From "Secure" to "Military-Grade": Exploring the Effect of App Descriptions on User Perceptions of Secure Messaging

Omer Akgul
University of Maryland
akgul@umd.edu

Ruba Abu-Salma
King's College London

Wei Bai
University of Maryland

Elissa M. Redmiles
Max Planck Institute for Software Systems

Michelle L. Mazurek
University of Maryland

Blase Ur
University of Chicago

ABSTRACT

Although end-to-end encryption (E2EE) is more widely available than ever before, many users remain confused about its security properties. As a result, even users with access to E2EE tools turn to less secure alternatives for sending private information. To investigate these issues, we conducted a 357-participant online user study analyzing how explanations of security impact user perceptions. In a between-subjects design, we varied the terminology used to detail the security mechanism, whether encryption was on by default, and the prominence of security in an app-store-style description page. We collected participants' perceptions of the tool's utility for privacy, security against adversaries, and whether use of the tool would be seen as "paranoid." Compared to "secure," describing the tool as "encrypted" or "military-grade encrypted" increased perceptions that it was appropriate for privacy-sensitive tasks, whereas describing it more precisely as "end-to-end encrypted" did not. However, "military-grade encrypted" was also associated with a greater perception of tool use as paranoid. Overall, we find that — compared to prior work from 2006 — the social stigma associated with encrypted communication has largely disappeared.

CCS CONCEPTS

• **Security and privacy** → Social aspects of security and privacy;

KEYWORDS

end-to-end encryption; military-grade; app descriptions

ACM Reference Format:
Omer Akgul, Ruba Abu-Salma, Wei Bai, Elissa M. Redmiles, Michelle L. Mazurek, and Blase Ur. 2021. From "Secure" to "Military-Grade": Exploring the Effect of App Descriptions on User Perceptions of Secure Messaging. In *Proceedings of the 20th Workshop on Privacy in the Electronic Society (WPES '21), November 15, 2021, Virtual Event, Republic of Korea.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3463676.3485602

WPES '21, November 15, 2021, Virtual Event, Republic of Korea.

ACM ISBN 978-1-4503-8527-5/21/11…$15.00
https://doi.org/10.1145/3463676.3485602

1 INTRODUCTION

As popular messaging tools like WhatsApp and iMessage have deployed end-to-end encryption (**E2EE**), the availability of encryption to non-expert users has increased dramatically. Other tools, including Signal andbTelegram, have launched with security, particularly E2EE, as an explicit selling point [18]. These tools have overcome what was previously the most important usability challenge in encryption — manual key management,– by leveraging centralized key-directory services. Building encryption into tools that are already popular, rather than requiring users to download security-specific tools, has also mitigated some adoption challenges [4, 51].

Nonetheless, this newfound encryption for the masses has not been a panacea for security. Users often do not realize their messages are end-to-end encrypted, do not understand the security properties this implies, or do not trust that this security is sufficient [4, 14, 24]. 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 [3, 24].

Toward addressing these challenges, we report on an online user study analyzing how a messaging tool's initial description of its security features impacts user perceptions. We gave 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 "encrypted," "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 how the varied descriptions affected participants' perceptions in three ways:

RQ1: Do participants perceive the tool as appropriate for people who value their privacy?
RQ2: How strong do participants perceive the tool to be against different potential adversaries?
RQ3: Do participants perceive using the tool as paranoid?

These questions were 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 encryption in an activist

organization [23]. We revisit their finding that "users saw universal, routine use of encryption as paranoid" now that E2EE is more widely available than ever before.

We found that two of the factors we varied affected perceptions in nuanced, important ways. Compared to "secure," describing the tool as "encrypted" or "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. This finding may help explain why users turn from E2EE tools to less-secure alternatives for sending confidential information [3, 14]. Participants were more likely to perceive users of a "military-grade encrypted" tool as paranoid, even though they were uncertain of the (nebulous) term's meaning.

Given prior findings that use of encryption can seem paranoid [23, 55], we hypothesized that encrypting messages by default would make a tool seem less appropriate for general tasks. We did not find this to be the case; we only observed a positive correlation between keeping the security mechanism on by default (vs turning it on manually) and perceived security against adversaries.

Gaw et al. predicted that automating encryption might remove some of its social stigma [23]. We found evidence this is now the case, as participants appeared to find security features to be a benefit, not annoyance [23]. Nonetheless, we still observed some association between specific descriptions of encryption and paranoia. For instance, the term "military-grade" was correlated with a stronger perception that tool uses were paranoid.

2 BACKGROUND AND RELATED WORK

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 [51]. Similar problems still exist in modern encrypted email clients [32]. Garfinkel and Miller found that automating key management and creating a more usable interface could improve email encryption outcomes [22]. More recently, researchers have argued for new points in the tradeoff space between more security and more usability [8, 39, 40].

As end-to-end encryption has been widely adopted in instant messaging systems, researchers have investigated the usability of these systems. In particular, researchers have explored the difficulty of understanding and performing authentication ceremonies [28, 42, 46, 48, 49]. Abu-Salma et al. also note that UI inconsistencies and technical jargon make it difficult to use these tools correctly and securely [2]. Further, researchers found that unique security and privacy problems appear in the context of group chats [34]. In this work, we explore factors related to perception and adoption of these tools, rather than their explicit usability.

Mental models and trust. Reachers have found that some users do not trust secure communication tools, in part because their mental models may be misaligned with the underlying technologies. Wu and Zappala identify perception of encryption for personal use as paranoid and doubts about its strength [55]. Other researchers have found that users overestimate the strength of adversaries [14], find SMS or landline phone calls more secure than E2EE communications [3, 4], or simply do not trust that chat apps can be secure [24]. On the other hand, strong design choices can contribute to well-aligned mental models, including of message deletion [41]. Preliminary attempts to clarify misaligned mental models were promising [7, 54] but further research offer mixed results [6, 45]. While misaligned mental models are not the focus of our study, we do further confirm prior findings.

Adoption and social factors. Much research suggests that social factors are critical to the adoption of secure communication tools. In early work, Gaw et al. found that members of an activist group saw encryption as useful only for very secret, highly important communications [23]. Overuse of encryption was seen as suspicious or paranoid. The authors argued that automated key management would improve these social factors. We revisit this idea, examining perceptions of secrecy and paranoia within the general population, where encrypted chat apps have become widespread.

Social factors are influential in secure behavior adoption broadly [10–12]. More specifically, De Luca et al. and Abu-Salma et al. found that peer influence significantly outweighs privacy protection in adoption of secure messaging systems [4, 13]. We add to this work by focusing on how the security description of a messaging tool affects user perceptions.

Defining and measuring paranoia. Psychological research has established a loose hierarchy of paranoia. At lower levels, individuals exhibit social concerns and thoughts of reference, believing that other people's actions or conversations focus on them. At higher levels, individuals experience thoughts of mild, moderate, and severe threats directed at themselves [20]. These are sometimes operationalized as two factors: **thoughts of social reference** (generally milder paranoia) and **thoughts of persecution** (generally more severe paranoia) [53]. Elevated levels of thoughts of reference often build up to thoughts of persecution [20]; however, the two can also exist independently [44].

Beliefs of surveillance are thought to be significantly associated with persecutory thoughts but not thoughts of reference. Further, most individuals with persecutory delusions adopt "security behaviors" like avoiding social gatherings and trying to anonymize themselves [19, 44]. These findings suggest that individuals' susceptibility to these thoughts may relate to their perceptions of secure communications.

