earlier in the survey (Section 3.2), such as foreign or local government (0.8%, similar to GOV) and the company/application (1.1%, similar to EMP). Similar observations were made in in prior work et al. [14, 23]. Interestingly, protection against the government was exclusively mentioned with MGE (4.3% of those participants).

MGE is for the military Almost a third (32.9%) of MGE participants defined the term as meaning up to the standards of, or even directly used by, the military (31.4%) or government (1.4%, one person). As an example, one participant said, "The encoding is good enough to be used by a very secretive organization such as the military." Unsurprisingly, this definition was not used in any other security term.

One participant captured the general sentiment with, "Messages are coded so that third parties cannot read them, and the encoding is good enough to be used by a very secretive organization such as the military".

4.5.2 Benefits and Drawbacks

We also asked participants to suggest benefits and drawbacks of Inara; we highlight some common responses below.

Privacy and security are not everything Participants noted a variety of benefits and drawbacks unrelated to security or privacy. Almost all participants (93.0%) noted non-security benefits, including that Inara is free (e.g., "It is free and loaded with features.", 47.3%), multi-platform (e.g., "can be used on both desktop and mobile devices.", 18.5%), and seems to have a user-friendly interface (e.g., "It's free, fun and looks easy to use.", 7.8%).

Almost a quarter of participants (22.7%) suggested there were no drawbacks, while just over half (51.5%) mentioned non-security drawbacks. The most common included lack of a large user base to communicate with (e.g., "...convincing others to download it as well", 21.2%), as well as the availability of other similar (or better) apps (e.g., "I have a hundred different ways to communicate as it is", 16.2%). Some were concerned about service quality or usability of the interface (e.g., "Not sure whether how user friendly it is", 11.5%).

These results align well with prior work; fractured user bases and low quality of service have previously been identified as critical factors inhibiting adoption of secure messaging tools [6, 13].

Security features are valuable, if you can trust them Many participants (43.7%) mentioned the specific security term associated with their condition as a benefit. For example, one participant said, "I think the military grade encryption is the primary benefit of using this program." In line with with prior work ([5, 23]) a smaller group of participants (6.2%) explicitly doubted the security or reliability of their security term by name. As one put it, "...there is no way to know how reliable their encryption or overall service quality will be." Notably, SEC was least frequently mentioned as a benefit (34.7% of SEC participants) and most frequently mentioned with doubt (15.3%); however, many participants in all conditions used this term in the general sense, making it difficult to draw a strong inference.

Participants also mentioned their assigned defaultness as a benefit or drawback. Among participants assigned on-by-default, 11.8% saw this feature as a benefit, and none listed it as a drawback. Among manual participants, 11.1% mentioned the inclusion of secure messages as a benefit; however, none explicitly mentioned opt-in as a valuable mechanism. In other words, it was not clear that these participants wouldn't prefer the on-by-default mechanism. On the other hand, a small number (1.7%) saw it as a drawback: "Not having the encryption being the default, but rather opt-in, can be a drawback for privacy. It should be on by default and people would have to opt-out instead."

Other indicators of trustworthiness Participants also described security- and privacy-related issues not directly associated with their assigned conditions.

Positive connotations for security included that Inara is "private" (21.3%), cannot be hacked or compromised (5.0%), "safe" (3.1%), independent from large companies (0.5%), and "anonymous" (0.2%). One respondent noted, "...you have the comfort knowing that your account is safe and private." Another mentioned"sending information, texts, videos without worry they will be hacked." These positive indicators appeared notably more frequently for MGE (48.6%) than for the other security terms (27.7% for E2EE, 23.0% for ENC, and 27.0% for SEC).

On the other hand, participants expressed a variety of doubts related to security and privacy. Some expressed distrust in the (unknown) company (7.8%) or were concerned about personal data collection (5.3%). Participants were also unsure about Inara's privacy and security claims (5.4%), or worried it was vulnerable to hacking (3.4%).

Some examples of participant comments include "I am unfamiliar with the company. How do I know I can trust them?", "It may be collecting information on me and selling to third parties," and "I always worry about the security and the vulnerability of this type of service. How do I know it won't be hacked?"

Considering security features and indicators of trustworthiness, we observe that 78.6% of MGE and 71.6% of ENC participants positively mentioned at least one security feature or an indicator of trustworthiness. In contrast this number was 61.0% for E2EE, and 48.6% for SEC. This observation is in line with our results in section 4.2: ENC and MGE seem to inspire more trustworthiness than SEC.