Psychologists have developed many paranoia metrics [16, 21, 25]. We use the Revised Green et al. Paranoid Thoughts Scale (**R-GPTS**), which concisely and separately measures thoughts of reference and thoughts of persecution [21].

3 METHODS

To investigate our research questions, we designed a survey-based experiment (n=357) using mock app-market description pages we

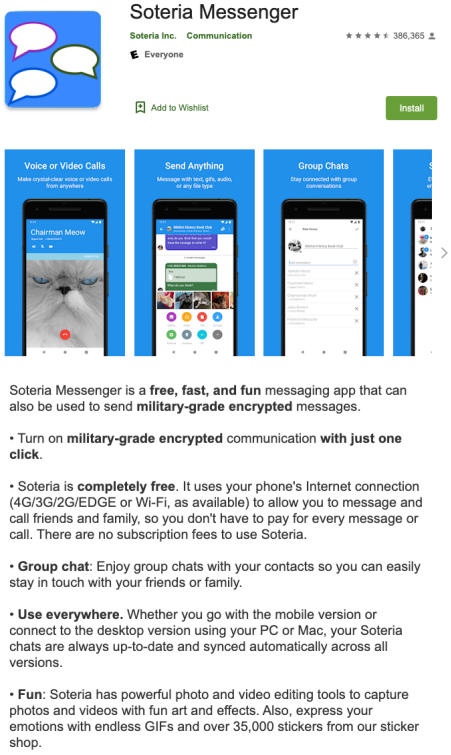


Figure 1: App description for the military-grade encrypted, manual, high-priority description version of Soteria

created for a fictional secure messaging app called *Soteria*. 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 specifically privacy-relevant tasks. We used a mock application to avoid confounding our experiment by invoking participants' (dis)trust in specific brands, which could have a large impact on privacy perceptions [29]. The study was approved by University of Maryland's Institutional Review Board (IRB).

3.1 Experimental conditions

Descriptions of Soteria vary across three key variables, summarized in Table 1. We use abbreviations for each (shown in bold) throughout the rest of the paper.

Security term was the high-level security mechanism mentioned in the description. We explored whether different security terms have different connotations, particularly since much of the security conversation around apps like Signal, WhatsApp, and iMessage has focused on it being E2EE. Security term had four possible options. The first one was intentionally vague: "secure communications" (secure, or **SEC**). Two options, "encrypted communications" (encrypted, or **ENC**) and "end-to-end encrypted communications" (end-to-end encrypted, or **E2EE**), have fairly precise meanings, particularly the latter. We also tested "military-grade encrypted communications" (military-grade, or **MGE**) as an example of a term that is seemingly technical, widely used to describe encryption, but ultimately meaningless.

Defaultness indicated whether the security term was described as "always" on "by default" (on by default, or **ON**) or whether users could "Turn on [security term] by just one click" (manual, or **MANUAL**). This variable was designed to evaluate whether on-by-default security suggests that an app is primarily designed for special circumstances rather than general-purpose communications. This variable revisits a key prediction from Gaw et al. that automating encryption might reduce social stigma [23].

Priority indicated whether the security mechanism was emphasized in the app description. Some E2EE tools are explicitly marketed as secure messengers (e.g., Signal); others are marketed as general-purpose messengers (e.g., WhatsApp, iMessage). Notably, some of these tools use the same E2EE protocols; for instance, WhatsApp uses the Signal Protocol. As a result, from the E2EE perspective, these tools differ not in features, but rather in marketing.[1] For high priority (high, or **HIGH**), we mentioned security term in the first sentence of the description, and defaultness is the top feature listed among many app features. This approach approximated tools for which strong security is the selling point, such as Signal. For low priority (low, or **LOW**), we did not include security term in the first sentence, and the defaultness statement appeared toward the end of the feature list. This approximated general communication tools that, from a typical user's perspective, incidentally offer strong security. As with defaultness, this variable investigated whether prioritizing security makes an app less palatable for general-purpose use.

We tested all eight combinations of **security term** and **defaultness**. To keep the number of conditions manageable and increase participants per experimental group, we varied **priority** only for end-to-end encrypted, our default security term. All other security terms were tested using the high-priority version only.

For realism, we mimicked the layout used in the Google Play Store application-description interface. We based our design on a 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. Thus, we included (in all conditions) mainstream features such as being free, multi-platform, supporting calls, supporting group chat, and supporting multimedia. A military-grade, manual, high-priority version of the description (as presented to participants) is shown in Figure 1.

All conditions were inspired by real-world privacy-tool descriptions. As of writing this paper, the Google Play store description page for Signal [43] closely mirrors *end-to-end encrypted, on by default, high*. Viber Messenger uses both *secure* and *end-to-end encrypted* with on by default, low [50]. Telegram mixes a variety of descriptions, including *end-to-end encrypted*, *secure*, and *encrypted*, together with *on by default*; security is mentioned in the first line but not the first feature, which could be considered a mix of our *low* and *high* priority conditions [47]. Although not used frequently in popular messaging applications, military-grade encryption is

[1] While WhatsApp and Signal share an E2EE protocol, other aspects of Signal, such as its open-source nature and more secure default backup settings, may ultimately make it more appropriate as a security tool [35].

Variable	Value	Abbreviation	Description
Security term	secure	**SEC**	"secure communications"
	encrypted	**ENC**	"encrypted communications"
	end-to-end encrypted	**E2EE**	"end-to-end encrypted communications"
	military-grade	**MGE**	"military-grade encrypted communications"
Defaultness	manual	**MANUAL**	"turn on [security term] with just one click"
	on by default	**ON**	"always on by default"
Priority	low	**LOW**	security not in first sentence; toward end of feature list
	high	**HIGH**	security mentioned in first sentence; top feature

Table 1: The three dimensions of the Soteria description we varied across our experimental conditions, along with short-form abbreviations (bolded) we use in the paper to refer to those particular settings.

RQ	Variable Name	Explanation
RQ1	**Privacy Likert**	Likert-scale response to "People who care about their privacy would use Soteria."
(Suitability	**# privacy users**	Number of privacy-sensitive options selected for who would use Soteria.
for privacy)	**# privacy cases**	Number of privacy-sensitive options selected for what Soteria could be used for.
RQ2	**Security Likert**	Likert-scale responses to "Soteria seems secure."
(Security against	**CS**	Aggregate strength score against "someone with a strong computer science degree"
adversaries)	**EMP**	Aggregate strength score against "people who work at Soteria"
	GOV	Aggregate strength score against "the United States government"
	ISP	Aggregate strength score against "your Internet Service Provider"
RQ3	**Paranoia Likert**	Likert-scale response to "People who might use Soteria are paranoid."
(Perceived	**# general users**	Number of general-purpose options selected for who would use Soteria.
as paranoid)	**# general cases**	Number of general-purpose options selected for what Soteria could be used for.

Table 2: Outcome variables associated with each research question, along with short-form variable names (bolded) we use in the paper. Explained in more detail in Section 3.2.

Variable	Explanation	Baseline
Description variations (see Table 1):		
Security term	Term used to describe the app's security properties	SEC
Defaultness	Whether the app encrypts messages by default	MANUAL
Priority	Whether security features were listed first	LOW
Demographic covariates:		
Thoughts of reference	Participant's paranoia: Thinking of others thinking about them	N/A
Thoughts of persecution	Participant's paranoia: Thinking of others trying to harm them	N/A
Technical expertise	How often the participant is asked for tech advice (less or more often)	LESS OFTEN
Age	The participant's age	N/A

Table 3: Independent variables (IVs) used in our regressions (Section 3.4), including dimensions of the description we varied and demographic covariates. For categorical variables, the baseline is listed.

commonly used by market leaders to describe other privacy tools, such as commercial VPNs.[2]

[2]At the time of data collection, NordVPN, with the largest share of the privacy-focused commercial VPN market [26], had a page dedicated to military-grade encryption (https://nordvpn.com/features/military-grade-encryption/). It now redirects to the equally vague "Next-generation encryption."

3.2 Questionnaire

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

Next, we addressed RQ1 and RQ3 by asking who participants thought would like to use Soteria, and for which purposes. Answer

choices related to both general-purpose communication (RQ3) (e.g., "People who need to keep in touch with a large group of friends," "Making plans") and more privacy-critical communication (RQ1) (e.g., "People who have something to hide," "Sharing health information/diagnoses"). Participants could select multiple answers.

In the next section, relating to all research questions, we asked Likert-type questions assessing whether Soteria was suitable for people who needed privacy (RQ1), whether it seemed secure (RQ2), and whether people who might use it were paranoid (RQ3). These were followed by free-response questions about the perceived upsides and downsides of Soteria.

In the next section, designed to address RQ2, we asked questions about how likely it was that different possible adversaries could intercept or otherwise interfere with Soteria communications. These adversaries — selected based on prior work investigating attitudes toward end-to-end encryption [3, 4, 8] — included "someone with a strong computer science background" (**CS**), "people who work at Soteria" (**EMP**), "the United States government" (**GOV**), and "your Internet Service Provider" (**ISP**). For each adversary, we asked six questions about different capabilities (see Appendix A).

We next asked the participant to explain, in their own words, their understanding of their assigned security term and rate their comfort with explaining it. Finally, we administered both sections (referred to as *thoughts of reference* and *thoughts of persecution*) of the R-GPTS paranoia scale (see Section 2) and asked about general demographics. As a proxy for technical expertise, we asked how frequently the participant is asked by family or friends for computer or technology advice.

Each research question is thus associated with several measurement variables, also known as outcome or dependent variables, summarized in Table 2 and described below.

RQ1: Suitability for privacy tasks. For RQ1, we analyzed Likert-scale responses to "People who care about their privacy would use Soteria." We also measured how many privacy-sensitive options the participant selected when answering "Who do you think would be interested in using Soteria", and "Which of the following can Soteria be used for?" (henceforth termed **# privacy users** and **# privacy cases**).

RQ2: Security against adversaries. For RQ2, we analyzed Likert-scale responses to "Soteria seems secure," as well as to the adversary-capability questions. We summed the six capability questions for each adversary into a single total,[3] leaving us with four adversary scores per participant.

RQ3: Perception that using Soteria is paranoid. For RQ3, we analyzed Likert-scale responses to "People who might use Soteria are paranoid."[4] Complementing RQ1, we also measured how many general-purpose users (**# general users**) and use cases (**# general cases**) the participant selected for "Who do you think would be interested in using Soteria?" and "Which of the following can Soteriabe used for?"

[3]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.

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

3.3 Recruitment and piloting

To refine our questionnaire, we conducted five cognitive interviews with demographically diverse lay users and five expert reviews with researchers with security/privacy and survey expertise, as well as our institution's research ethics consultant [52]. We then piloted the survey on 20 participants from the Prolific crowdsourcing platform[5] to validate survey flow and randomizations, check for floor and ceiling effects, and look for other abnormalities. We found no major issues.

Participants for the main study were also recruited using Prolific. Using Prolific's screen tools, we selected participants who lived in the United States, were 18 or older, and had a 95% approval rate on the platform. We advertised a "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), only participants with at least two incorrect comprehension answers and/or nonsensical free-response answers were not paid (n=7).

3.4 Analysis

We analyzed quantitative responses using regression models. For Likert scores, we used ordinal logistic regression, which is appropriate for ordinal data. For the counts of privacy-sensitive and general options (**# privacy users**, **# privacy cases**, **# general users**, **# general cases**) we used poisson regression, which is appropriate for count data. Finally, for adversary scores, we used linear regression. For goodness-of-fit with linear regressions, we report R^2; we report corrected Aldrich-Nelson pseudo-R^2 for others [27].

For each regression, we considered several independent variables, summarized in Table 3. We included the three condition variables representing aspects of the app description: security term, defaultness, and priority. We also included demographic covariates: thoughts of reference, thoughts of persecution, how often the participant gave tech advice, and participant age. We included tech advice as a proxy for tech savviness, to understand whether more tech knowledge affected perceptions of encryption and secure messaging. We assigned participants to one of two groups, based on their Likert-scale answers: **LESS OFTEN** (never, rarely, sometimes) or **MORE OFTEN** (often, always). We included age because we hypothesized that perceptions might have changed over time, which might be reflected in different age cohorts.

To avoid overfitting, we constructed models with different subsets of these covariates and selected a 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 considered models that included both security term and defaultness, as these are were main variables of interest.

To compare participants' confidence in their own definitions of their assigned security term, we considered the response options as an ordinal scale: having heard of the term and feeling confident explaining it, having heard of it but not confident explaining it, and

[5]https://www.prolific.co

not having heard of it. After a significant omnibus test (Kruskal-Wallis $\chi^2, p < 0.001$), we ran pairwise tests (two-tailed Mann-Whitney U) with Bonferroni correction, comparing secure to all other security terms.

To analyze free-text responses, we employed exploratory, inductive qualitative coding. For each question, two researchers worked together to create a codebook using the a random 10% of responses, then independently coded the rest in random batches of 10%. Between batches, we calculated inter-rater reliability using Cohen's κ. If agreement was not yet sufficient, we iteratively updated the codebook and previously coded responses, then moved to the next batch. Once acceptable reliability was achieved, one researcher coded the remaining responses for that question. Our κ values of 0.85, 0.80, and 0.82 represents "excellent" agreement [17].

3.5 Limitations

Our work has several limitations common to human-subjects research. We use self-report data, which can suffer from biases related to satisficing [30], social desirability [31], and demand effects [33]. We mitigated this by extensively testing the questionnaire, using comprehension and free-response questions to identify and exclude low-quality data, and 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 [38].

Privacy and security are difficult to universally define; differences between participants' perceptions of these concepts could affect our results. To mitigate this, we deployed 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 used only an app store description page, based on the Google Play store, excluding other ways someone might learn about an app's features. We made small modifications (such as increasing font size for readability) to 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.

Typically for Prolific, our sample is younger and more educated than the overall U.S. population, somewhat limiting generalizability. On the other hand, prior work has found crowd-worker samples can be reasonably representative of the U.S. population when it comes to privacy- and security-related topics [37]. We limited our sample to the U.S. to reduce variability, which also limits generalizability.

4 RESULTS

We first describe our participants. We then present quantitative results for each of our three research questions, followed by qualitative results drawn from free-response questions.

Table 4 provides a high-level summary of our quantitative results, showing how the different input factors (independent variables) correlate with outcomes (dependent variables) corresponding to our research questions. The independent variables include both the three dimensions of the description of Soteria we varied across participants and covariates capturing participants' demographics and degree of inherent paranoia.