Similarly, if we consider the how many participants made at least one negative mention of a security feature or expressed doubt related to security and privacy of Inara, we see that MGE, E2EE, and SEC had about the same percentage of mentions (40.0%, 40.4%, and 40.3% respectively) while ENC was slightly lower (32.4%).

Inara is for criminals Again, in a similar theme observed in prior work ([40]), 7.6% of participants mentioned that Inara would be useful for illegal activities. (We note that illegal activities were mentioned as one option in the "who" and "what" questions participants saw before answering the free response questions, which may have caused this answer to be somewhat overrepresented; however, participants did not similarly repeat other listed activities, such as political activism or sending sexts.) One respondent noted, "It's great for the people who actively engage in shady or illegal activities in general." Another expressed concern about law enforcement: "I don't see any personal negatives for myself. However, I can readily picture a major objection from police and governmental departments who would be un-

able to monitor or tap conversations during criminal and national security investigations."

5 Discussion

We compared different descriptions of a secure-messaging tool to understand how the terminology used, defaultness settings, and prioritization of privacy among app features affect users' perceptions of a secure instant-messaging tool. Based on their responses, we revisit findings from 2006 about perceptions of encrypted tool use as paranoid; further, we explore how people's levels of psychological paranoia contribute to perceptions about secure messaging.

MGE is poorly understood but most influential Participants were least confident in defining MGE compared to the other security terms; however, a plurality of users interpreted is used by, or up the standards of, the military or government. (We note that unlike ENC or E2EE, this term has no precise or well-defined meaning.) Participants' intuition seems to provide a strong association with privacy: MGE was also correlated with perception of more privacy-sensitive users and tasks. On the other hand, MGE was also associated with a stronger perception that tool users are paranoid. Somewhat surprisingly, however, MGE had no effect on perception of strength against adversaries. Overall, this phrasing may convey a tool suitable for privacy uses but not necessarily for everyday use, and not necessarily more secure than other tools.

The only other term that seemed to have an effect was ENC, which also had an increased association with privacy relative to the baseline SEC, but to a lesser extent than MGE. Somewhat to our surprise, E2EE did not have significant effect for any of our research questions. This may relate to the relatively high frequency of misunderstanding of this concept, similar to that pointed out in prior work [5, 14].

Defaultness only matters for security against adversaries; priority does not matter at all We initially hypothesized that having security on by default might affect impressions of whether Inara was useful for general-purpose and/or privacy tasks. For example, based on Gaw et al.'s results [22], it seemed plausible that participants might view on-by-default security as overkill. Instead, however, we found that defaultness only correlated with security against adversaries, with manual security seen (appropriately) as less secure than automatic. Our qualitative results align with this finding: no one mentioned always-on security as a drawback, whereas some participants complained about manual security. Qualitatively, defaultness was not mentioned in conjunction with illegal activity or other indicators of illicitness.

Varying whether security received high or low emphasis in the app decription had no effect for any of our research questions; this may be because this nuance was too subtle to register in this experiment (see e.g., Redmiles et al. [30]).

Personal paranoia is a factor We find that higher levels of persecutory thoughts are associated both with stronger belief that Inara is useful for privacy-relevant tasks, and with stronger belief that people who use Inara are paranoid. This aligns well with prior work suggesting that persecutory thoughts can be associated with fear of surveillance and use of coping "safety" behaviors [18]: fears of surveillance might motivate the importance of secure messaging for privacy-sensitive tasks.

In contrast, participants with higher levels of thoughts of reference were less likely to associate Inara with privacy tasks and less likely to believe it provided strong protection from adversaries. We hypothesize that these two correlations, taken together, indicate lack of trust in Inara to properly handle privacy-sensitive communications. Because the two paranoia metrics measure different underlying factors — and particularly differ with respect to fears of surveillance — it is not unexpected that they point in somewhat different directions in this context.

Secrecy, flagging, and paranoia revisited The encrypted communications landscape has changed drastically since Gaw et al.'s seminal paper. Instant messaging is a major mode of communications, and many of the most popular messaging applications (WhatsApp, Facebook Messenger, iMessage) have adopted end-to-end encryption either by default or via opt-in. This has been made possible, in part, by centralizing key management in service-level directories and embracing automation, so that end-to-end encryption is (nearly) transparent to users. In 2006, Gaw et al. suggested that such automation and transparency might improve social factors that hindered adoption.