4.1 Participants

In total, 375 participants completed the study. We discarded 18 invalid responses due to an incorrect answer to the security-related comprehension question, incorrect answers to two other comprehension questions, or nonsensical free-response answers. We analyzed the remaining 357 responses, with 33–38 responses in each of the 10 conditions. (The complete distribution is given in Table 8 in Appendix B.)

Participant demographics are shown in Table 5. As is typical for a crowdworker sample, compared to American Community Survey data[6], our population is much younger, significantly more educated, less Hispanic, and slightly more Asian. As expected, our population generally aligns with the non-clinical population from the R-GPTS paranoia study [21].

4.2 Using Soteria with privacy in mind (RQ1)

Three questions in our survey targeted perceptions of whether Soteria was appropriate for privacy-sensitive tasks: two multiple-answer questions about who the participant thought would use Soteria and why, and a Likert-scale question directly asking if privacy-sensitive people would use Soteria. As shown in Table 4, all three questions yielded parallel results.

Summary of results. The description significantly affected perceptions of whether Soteria was appropriate for privacy-sensitive tasks. Participants would be more likely to see Soteria as appropriate for privacy if it were described as "encrypted" or "military-grade encrypted" than as "secure." Describing Soteria as "end-to-end encrypted" would most precisely describe the strongest security property we studied. However, perceptions of Soteria's appropriateness for privacy did not differ significantly between describing it as "end-to-end encrypted" or as "secure." This result is particularly important given that messaging apps like WhatsApp, iMessage, and Signal all prominently note that they are end-to-end encrypted, which this result suggests is ineffective messaging. Regardless of condition, participants with greater thoughts of persecution were more likely to find Soteria appropriate for privacy. Surprisingly, participants with greater thoughts of reference scores were instead less likely to associate Soteria with privacy.

Privacy Likert. Median agreement with "People who care about their privacy would use Soteria" was "somewhat agree" for each security term (Figure 2). In the final regression model (Table **??**), "military-grade" and "encrypted" were associated with significantly stronger privacy responses than the baseline of "secure" ($p = 0.006$, $p = 0.028$). On average, participants told the app was "military-grade encrypted" were 2.4× as likely than those told it was "secure" to increase one step on the Likert scale. Participants told it was encrypted were 2.0× as likely. We did not observe significant differences between participants in the "end-to-end encrypted" and "secure" conditions. While we required defaultness to be retained in the final model, it was not statistically significant. Priority was not retained, suggesting it was not an important factor.

[6] https://www.census.gov/acs/www/data/data-tables-and-tools/data-profiles/2018/

	RQ1: Utility for Privacy			RQ2: Strength Against Adversaries					RQ3: Perceived as Paranoid		
	Privacy Likert	# Privacy Users	# Privacy Cases	Security Likert	CS*	EMP*	GOV*	ISP*	Paranoid Likert	# General Users*	# General Cases*
Description vs. secure (SEC)											
encrypted (ENC)	↑	↑	↑	–	–	–	–	–	–	↑	–
end-to-end encrypted (E2EE)	–	–	–	–	–	–	–	–	–	–	–
military-grade (MGE)	↑	↑	↑	–	–	–	–	–	↑	–	–
Defaultness vs. manual (MANUAL)											
on by default (ON)	–	–	–	↑	↑	–	–	↑	–	–	–
Priority vs. low (LOW)											
high (HIGH)	–	–	–	–	–	–	–	–	–	–	–
Demographic covariates											
thoughts of reference	↓	↓	↓	–	↓	–	↓	↓	–	–	–
thoughts of persecution	↑	↑	–	–	–	–	–	–	↑	–	–
technical expertise	–	–	–	–	↑	–	–	–	–	–	–
age	–	–	↑	–	–	–	–	–	↓	–	–

Table 4: Statistically significant ($p < 0.05$) positive (↑) or negative (↓) correlations between the outcome and independent variables. Positive (↑) correlations indicate greater privacy utility, greater strength against adversaries, and more perception as paranoid. (Some outcome variables, indicated with *, are shown reversed to maintain consistent direction of correlation.)

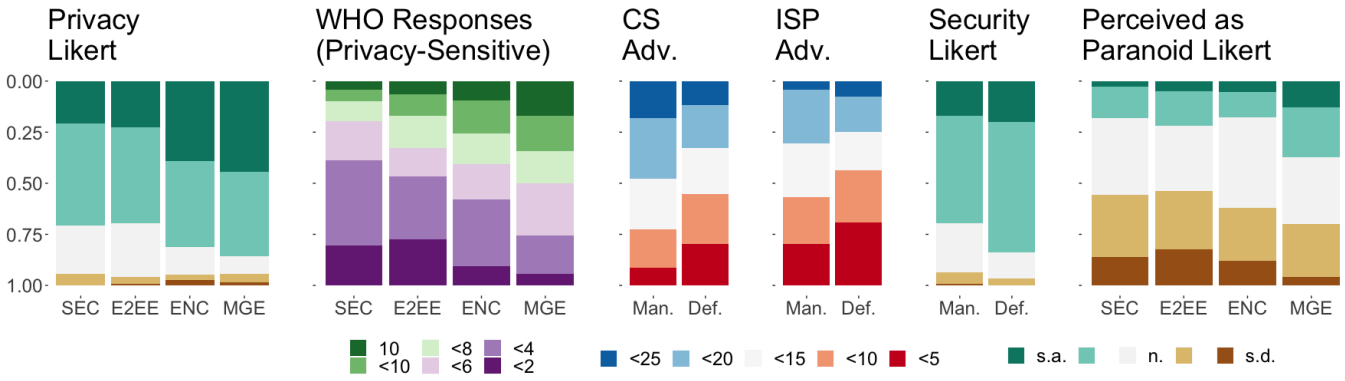


Figure 2: Illustrations of some key quantitative differences: Utility for Privacy on the left (Privacy Likert and # privacy users responses), Security Against Adversaries in the middle (ISP and GOV adversary-capability scores, Security Likert), Perceived as Paranoid Likert on the right. For # privacy users, 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.

The demographic covariates capturing inherent paranoia were also statistically significant. Thoughts of reference was associated with a weaker privacy response, while thoughts of persecution was associated with a stronger one ($p = 0.014, p = 0.006$). An increase of 10 points on the thoughts of persecution scale (corresponding to an increase of 1-2 levels in paranoia severity [21]) was associated with a 1.7× increased likelihood to move up a Likert step. In contrast, an increase of 10 in the thoughts of reference scale (again a change of 1-2 levels) yielded an estimate of 0.6×. No other covariates were retained in the final model.

Who would use Soteria. The analysis of # privacy users closely resembled the Privacy Likert results. The mean numbers of privacy-sensitive options selected were 3.7, 4.2, 5.2, and 6.1 (out of a possible maximum of 10) for "secure," "end-to-end encrypted," "encrypted," and "military-grade," respectively (Figure 2). Our fitted model aligned with this trend (Table 7). On average compared to "secure," "encrypted" resulted in 39.2% more privacy-sensitive options selected ($p < 0.001$), while "military-grade" was associated with 59.3% more options selected ($p < 0.001$). "End-to-end encrypted" again did not differ significantly from "secure."

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 AK Native	2.0 %
	Nat. Hawaiian or Other Pac. Islander	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 Bkgrd.	Yes	22.7%
	No	75.1%

Table 5: Participant demographics. Values may not sum to 100% due to "other" categories and multiple selection.

As in the Privacy Likert results, thoughts of reference and thoughts of persecution were significantly correlated with the number of privacy-sensitive options selected (both $p < 0.001$). A 10 point jump in thoughts of reference scores yielded 19.1% fewer privacy-sensitive options. Conversely, the same increase in thoughts of persecution scores resulted in 21.9% more options. The final model similarly included defaultness, but it was not significant. It did not include priority.

Purposes for which Soteria would be used. Results for # privacy cases (Table 7) were generally in line with # privacy users, but with smaller effect sizes. For instance, compared to "secure," "encrypted" was associated with 19.7% more selections ($p = 0.020$), "military-grade" with 25.8% ($p = .003$). For # privacy users, the numbers were 39.2% and 59.3%, respectively. Similarly, a 10-point increase in thoughts of reference decreased the number of selected privacy options by 9.6% ($p = 0.009$), compared to 19.1% for # privacy users.

The relatively lower effect size of # privacy cases might be due to ceiling effects. Participants on average chose 9.3 of the 12 choices ($\sigma = 3.3$), and 46.7% of participants selected all 12. For # privacy users, only 7.6% selected all options.

Unlike in the other models, age was also significantly associated with # privacy cases selected. The final model estimated that 10 additional years of age corresponded to 4.9% more privacy-relevant selections ($p = 0.025$).

4.3 Perceptions of security (RQ2)

We examined perceptions of security against adversaries using the Security Likert question "Soteria seems secure," as well as scores generated from the adversary-capability questions. The results (Table 4) were consistent across these metrics.

Privacy Likert	OR	$CI_{95\%}$	T-value	p-value
Description (vs. secure)				
end-to-end	1.0	[0.6, 1.6]	-0.119	0.905
encrypted	2.0	[1.1, 3.7]	2.202	0.028*
military-grade	2.4	[1.3, 4.6]	2.735	0.006*
Defaultness (vs. manual)				
on-by-default	1.1	[0.7, 1.6]	0.400	0.689
Demographic covariates				
reference	0.9	[0.9, 1.0]	-2.460	0.014*
persecution	1.1	[1.0, 1.1]	2.737	0.006*
Security Likert				
Description (vs. secure)				
end-to-end	0.9	[0.5, 1.5]	-0.522	0.602
encrypted	1.4	[0.8, 2.7]	1.109	0.267
military-grade	1.5	[0.8, 2.9]	1.225	0.221
Defaultness (vs. manual)				
on-by-default	1.8	[1.2, 2.7]	2.665	0.008*
Demographic covariates				
persecution	1.0	[1.0, 1.0]	1.899	0.058
Paranoid Likert				
Description (vs. secure)				
end-to-end	1.0	[0.6, 1.6]	-0.090	0.928
encrypted	1.2	[0.7, 2.1]	0.585	0.558
military-grade	2.5	[1.4, 4.6]	2.955	0.003*
Defaultness (vs. manual)				
on-by-default	0.9	[0.6, 1.4]	-0.353	0.724
Demographic covariates				
persecution	1.0	[1.0, 1.1]	2.660	0.008*
tech. exp.	0.7	[0.4, 1.0]	-1.810	0.070
age	1.0	[1.0, 1.0]	-2.219	0.026*

Table 6: Final regression models for Privacy, Paranoid, and Security Likerts. Pseudo-R^2's are 0.11, 0.07, and 0.13 respectively. Confidence intervals (CI) were exponentiated to correspond to odds ratios (OR). * indicates statistical significance.

Summary of results. For these metrics, security term did not show any significant effects. However, participants in on-by-default conditions were more likely to agree Soteria was 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 "Soteria seems secure," with median responses of "somewhat agree" for both manual and on-by-default (see Figure 2). In our final regression

# Privacy Cases	IRR	$CI_{95\%}$	Z-value	p-value
Description (vs. secure)				
end-to-end	1.1	[1.0, 1.3]	1.336	0.182
encrypted	1.2	[1.0, 1.4]	2.335	0.020*
military-grade	1.3	[1.1, 1.5]	2.980	0.003*
Defaultness (vs. manual)				
on-by-default	1.0	[0.9, 1.1]	0.401	0.688
Demographic covariates				
reference	1.0	[1.0, 1.0]	-2.622	0.009*
age	1.0	[1.0, 1.0]	-2.249	0.025*
# Privacy Users				
Description (vs. secure)				
end-to-end	1.1	[1.0, 1.3]	1.427	0.154
encrypted	1.4	[1.2, 1.6]	4.052	< 0.001*
military-grade	1.6	[1.4, 1.9]	5.763	< 0.001*
Defaultness (vs. manual)				
on-by-default	1.1	[1.0, 1.2]	1.102	0.270
Demographic covariates				
reference	1.0	[1.0, 1.0]	-3.772	< 0.001*
persecution	1.0	[1.0, 1.0]	4.293	< 0.001*

Table 7: Final regression models for # privacy users and # privacy cases. Pseudo-R^2's are 0.31 and 0.10 respectively. Confidence intervals (CI) were exponentiated to correspond to incidence rate ratios (IRR). * indicates statistical significance.

model (Table **??**), participants in the on-by-default condition were 1.8× more likely than manual participants to increase one point on the Likert scale ($p = 0.008$).

The final model was required to retain the security term variable but did not find them to have a significant effect on security perceptions. The final model also included thoughts of persecution as a non-significant factor. No other variables were retained in model selection.

Adversary scores. We calculated adversary scores for "someone with a strong computer science background" (CS), "people who work at Soteria" (EMP), "the United States government" (GOV), and "your Internet Service Provider" (ISP). We found 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 model for EMP didn't explain much of the variance (adjusted $R^2 < 0.02$ with no statistically significant variables), so we do not discuss it further.

Across all adversaries, capability scores were slightly lower on average for on-by-default than for manual (Figure 2). In the CS model, on-by-default was associated with an estimated 2.8-point drop in adversary score ($p < 0.001$). For ISP, it was 1.5 ($p = 0.025$); for GOV, it was not significant. Details are given in Tables 10, 11, and 12 in Appendix B.

The thoughts of reference score also appeared in all three final adversary models as a small but significant factor (all $p \leq 0.001$). An increase of 10 points in thoughts of reference score was associated with 1.8 additional points of adversary capability rating in each case.

In the CS model only, participants who reported being frequently being asked for computer advice were associated with a 1.6-point drop in adversary capability relative to less-frequent advice givers ($p = 0.045$). Intuitively, 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 beyond those mentioned above appeared in any adversary model.

4.4 Is using Soteria seen as paranoid? (RQ3)

Prior work suggested use of encrypted communication tools can be viewed as paranoid, or only appropriate for illicit or secretive communications [23, 55]. This may manifest as reluctance to use encrypted communications for fear of appearing odd to others. We measured this factor with a Likert-scale question ("People who might use Soteria are paranoid"), as well as by measuring # general users and # general cases.

Summary of results. On average, participants were neutral as to whether the use of Soteria was paranoid. Participants in the military-grade condition were most likely to view Soteria users as paranoid. In smaller effects, participants with higher thoughts of persecution scores and younger participants were more likely to view Soteria users as paranoid.

Paranoid Likert. Overall and for each security term, the median Likert response was "Neither agree nor disagree" (Figure 2). However, in the final regression model (Table **??**), "military-grade" participants were 2.5× as likely as "secure" participants to move up one point on the scale ($p = 0.003$). In contrast, "encrypted" and "end-to-end encrypted" were not significantly different from "secure."