We find that this prediction has, to an extent, come true, at least for the more general population that we study. Our participants overwhelmingly agreed that Inara could be used for common, general-purpose communication tasks. Security features were seen almost en-

tirely as a benefit, and a key drawback was concern over whether the tool could live up to its security promises. Few to none mentioned that encryption should only be used for secret or important messages. Some in the manual condition requested that security be turned on by deault. We therefore argue that secrecy and flagging are no longer critical social factors.

On the other hand, we find that Gaw et al.'s concept of paranoia — that using encrypted communication might cause one to be perceived as overly fearful or unreasonable – is not entirely gone. Use of the MGE term, for example, was associated with viewing Inara users as somewhat paranoid. A smaller number of free-response answers align with this: one participant did comment that Inara "might make you look suspicious," and others brought up the potential for illicit activities as a drawback, although other concerns seemed more salient. Our findings suggest, then, that while paranoia remains a social factor, it is a minor and likely manageable one.

Implications Our results have implications for designers of secure communications tools. While security features have generally been seen as less important than user base or quality of service (ideas that also recur in our data), perceptions of what a tool is (not) useful for are themselves a social factor than can feed back into development of a user base.

The wording chosen to describe security can influence users' perceptions of the tool, both positively and negatively. Something like "military-grade," in addition to being imprecise, may be overdoing it, making a tool seem fraught. "End-to-end encryption," while more precise, does not appear to mean enough to people to be useful as a security or privacy indicator. Further work is needed to explore how to provide stronger association with privacy without tipping over into paranoia; in our results, "encrypted" came closest to this balance.

We also find that turning security on-by-default has a small but positive effect on participants' perceptions of a tool's security, without activating fears of being seen as paranoid. Making encryption automatic might seem to be an obvious recommendation, but some companies have resisted the idea, either in the name of consumer "choice" or because encrypted messaging cannot easily support features like automating suggestions or ads based on the content of conversations. We hope that demonstrating a positive association with on-by-default will provide an incentive for companies to move in this direction.

6 Conclusion

Nowadays, popular instant messaging tools (e.g., WhatsApp) support end-to-end encrypted communications, making the use of encryption ubiquitous to mainstream users. However, user perceptions of end-to-end encryption and it's security properties lag behind.

In this paper, we designed and conducted an online user study with 357 participants to analyze how a messaging tool's description of it's encryption features impacted participant perceptions. We found that describing a messaging tool as "encrypted" or "military-grade encrypted" — as opposed to "secure" — resulted in participants perceiving the tool as more appropriate for sending information that is sensitive. However, describing the same messaging tool as 'end-to-end encrypted' did not show any effects. Further, we found that participants saw messaging tools that supported encryption by default more secure against adversaries than tools with encryption that they need to turn on. We also observed some association between perceptions of paranoia relating to secure communication tools and specific phrasings of encryption (e.g., , participants were more willing to think that users of a 'military-grade encrypted' messaging tool were paranoid). Finally, we see that users' own psychological paranoia levels affect how secure they think privacy-sensitive communication tools are against adversaries, how much utility for privacy they provide, and how paranoid they think using such tools are.

We recommend that designers of secure communication tools turn ubiquitous security features on-by-default, and carefully chose how they describe their tools as to not be vague but also not seem overly technical and mysterious.

References

[1] Signal private messenger - apps on google play. https://play.google.com/store/apps/details?id=org.thoughtcrime.securesms. Accessed on: 02.29.2020.

[2] Telegram - apps on google play. https://play.google.com/store/apps/details?id=org.telegram.messenger. Accessed on: 02.29.2020.

[3] Viber messenger - messages, group chats & calls - apps on google play. https://play.google.com/store/apps/details?id=com.viber.voip. Accessed on: 02.29.2020.

[4] Ruba Abu-Salma, Kat Krol, Simon Parkin, Victoria Koh, Kevin Kwan, Jazib Mahboob, Zahra Traboulsi, and M Angela Sasse. The security blanket of the chat world: An analytic evaluation and a user study of telegram. Internet Society, 2017.

[5] Ruba Abu-Salma, Elissa M Redmiles, Blase Ur, and Miranda Wei. Exploring User Mental Models of End-to-End Encrypted Communication Tools. In *Proc. USENIX Workshop on Open and Free Communications on the Internet*, 2018.

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

[7] Wei Bai. *User Perceptions of and Attitudes toward Encrypted Communication*. PhD thesis, 2019.

[8] Wei Bai, Moses Namara, Yichen Qian, Patrick Gage Kelley, Michelle L. Mazurek, and Doowon Kim. An inconvenient trust: User attitudes toward security and usability tradeoffs for key-directory encryption systems. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*, pages 113–130, Denver, CO, June 2016. USENIX Association.

[9] Hamparsum Bozdogan. Model selection and akaike's information criterion (aic): The general theory and its analytical extensions. *Psychometrika*, 52(3):345–370, 1987.

[10] Sauvik Das, Tiffany Hyun-Jin Kim, Laura A Dabbish, and Jason I Hong. The Effect of Social Influence on Security Sensitivity. In *ACM Symposium on Usable Privacy and Security*, volume 14, 2014.

[11] Sauvik Das, Adam DI Kramer, Laura A Dabbish, and Jason I Hong. Increasing Security Sensitivity with Social Proof: A Large-scale Experimental Confirmation. In *ACM Conference on Computer and Communications Security*, pages 739–749, 2014.

[12] Sauvik Das, Adam DI Kramer, Laura A Dabbish, and Jason I Hong. The Role of Social Influence in Security Feature Adoption. In *ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 1416–1426, 2015.

[13] Alexander De Luca, Sauvik Das, Martin Ortlieb, Iulia Ion, and Ben Laurie. Expert and non-expert attitudes towards (secure) instant messaging. In *Twelfth Symposium on Usable Privacy and Security ({SOUPS} 2016)*, pages 147–157, 2016.

[14] Sergej Dechand, Alena Naiakshina, Anastasia Danilova, and Matthew Smith. In encryption we don't trust: The effect of end-to-end encryption to the masses on user perception. In *2019 IEEE European Symposium on Security and Privacy (EuroS&P)*, pages 401–415. IEEE, 2019.

[15] Electronic Frontier Foundation (EFF). Secure Messaging Scorecard. https://www.eff.org/secure-messaging-scorecard. Accessed on: 09.07.2016.

[16] Allan Fenigstein and Peter A Vanable. Paranoia and self-consciousness. *Journal of personality and social psychology*, 62(1):129, 1992.

[17] Joseph L Fleiss, Bruce Levin, and Myunghee Cho Paik. *Statistical methods for rates and proportions*. john wiley & sons, 2013.

[18] Daniel Freeman, PA Garety, and E Kuipers. Persecutory delusions: developing the understanding of belief maintenance and emotional distress. *Psychological medicine*, 31(7):1293–1306, 2001.

[19] Daniel Freeman, Philippa A Garety, Paul E Bebbington, Benjamin Smith, Rebecca Rollinson, David Fowler, Elizabeth Kuipers, Katarzyna Ray, and Graham Dunn. Psychological investigation of the structure of paranoia in a non-clinical population. *The British Journal of Psychiatry*, 186(5):427–435, 2005.

[20] Daniel Freeman, Bao S Loe, David Kingdon, Helen Startup, Andrew Molodynski, Laina Rosebrock, Poppy Brown, Bryony Sheaves, Felicity Waite, and Jessica C Bird. The revised green et al., paranoid thoughts scale (r-gpts): psychometric properties, severity ranges, and clinical cut-offs. *Psychological medicine*, pages 1–10, 2019.

[21] Simson L Garfinkel and Robert C Miller. Johnny 2: A User Test of Key Continuity Management with S/MIME and Outlook Express. In *ACM Symposium on Usable Privacy and Security*, pages 13–24, 2005.

[22] Shirley Gaw, Edward W Felten, and Patricia Fernandez-Kelly. Secrecy, flagging, and paranoia: adoption criteria in encrypted email. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 591–600, 2006.

[23] Nina Gerber, Verena Zimmermann, Birgit Henhapl, Sinem Emeröz, and Melanie Volkamer. Finally johnny can encrypt: But does this make him feel more secure? In *Proceedings of the 13th International Conference on Availability, Reliability and Security*, pages 1–10, 2018.