In a smaller effect, a 10-point jump in thoughts of persecution increased the likelihood of a higher rating by 1.4× ($p = 0.008$). A 10-year increase in age corresponded to 0.8× the likelihood of increasing agreement ($p = 0.26$). That is, older participants were less likely than younger participants to view using Soteria 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. Defaultness (required) and tech advice frequency were retained in the final model but were not statistically significant.

Soteria for general users and use cases. According to the final model, participants told that Soteria was "encrypted" selected about 85.6% as many general users ($p = 0.044$, Table 9 in Appendix B). This suggests that use of an "encrypted" tool was seen by participants as more paranoid than a "secure" one.

The regression model for # general cases explained little variance (Pseudo-$R^2 < 0.02$, no significant terms), so we do not discuss it further. As with # privacy cases, this is likely due to ceiling effects: participants selected on average 4.5 of 5 options, with 77.3%

selecting all five. We observed similar, albeit less extreme, ceiling effects for # general users (mean 4.7 out of 6, 47.3% selected all).

4.5 Free-Text Responses

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

4.5.1 Participant definitions of security terms. We asked participants 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.

Participants were least comfortable defining "military-grade" (18.5% confident). This was significantly different from "secure" (34.7% confident, corrected $p < 0.001$). Other security terms did not differ significantly from "secure." Full confidence ratings are shown in Figure 3 in Appendix B.

We highlight below several key themes from participants' free responses, which broadly align with prior work in mental models of encryption [1, 3–5, 14, 24, 32, 55].

Technical details. When defining the security terms, about a third of our participants (30.3%) mentioned specific technical details. Participants described transformation from plain-text to ciphertext (12.9%) in a variety of ways, including "scrambled," "special coding language," and "random letters, symbols, and numbers." Mentions of transformation to ciphertext were most common when defining "encrypted" (37.8%, compared to less than 10% for any other security term). These results align well with prior work [55].

Other participants focused on the need for a secret (8.4%), aligning with prior finding that mental models approximate symmetric-key encryption [55]. Many of these referred to the secret as a "key," but some implied a secret algorithm instead (e.g., "know how to decrypt it"). A small fraction (0.8%) mentioned or described asymmetric encryption. The need for a secret was mentioned most frequently for "encrypted" (24.3%), followed by "end-to-end encrypted" (7.1%), and "military-grade" (5.7%). It was never mentioned for "secure."

Among participants told Soteria is "secure," 30.1% implied or hoped that "secure" involved encryption (e.g., "the communications are encrypted in some way"). Less than 1% of participants mentioned each of specific encryption algorithms (e.g., AES256), account protection (e.g., "Only administered users can access it"), or (non-)protection of metadata.

Protection from whom. Again consistent with prior work [1, 3], 34.5% of participants mentioned general or specific adversaries when describing Soteria's security properties. Some (14.3%) specifically noted that only the sender and receiver could see messages (e.g., "any third party . . . 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, 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 prior work [14, 24]. Interestingly, protection against the government was exclusively mentioned by "military-grade" participants.

Military-grade is for the military. Almost a third of "military-grade" 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). Unsurprisingly, this definition was not used for any other security term.

4.5.2 Benefits and Drawbacks. We also asked participants to suggest benefits and drawbacks of Soteria. 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 (93.0%) noted non-security benefits, such as being free, multi-platform, or user-friendly. Just over half (51.5%) mentioned non-security drawbacks, including the lack of a large user base to communicate with, already having other similar apps, and concerns about service quality. Fractured user bases and low quality of service have previously been identified as critical factors inhibiting adoption of secure messaging tools [4, 13]. Almost one quarter (22.7%) listed no drawbacks.

Security is valuable, if you can trust it. Many participants (43.7%) mentioned the specific security term associated with their condition as a benefit (e.g., "I think the military grade encryption is the primary benefit of using this program"). In line with with prior work [3, 24], a smaller group (6.2%) explicitly doubted the security of Soteria. As one put it, ". . . there is no way to know how reliable their encryption . . . will be." Notably, secure was least frequently mentioned as a benefit (34.7% of secure 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 defaultness as a benefit or drawback. Among those assigned on-by-default, 11.8% saw it as a benefit, and none listed it as a drawback. Among manual participants, 11.1% mentioned secure messages as a benefit; however, none explicitly mentioned opt-in as valuable. A small number (1.7%) explicitly saw it as a drawback: "Not having the encryption being the default, but rather opt-in, can be a drawback for privacy."

Other indicators of trustworthiness. Participants also described security- and privacy-related issues not directly associated with their assigned conditions. Positive security connotations included that Soteria was "private" (21.3%), could not be compromised (5.0%), was "safe" (3.1%), was independent from large companies (0.5%), and was "anonymous" (0.2%). Notably, these benefits appeared more frequently for "military-grade" (48.6%) than for the other security terms (23–27% each). We note that 78.6% of "military-grade" and 71.6% of "encrypted" participants positively mentioned at least one security feature or indicator of trustworthiness, compared to 61.0% for "end-to-end encrypted" and 48.6% for "secure."

This aligns with our results in section 4.2: "encrypted" and "military-grade" seemed to inspire more trustworthiness than "secure."

Doubts related to security and privacy included distrust in the (unknown) company (7.8%), concern about personal data collection (5.3%), skepticism about Soteria's privacy and security claims (5.4%), and worry about being vulnerable to hacking (3.4%). Participants in "military-grade," "end-to-end encrypted," and "secure" were about equally likely to mention such doubts (40.0%, 40.4%, and 40.3% respectively), slightly more than "encrypted" (32.4%).

Soteria is for criminals. Similar to prior work [55], and reflecting common discussion in the news, 7.6% of participants mentioned that Soteria would be useful for illegal activities. One respondent noted, "It's great for the people who actively engage in shady or illegal activities." Another wrote: "I can readily picture a major objection from police and governmental departments who would be unable to monitor or tap conversations." Illegal activities were mentioned as one option in the "who" and "what" questions, which might have primed participants toward this answer. However, participants did not similarly repeat other listed activities, such as political activism or sending sexts.

5 DISCUSSION

We compared different descriptions of a secure-messaging tool to understand how the terminology used to describe its security mechanism, whether or not security is on-by-default, and the prioritization of privacy among app features affect users' perceptions. Our results shed light on how users form opinions about messaging apps and allow us to revisit a finding from 2006 that encrypted tool use is often seen as paranoid. Further, we explore how people's levels of psychological paranoia contribute to perceptions of secure messaging.

Military-grade is poorly understood but influential. Participants were least confident in defining "military-grade" compared to the other security terms; a plurality of participants interpreted it as used by, or up to the standards of, the military or government. (We note that this term has no precise meaning, so participants' confusion is understandable.) This intuition seems to provide a strong association with privacy, perhaps to an unsettling degree: "military-grade" was correlated with more utility for privacy, but also a stronger perception that tool users were paranoid. Somewhat surprisingly; however, "military-grade" had no effect on perception of strength against adversaries. Overall, this phrasing may convey a tool is suitable for privacy uses but not necessarily for everyday use, and it is not necessarily more secure than other tools.

"Encrypted" was also correlated with greater privacy utility than the "secure" baseline, but to a lesser extent than "military-grade." Somewhat to our surprise, "end-to-end encrypted" did not differ significantly from "secure" on any metric. This may relate to the relatively high frequency of misunderstanding of this concept, as pointed out in prior work [3, 14, 32].