[24] CEL Green, D Freeman, E Kuipers, P Bebbington, D Fowler, G Dunn, and PA Garety. Measuring ideas of persecution and social reference: the green et al. paranoid thought scales (gpts). *Psychological medicine*, 38(1):101–111, 2008.

[25] Amir Herzberg and Hemi Leibowitz. Can johnny finally encrypt? evaluating e2e-encryption in popular im applications. In *Proceedings of the 6th Workshop on Socio-Technical Aspects in Security and Trust*, pages 17–28, 2016.

[26] Jon A Krosnick, Sowmya Narayan, and Wendy R Smith. Satisficing in surveys: Initial evidence. *New directions for evaluation*, 1996(70):29–44, 1996.

[27] Ivar Krumpal. Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity*, 47(4):2025–2047, 2013.

[28] Jonathan Mummolo and Erik Peterson. Demand effects in survey experiments: An empirical assessment. *American Political Science Review*, 113(2):517–529, 2019.

[29] Elissa M Redmiles, Sean Kross, and Michelle L Mazurek. How well do my results generalize? comparing security and privacy survey results from mturk, web, and telephone samples. In *2019 IEEE Symposium on Security and Privacy (SP)*, pages 1326–1343. IEEE, 2019.

[30] Elissa M Redmiles, Ziyun Zhu, Sean Kross, Dhruv Kuchhal, Tudor Dumitras, and Michelle L Mazurek. Asking for a friend: Evaluating response biases in security user studies. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, pages 1238–1255, 2018.

[31] Scott Ruoti, Jeff Andersen, Scott Heidbrink, Mark O'Neill, Elham Vaziripour, Justin Wu, Daniel Zappala, and Kent Seamons. "We're on the Same Page: A Usability Study of Secure Email Using Pairs of Novice Users. In *ACM Conference on Human Factors and Computing Systems*, 2016.

[32] Scott Ruoti, Nathan Kim, Ben Burgon, Timothy Van Der Horst, and Kent Seamons. Confused Johnny: When Automatic Encryption Leads to Confusion and Mistakes. In *ACM Symposium on Usable Privacy and Security*, page 5, 2013.

[33] Theodor Schnitzler, Christine Utz, Florian M Farke, Christina Pöpper, and Markus Dürmuth. Exploring user perceptions of deletion in mobile instant messaging applications. *Journal of Cybersecurity*, 6(1):tyz016, 2020.

[34] Svenja Schröder, Markus Huber, David Wind, and Christoph Rottermanner. When signal hits the fan: On the usability and security of state-of-the-art secure mobile messaging. In *European Workshop on Usable Security. IEEE*, 2016.

[35] Mike Startup and Sue Startup. On two kinds of delusion of reference. *Psychiatry Research*, 137(1-2):87–92, 2005.

[36] Elham Vaziripour, Justin Wu, Mark O'Neill, Jordan Whitehead, Scott Heidbrink, Kent Seamons, and Daniel Zappala. Is that you, alice? a usability study of the authentication ceremony of secure messaging applications. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*, pages 29–47, 2017.

[37] Alma Whitten and J. Doug Tygar. Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0. In *USENIX Security Symposium*, 1999.

[38] Gordon B Willis. Cognitive interviewing revisited: A useful technique, in theory. *Methods for testing and evaluating survey questionnaires*, pages 23–43, 2004.

[39] Gloria HY Wong, Christy LM Hui, Jennifer YM Tang, Cindy PY Chiu, May ML Lam, Sherry KW Chan, WC Chang, and Eric YH Chen. Screening and assessing ideas and delusions of reference using a semi-structured interview scale: A validation study of the ideas of reference interview scale (iris) in early psychosis patients. *Schizophrenia research*, 135(1-3):158–163, 2012.

[40] Justin Wu and Daniel Zappala. When is a tree really a truck? exploring mental models of encryption. In *Fourteenth Symposium on Usable Privacy and Security ({SOUPS} 2018)*, pages 395–409, 2018.