Defaultness only matters for security against adversaries; priority does not matter at all. We initially hypothesized that defaultness might affect whether use of Soteria was seen as paranoid, or as useful for privacy-sensitive but not general-purpose tasks. Based on Gaw et al.'s results [23], it seemed plausible that participants might view on-by-default security as overkill. However, we found that defaultness only correlated with perceived security against adversaries, with manual security seen (appropriately) as less secure than on-by-default. In free responses, defaultness was not mentioned in conjunction with illegal activity or other indicators of illicitness.

Varying whether security received high or low priority in the app description had no effect for any metric; this nuance may have been too subtle to register in this experiment (see e.g., Redmiles et al. [38]).

Personal paranoia is a factor. We find that higher levels of persecutory thoughts are associated with stronger beliefs that Soteria is useful for privacy-relevant tasks and that people who use Soteria are paranoid. This aligns well with prior work suggesting that persecutory thoughts are associated with fear of surveillance and use of coping "safety" behaviors [19]: 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 Soteria with privacy tasks and less likely to believe it provided strong protection from adversaries. We hypothesize that together these two correlations indicate lack of trust in Soteria to properly handle privacy-sensitive communications. Because the two paranoia metrics measure different underlying factors — differing specifically 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. Although our work is not directly comparable to Gaw et al.'s seminal paper, we make observations with regard to similar themes in a world with a drastically different encrypted communication landscape. Centralized key management and near-transparency to users have brought end-to-end-encrypted communication to many of the most popular messaging applications. 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 Soteria could be used for general-purpose communication tasks. Security features were seen almost entirely as a benefit, and a key drawback was concern about 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. On the contrary, some in the manual condition requested that security be turned on by default. 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 to be perceived as overly fearful or unreasonable — is not entirely gone. Use of the military-grade security term was associated with viewing Soteria users as somewhat paranoid. One free-response participant commented that Soteria "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) [13, 15], perceptions of which users and use cases a tool is (not) appropriate 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, and thereby influence the likelihood of adopting it [36]. A term like "military-grade," in addition to being imprecise, may be overdoing it, making a tool seem fraught or even illicit. "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" — a relatively well-understood term — came closest to this balance.

We also find that turning security on by default has a small positive effect on 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, ads based on the content of conversations, or group chat [2, 34]. We hope that demonstrating a positive association with on-by-default will provide an incentive for companies to move in this direction.

6 ACKNOWLEDGEMENTS

We thank our participants. This material is based upon work supported by the United States Air Force and DARPA under Contract No FA8750-16-C-0022. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the United States Air Force and DARPA.

REFERENCES

[1] Ruba Abu-Salma. 2020. *Designing User-Centered Privacy-Enhancing Technologies.* Ph.D. Dissertation. UCL (University College London).

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

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

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

[5] Ruba Abu-Salma, M Angela Sasse, Joseph Bonneau, and Matthew Smith. 2015. POSTER: Secure Chat for the Masses? User-centered Security to the Rescue. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security.* 1623–1625.

[6] Omer Akgul, Wei Bai, Shruti Das, and Michelle L. Mazurek. 2021. Evaluating In-Workflow Messages for Improving Mental Models of End-to-End Encryption. In *30th USENIX Security Symposium (USENIX Security 21).* USENIX Association, 447–464. https://www.usenix.org/conference/usenixsecurity21/presentation/akgul

[7] Wei Bai. 2019. *User Perceptions of and Attitudes toward Encrypted Communication.* Ph.D. Dissertation.

[8] Wei Bai, Moses Namara, Yichen Qian, Patrick Gage Kelley, Michelle L. Mazurek, and Doowon Kim. 2016. 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).* USENIX Association, Denver, CO, 113–130. https://www.usenix.org/conference/soups2016/technical-sessions/presentation/bai

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

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

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

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

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

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

[15] Steve Dodier-Lazaro, Ruba Abu-Salma, Ingolf Becker, and M Angela Sasse. 2017. From Paternalistic to User-Centred Security: Putting Users First with Value-Sensitive Design. In *CHI 2017 Workshop on Values in Computing.* Values In Computing.

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

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

[18] Electronic Frontier Foundation. 2014. Secure MessagingScorecard. (2014). https://www.eff.org/secure-messaging-scorecard Accessed on: 09.07.2016.

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

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

[21] Daniel Freeman, Bao S Loe, David Kingdon, Helen Startup, Andrew Molodynski, Laina Rosebrock, Poppy Brown, Bryony Sheaves, Felicity Waite, and Jessica C Bird. 2019. The revised Green et al., Paranoid Thoughts Scale (R-GPTS): psychometric properties, severity ranges, and clinical cut-offs. *Psychological medicine* (2019), 1–10.

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

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

[24] Nina Gerber, Verena Zimmermann, Birgit Henhapl, Sinem Emeröz, and Melanie Volkamer. 2018. Finally Johnny Can Encrypt: But Does This Make Him Feel More Secure?. In *Proceedings of the 13th International Conference on Availability, Reliability and Security.* 1–10.

[25] CEL Green, D Freeman, E Kuipers, P Bebbington, D Fowler, G Dunn, and PA Garety. 2008. Measuring ideas of persecution and social reference: the Green et al. Paranoid Thought Scales (GPTS). *Psychological medicine* 38, 1 (2008), 101–111.

[26] Eric Griffith. 2019. NordVPN dominates VPN market share, and that will likely continue. *PC Magazine* (2019). https://www.pcmag.com/news/nordvpn-dominates-vpn-market-share-and-that-will-likely-continue

[27] Timothy M Hagle and Glenn E Mitchell. 1992. Goodness-of-fit measures for probit and logit. *American Journal of Political Science* (1992), 762–784.

[28] Amir Herzberg and Hemi Leibowitz. 2016. 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.* 17–28.

[29] Patrick Gage Kelley, Lorrie Faith Cranor, and Norman Sadeh. 2013. Privacy as part of the app decision-making process. In *Proceedings of the SIGCHI conference on human factors in computing systems.* 3393–3402.

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

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

[32] Juan Ramón Ponce Mauriés, Kat Krol, Simon Parkin, Ruba Abu-Salma, and M Angela Sasse. 2017. Dead on Arrival: Recovering from Fatal Flaws in Email Encryption Tools. In *The {LASER} Workshop: Learning from Authoritative Security Experiment Results ({LASER} 2017).* 49–57.

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

[34] Sean Oesch, Ruba Abu-Salma, Oumar Diallo, Juliane Krämer, James Simmons, Justin Wu, and Scott Ruoti. 2020. Understanding User Perceptions of Security

and Privacy for Group Chat: A Survey of Users in the US and UK. In *Annual Computer Security Applications Conference*. 234–248.

[35] K.G. Orphanides. 2020. Why everyone should be using Signal instead of WhatsApp. *Wired* https://www.wired.co.uk/article/signal-vs-whatsapp. (April 16, 2020).

[36] Robert W Proctor and Jing Chen. 2015. The role of human factors/ergonomics in the science of security: decision making and action selection in cyberspace. *Human factors* 57, 5 (2015), 721–727.

[37] Elissa M Redmiles, Sean Kross, and Michelle L Mazurek. 2019. 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)*. IEEE, 1326–1343.

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

[39] Scott Ruoti, Jeff Andersen, Scott Heidbrink, Mark O'Neill, Elham Vaziripour, Justin Wu, Daniel Zappala, and Kent Seamons. 2016. "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*.