Appendix

A The survey

Consent and validation

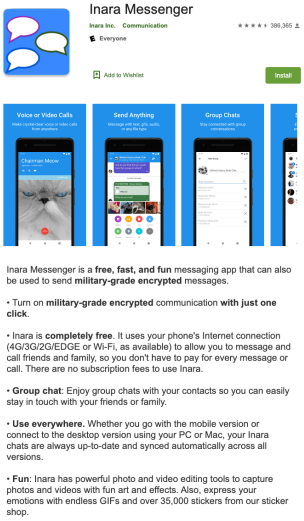
1. *Consent form is shown, and consent is given*
2. In what country do you currently reside?
   - United Kingdom
   - United States
   - Ireland
   - Germany
   - France
   - Spain
   - Other [Free text]

   End survey if not United States
3. Please enter your Prolific ID here
   [Free text]

Part 1: Who would use Inara and Likerts

1. Imagine that you are looking for a new messaging app to communicate with your family members, friends, colleagues, and others. You search in your mobile phone app store (e.g., Apple Store, Google Play Store) and discover an app named Inara.

   To see the app store description of Inara please proceed.
   Description is shown



   Based on the screenshot above, please answer the questions below about Inara.

2. Can you use Inara on a desktop or only on a mobile phone?
   - On a desktop or mobile phone
   - On a mobile phone only
3. How much do phone calls cost in Inara?
   - 2 cents/minute
   - Free except for countries in Europe
   - Always free
4. Which of the following statements is true?
   - To use the [security term] communication in Inara, you need to turn it on.
   - In Inara, the [security term] communication is turned on by default.
   - None of the above

   Page Break

   Based on your understanding of Inara, please answer the following questions.

5. Who do you think would be interested in using Inara? (Select all that apply)
   - People who talk to their family members, friends, and/or colleagues
   - People who live far from their family
   - People who want privacy
   - People who have something to hide
   - People who need to keep in touch with a large group of friends
   - People who want a free method to communicate with their friends
   - People who feel paranoid
   - People who like to use gifs, emojis, etc. in their conversations
   - People who like using messaging apps interchangeably between mobile phones and PC or MAC
   - People who are up to no good (e.g. organized criminals, hackers)
   - People who live in the United States of America
   - People who live under an oppressive government
   - Government employees hoping to protect national secrets
   - Employees of a corporation hoping to keep business secrets confidential from their competitors
   - Doctors and patients
   - Other [Free text]

   Privacy-sensiteve options: {3, 4, 7, 9, 10, 11, 12, 13, 14, 15}

   Page Break

6. Who would you talk to on Inara if you decided to use the app? (select all that apply)
   - Spouse or partner
   - Family members
   - Friends
   - Work colleagues
   - Acquaintances
   - People I have met on other platforms (e.g. Facebook, Twitter, Reddit, Quora), but whom I do not necessarily know
   - Other [Free text]

   Page Break

7. Which of the following can Inara be used for **regardless of whether or not you would do each of these things**? (Select all that apply)
   - Chatting with family members, friends, and/or colleagues
   - Gossiping
   - Making plans
   - Arranging meetings with work colleagues
   - Discussing work
   - Sending the username and password of a personal account
   - Discussing politics
   - Sending bank card details (account number, PIN)
   - Doing illegal things (e.g. buying/selling drugs)
   - Campaigning for a cause (e.g., Black Lives Matter)
   - Sending sexts or nude pictures
   - Sharing health information/diagnoses/medications
   - Other [Free text]

   Privacy-sensiteve options: {6, 7, 8, 9, 10, 11, 12}

   Page Break

8. Who do you think would be interested in using Inara? (Select all that apply)
   - Send/receive text messages
   - Send/receive images
   - Send/receive videos
   - Send/receive file attachments
   - Send/receive voice notes
   - Make phone calls
   - Make video calls
   - Other [Free text]

   Page Break

9. Please answer the following questions.

   What do you see as the major benefits of using Inara?
   [Free text]
10. What do you see as the major drawbacks of using Inara?
    [Free text]
11. To what extent do you agree with the following statements:
    optoins: {Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree}
    - People who might use Inara are paranoid.
    - Based on the screenshot given, Inara looks professionally designed.
    - Inara seems secure.
    - Inara seems fun to use.
    - People who care about their privacy would use Inara.

Page Break

Part 2: Adversary capabilities

1. In this section you will be asked about what different people or groups could do in relation to your Inara communications or your Inara account. Please rate all of the actions that you think each of the people or groups could do. The same question will be asked for four different groups or people.