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

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

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

[43] Signal. 2020. Signal Private Messenger - Apps on Google Play. https://play.google.com/store/apps/details?id=org.thoughtcrime.securesms. (2020). Accessed on: 02.29.2020.

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

[45] Christian Stransky, Dominik Wermke, Johanna Schrader, Nicolas Huaman, Yasemin Acar, Anna Lena Fehlhaber, Miranda Wei, Blase Ur, and Sascha Fahl. 2021. On the Limited Impact of Visualizing Encryption: Perceptions of E2E Messaging Security. In *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*. USENIX Association, 437–454. https://www.usenix.org/conference/soups2021/presentation/stransky

[46] Joshua Tan, Lujo Bauer, Joe Bonneau, Lorrie Cranor, Jeremy Thomas, and Blase Ur. 2017. Can Unicorns Help Users Compare Crypto Key Fingerprints?. In *ACM Conference on Human Factors and Computing Systems*.

[47] Telegram. 2020. Telegram - Apps on Google Play. https://play.google.com/store/apps/details?id=org.telegram.messenger. (2020). Accessed on: 02.29.2020.

[48] Elham Vaziripour, Devon Howard, Jake Tyler, Mark O'Neill, Justin Wu, Kent Seamons, and Daniel Zappala. 2019. I Don't Even Have to Bother Them! Using Social Media to Automate the Authentication Ceremony in Secure Messaging. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300323

[49] Elham Vaziripour, Justin Wu, Mark O'Neill, Jordan Whitehead, Scott Heidbrink, Kent Seamons, and Daniel Zappala. 2017. 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)*. 29–47.

[50] Viber. 2020. Viber Messenger - Messages, Group Chats & Calls - Apps on Google Play. https://play.google.com/store/apps/details?id=com.viber.voip. (2020). Accessed on: 02.29.2020.

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

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

[53] Gloria HY Wong, Christy LM Hui, Jennifer YM Tang, Cindy PY Chiu, May ML Lam, Sherry KW Chan, WC Chang, and Eric YH Chen. 2012. 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 (2012), 158–163.

[54] Justin Wu, Cyrus Gattrell, Devon Howard, Jake Tyler, Elham Vaziripour, Daniel Zappala, and Kent Seamons. 2019. "Something isn't secure, but I'm not sure how that translates into a problem": Promoting autonomy by designing for understanding in Signal. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. USENIX Association, Santa Clara, CA. https://www.usenix.org/conference/soups2019/presentation/wu

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

A QUESTIONNAIRE

Consent and validation

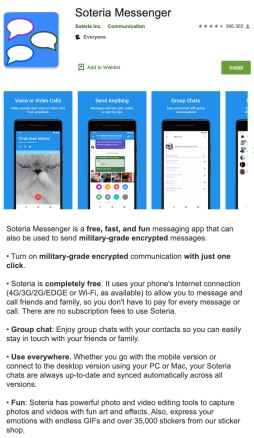
(1) *Consent form is shown, and consent is given*
(2) In what country do you currently reside?
  - United Kingdom
  - United States
  - Ireland
  - Germany
  - France
  - Spain
  - Other [Free text]

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

Part 1: Who would use Soteria 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 Soteria.

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



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

(2) Can you use Soteria 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 Soteria?
  - 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 Soteria, you need to turn it on.
  - In Soteria, the [security term] communication is turned on by default.
  - None of the above

  Page Break

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

(5) Who do you think would be interested in using Soteria? (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-sensitive options: {3, 4, 7, 9, 10, 11, 12, 13, 14, 15}

  Page Break

(6) Who would you talk to on Soteria 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 Soteria 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-sensitive options: {6, 7, 8, 9, 10, 11, 12}

Page Break

(8) If you decided to use Soteria, which of the following would you do on Soteria? (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 Soteria? [Free text]

(10) What do you see as the major drawbacks of using Soteria? [Free text]

(11) To what extent do you agree with the following statements: *options: {Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree}*
- People who might use Soteria are paranoid.
- Based on the screenshot given, Soteria looks professionally designed.
- Soteria seems secure.
- Soteria seems fun to use.
- People who care about their privacy would use Soteria.

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 Soteria communications or your Soteria 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 Soteria, 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:
options: {Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree}
- read the content of your Soteria messages
- listen to your Soteria phone calls
- modify your Soteria communications
- impersonate you on Soteria
- determine who you are communicating with on Soteria
- determine how long you are communicating with someone on Soteria

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 assigned 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? *options: {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 inaccurate 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 [21])

Page Break

(5) R-GPTS Part B (as it appears in [21])

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? *options: {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]

	β	95% CI		T-value	p-value
Description (vs. secure)					
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
Defaultness (vs. manual)					
on-by-default	−1.522	[−2.854	−0.190]	−2.247	0.025*
Demographic covariates					
reference	0.175	[0.081	0.269]	3.663	< 0.001*
technical expertise	−1.173	[−2.643	0.296]	−1.570	0.117

Table 11: Final model for adversary-capability score of "Your Internet Service Provider (ISP, e.g. Verizon, Article AT&T)" (ISP). Adjusted $R^2 = 0.047$.

	β	95% CI		T-value	p-value
Description (vs. secure)					
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
Defaultness (vs. manual)					
on-by-default	−1.033	[−2.470	0.404]	−1.414	0.158
Demographic covariates					
reference	0.175	[0.074	0.277]	3.396	< 0.001*
technical expertise	−1.430	[−3.015	0.156]	−1.773	0.077

Table 12: Final model for adversary-capability score of "The United States government" (GOV). Adjusted $R^2 = 0.034$.

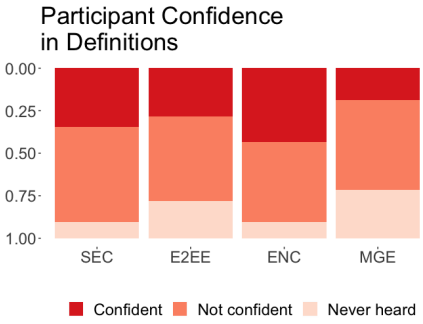


Figure 3: Participant confidence in explaining the security term assigned to them. Darker colors indicate more confidence.

- I do not have digital security requirements (or precautions)
- Prefer not to say

End of survey

B ADDITIONAL FIGURES AND TABLES

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 8: The number of participants who saw each description.

	IRR	95% CI		Z-value	p-value
Description (vs. secure)					
end-to-end	0.943	[0.830	1.073]	−0.895	0.371
encrypted	0.856	[0.736	0.995]	−2.017	0.044*
military-grade	0.921	[0.792	1.071]	−1.069	0.285
Defaultness (vs. manual)					
on-by-default	1.013	[0.920	1.116]	0.269	0.788
Demographic covariates					
reference	0.994	[0.987	1.001]	−1.729	0.084

Table 9: Final model for who would use (general purpose) Soteria (# general users). Pseudo-$R^2 = 0.033$. CI obtained by exponentiating the CI of the regression coefficients.

	β	95% CI		T-value	p-value
Description (vs. secure)					
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
Defaultness (vs. manual)					
on-by-default	−2.765	[−4.164	−1.366]	−3.887	< 0.001*
Demographic covariates					
reference	0.176	[0.078	0.275]	3.513	< 0.001*
technical expertise	−1.577	[−3.120	−0.033]	−2.009	0.045*

Table 10: Final model for adversary-capability score of "Someone with a strong computer science background" (CS). Adjusted $R^2 = 0.073$.