   Page Break

   The following question is asked six times in total. One for each of: ADVERSARY = {People who work at Inara, Someone with a strong computer science background, The United States government, Your Internet Service Provider (ISP, e.g. Verizon, AT&T)}. The order of adversaries are randomized

2. ADVERSARY could:
   optoins: {Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree}
   - read the content of your Inara messages
   - listen to your Inara phone calls
   - modify your Inara communications
   - impersonate you on Inara
   - determine who you are communicating with on Inara
   - determine how long you are communicating with someone on Inara

   Page Break

Part 3: App usage

1. Which of the following messaging apps have you heard of? (select all that apply)
   [Adium, Silent Phone/Silent text, Blackberry Messenger (BBM), Skype, Blackberry Protect, Snapchat (Direct Snaps), ChatSecure, Surespot, Confide, Telegram, eBuddy XMS, TextSecure, Facebook Messenger, Threema, FaceTime, Viber, Google Hangouts, WhatsApp, iMessage, Wickr, Jitsi, WeChat, Kit Messenger, Yahoo! Messenger, Ostel, Instagram Direct Messages, Pidgin, LinkedIn InMail, QQ, Signal, Other: *free text*]

   Page Break

2. How often do you use the following apps?
   all messaging apps selected in the previous question are listed for each messaging app options: {Have heard of it, but not used it; Used it before, but stopped using it; Use it currently}
   - {messaging app}

   Page Break

   the following is only displayed if more than one app is {Used it before, but stopped using it; Use it currently}
3. You mentioned you used the following apps: [selected apps listed] What made you decide to use multiple apps?
   [Free text]

   Page Break

Part 4: security term definitions

The questions in this part are customized based on the security term assinged to the participant

1. Have you heard of the term [assigned security term]?
   - Yes, I have heard of the term [assigned security term] and I feel confident explaining what it means.
   - Yes, I have heard of the term [assigned security term] However, I do not feel confident explaining what it means.
   - No, I have not heard of the term [assigned security term]
2. As far as you know, what does it mean that communications are [assigned security term]?
   [Free text]

Part 5: Paranoia/Risk

1. Do you feel at risk due to your job duties, political beliefs, or public status?
   - Yes
   - No
   - Prefer not to say

   Page Break

if Yes is selected we display the next question, if not the next question is skipped.

2. The risk I feel is
   - physical risk due to stalking, threats, or attacks from people who do not like what I do or say.
   - cyber risk due to stalking, threats, or attacks from people who do not like what I do or say.
   - Other [free text]

Page Break

3. As far as you know, have you ever had any of these experiences? *optoins: {I have had this experience, I have not had this experience, I don't know}*
   - Had important personal information stolen such as your Social Security Number, your credit card, or bank account information
   - Had medical or health information stolen
   - Had innacurate information show up in your credit report
   - Had an email or social networking account of yours compromised or taken over without your permission by someone else
   - Had difficulty paying off a loan or cash advance that you signed up for online
   - Had been the victim of an online scam and lost money
   - Had experienced persistent and unwanted contact from someone online
   - Had lost a job opportunity or educational opportunity because of something that was posted online
   - Had experienced trouble in a relationship or friendship because of something that was posted online
   - Had someone post something about you online that you didn't want shared

Page Break

4. R-GPTS Part A (as it appears in [20])

Page Break

5. R-GPTS Part B (as it appears in [20])

Page Break

Part 6: Demographics

1. What is your age?
   [numeric free text]
2. What is your gender?
   - Male
   - Female
   - Other [free text]
3. Are you of Hispanic, Latino, or Spanish origin?
   - No
   - Yes
   - Prefer not to say
4. Which of the following best describes your ethnicity? (select all that apply)
   - White
   - Black or African American
   - American Indian or Alaska Native
   - Asian
   - Native Hawaiian or Other Pacific Islander
   - Some other race: [free text]
   - Prefer not to say

Page Break

5. What is your highest level of education? If you are currently enrolled, please specify the highest level/degree completed.
   - Less than 9th grade
   - 9th to 12th grade, no diploma
   - High school graduate
   - Some college, no degree
   - Associate's degree
   - Bachelor's degree
   - Graduate or Professional degree
   - Other [free text]
6. Which of the following best describes your educational background or job field?
   - I have an education in, or work in, the field of computer science, computer engineering, or IT.
   - I do not have an education in, nor do I work in, the field of computer science, computer engineering, or IT.
   - Prefer not to say
7. Have you ever written a computer program?
   - Yes
   - No
   - Do not know
8. How often do people ask you for technology-related advice?
   optoins: {Never, Rarely, Sometimes, Often, Always}

9. Please select the digital security behaviors (or precautions) that are required by your school and/or work, if any.
   - Sending emails with encryption
   - Using a dedicated phone for work tasks
   - Using two-factor authentication to access your work device (Note: Two-factor authentication uses not only a password and a username but also an additional verification code, such as a 4-digit code texted to your phone.)
   - Using two-factor authentication to access your online accounts (Note: Two-factor authentication uses not only a password and a username but also an additional verification code, such as a 4-digit code texted to your phone.)
   - Using a VPN when working on work activities
   - Other [free text]
   - I do not have digital security requirements (or precautions)
   - Prefer not to say

end of survey

B Additional Regression Tables

This Appendix includes regression tables that are not included in the main results section.

	β	95% CI		T-value	p-value
end-to-end	0.400	[−0.407	1.208]	0.975	0.330
encrypted	1.411	[0.488	2.334]	3.007	0.003*
military-grade	2.171	[1.227	3.115]	4.522	< 0.001*
on-by-default	0.282	[−0.307	0.871]	0.942	0.347
reference	−0.092	[−0.156	−0.029]	−2.858	0.005*
persecution	0.088	[0.034	0.142]	3.201	0.001*

Table 7. Regression table for final selected model for who (privacy-sensitive) would use Inara (WHO) regression output. Adjusted $R^2 = 0.097$. Statistically significant covariates are indicated with *

	β	95% CI		T-value	p-value
end-to-end	0.412	[−0.285	1.109]	1.162	0.246
encrypted	0.838	[0.040	1.637]	2.065	0.040*
military-grade	1.109	[0.297	1.920]	2.687	0.008*
on-by-default	0.101	[−0.409	0.611]	0.389	0.698
reference	−0.045	[−0.083	−0.007]	−2.351	0.019*
age	−0.023	[−0.046	−0.001]	−2.011	0.045*

Table 8. Regression table for final selected model for what purpose (privacy-sensitive) Inara would be used for (WHAT-FOR) regression output. Adjusted $R^2 = 0.031$. Statistically significant covariates are indicated with *

	β	95% CI		T-value	p-value
end-to-end	−0.956	[−2.870	0.958]	−0.982	0.327
encrypted	−1.149	[−3.341	1.042]	−1.031	0.303
military-grade	−1.386	[−3.614	0.842]	−1.223	0.222
on-by-default	−2.765	[−4.164	−1.366]	−3.887	< 0.001*
reference	0.176	[0.078	0.275]	3.513	0.001*
oft. advice	−1.577	[−3.120	−0.033]	−2.009	0.045*

Table 9. Regression table for final selected model for adversary-capability score of "Someone with a strong computer science background" (CS) regression output. Adjusted $R^2 = 0.073$. Statistically significant covariates are indicated with *

	β	95% CI		T-value	p-value
end-to-end	0.260	[−1.706	2.227]	0.261	0.795
encrypted	−1.164	[−3.415	1.088]	−1.017	0.310
military-grade	−0.097	[−2.386	2.192]	−0.083	0.934
on-by-default	−1.033	[−2.470	0.404]	−1.414	0.158
reference	0.175	[0.074	0.277]	3.396	0.001*
oft. advice	−1.430	[−3.015	0.156]	−1.773	0.077

Table 10. Regression table for final selected model for adversary-capability score of "The United States government" (GOV) regression output. Adjusted $R^2 = 0.034$. Statistically significant covariates are indicated with *

	β	95% CI		T-value	p-value
end-to-end	−0.613	[−2.436	1.209]	−0.662	0.508
encrypted	−1.648	[−3.735	0.439]	−1.553	0.121
military-grade	−0.846	[−2.968	1.276]	−0.784	0.434
on-by-default	−1.522	[−2.854	−0.190]	−2.247	0.025*
reference	0.175	[0.081	0.269]	3.663	0.000*
oft. advice	−1.173	[−2.643	0.296]	−1.570	0.117

Table 11. Regression table for final selected model for adversary-capability score of "Your Internet Service Provider (ISP, e.g. Verizon, Article AT&T)" (ISP) regression output. Adjusted $R^2 = 0.047$. Statistically significant covariates are indicated with *