How Language Formality in Security and Privacy Interfaces Impacts Intended Compliance

JACKSON STOKES* and TAL AUGUST*, University of Washington, USA
ROBERT MARVER, University of Washington, USA
ALEXEI CZESKIS, ID.me, USA
FRANZISKA ROESNER, University of Washington, USA
TADAYOSHI KOHNO, University of Washington, USA
KATHARINA REINECKE, University of Washington, USA

Strong end-user security practices benefit both the user and hosting platform, but it is not well understood how companies communicate with their users to encourage these practices. This paper explores whether web companies and their platforms use different levels of language formality in these communications and tests the hypothesis that higher language formality leads to users' increased intention to comply. We contribute a dataset and systematic analysis of 1,817 English language strings in web security and privacy interfaces across 13 web platforms, showing strong variations in language. An online study with 512 participants further demonstrated that people perceive differences in the language formality across platforms and that a higher language formality is associated with higher self-reported intention to comply. Our findings suggest that formality can be an important factor in designing effective security and privacy prompts. We discuss implications of these results, including how to balance formality with platform language style. In addition to being the first piece of work to analyze language formality in user security, these findings provide valuable insights into how platforms can best communicate with users about account security.

Additional Key Words and Phrases: Privacy, Security, Language, Formality

ACM Reference Format:
Jackson Stokes, Tal August, Robert Marver, Alexei Czeskis, Franziska Roesner, Tadayoshi Kohno, and Katharina Reinecke. 2023. How Language Formality in Security and Privacy Interfaces Impacts Intended Compliance. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3544548.3581275

1 INTRODUCTION

As our world becomes more digitally integrated, individuals store an increasing amount of personal information in online accounts. From identity to financial information, these accounts are clear targets for hackers. There are many different methods of exploit for hackers to compromise digital accounts, from guessing simple passwords to using social engineering. To combat this, there are numerous security measures that individuals and platforms can take to protect their accounts. However, these measures often have imperfect adoption among users [53]. While there are many reasons for individuals not to use certain security practices, some of these are due to a lack of knowledge or understanding on the part of the users. For example, many users are unaware of what security practices an individual should perform, or

*Both authors contributed equally to this research.

Manuscript submitted to ACM

why they improve security [17, 40, 46]. An individual might also not adopt security practices because they do not have clear directions of what they can do to secure their account or because of the additional time it takes to do so [48, 51].

It is in the interest of platforms to promote strong security practices among their users to prevent data leaks of private information and to promote trust. Platforms have the opportunity to convince their users to make secure choices on their accounts through security recommendations (e.g., by suggesting a password with a minimum number of characters). These recommendations come in the form of explicit requirements, such as using strong passwords, or they might bring up opt-in security measures, such as suggesting that a user signs up for two-factor authentication. The security recommendations that account providers give must clearly convey the proper steps that an individual should take, and some argue that providers must also give an explanation of why users should take these steps [25].

How this kind of text is written may play a role in achieving these goals. The style of language can have a significant impact on its ability to persuade or inform a reader [10, 23, 52]. For example, the formality of prompts can influence people's attention to those prompts [3]. Formality is also associated with trust and authority [10, 36] and its use varies widely depending on context (e.g., talking to a friend or boss) and population groups (e.g., older and younger people) [23]. This suggests that the wording of security prompts, specifically their formality, may have an impact on people's reaction to these prompts. In this study, we examine how different levels of formality in security guidelines can have an impact on user's reported intent to comply. We have two main research questions:

R.Q.1 Do people perceive the language formality of security and privacy prompts to be different across platforms?
R.Q.2 Does the language formality of security and privacy prompts impact whether people intend to comply?

To answer these research questions, we began by collecting a dataset of 1,817 English prompts from the web (including both desktop and mobile) security and privacy interfaces of 13 large U.S. technology platforms. We coded these prompts for aspects of language we hypothesized were associated with perceived formality and intention to comply (e.g., professional tone). In an online experiment (n=512), we collected participant ratings of formality and their likelihood to comply with a sample of 135 prompts.

We found that perceived formality was associated with our codes for professional language and significantly varied across platforms, with Amazon using the most formal language and Instagram using the least. We also found that formality had a significant effect on participants' reported intended compliance. The most formal prompts were associated with the highest intention to comply, even when controlling for different types of prompts (e.g., two-factor authentication vs. using a unique password). The results suggest that formality can be an important factor in designing effective security and privacy prompts.

Our paper makes the following contributions:

(1) A dataset of 1,817 English web security prompts from 13 large U.S. technology platforms, manually coded with language tones related to formality (e.g., authoritativeness) available at https://www.labinthewild.org/data/.
(2) Empirical findings based on 512 participant ratings of security and privacy strings showing that (i) perceived formality differs across platform and (ii) formality is positively associated with intent to comply.

2 RELATED WORK

This research fits into broader research about optimal forms of communication on the web, the impact of text formality on its audience, and personalized interfaces.

2.1 Effects of Language on User Behavior

Language is particularly powerful in user interfaces because it allows for direct requests of user action and can persuade users to take action that they might not have done on their own [12]. The persuasiveness of a piece of text is complex: formality, dialect, and jargon have all been shown to affect user behavior, such as their likelihood to understand and follow requests [2, 3, 6, 49]. In particular, formality has been shown to impact attention to instructions in online experiments [3]. Research also indicates that text that contains a request tends to be more formal [36].

Nudging can affect the way users interact with these interfaces. Effective nudging can alter behavior in predictable ways by leveraging cognitive biases [7]. Nudging has also been shown to be effective in encouraging good security behavior [1, 4]. This work can contribute to our understanding of how formality plays a role in effective security nudging.

2.2 Interface Design in Usable Security and Privacy

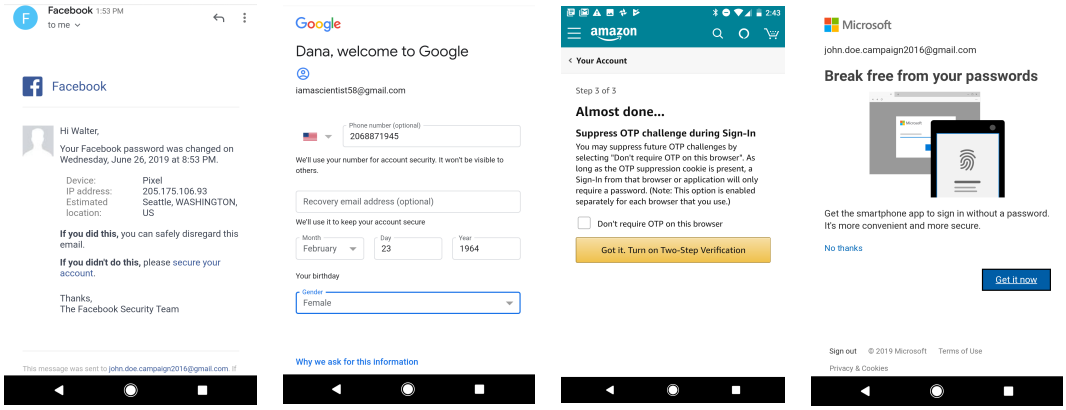
Our work here contributes to our understanding of challenges related to communicating with users who should ideally comply with security and privacy prompts but may not do so if those prompts are not well-designed. For example, a user might not understand what security practices they should employ or how to employ them, or they might hesitate on following security practices because they seem too burdensome [31, 44]. Prior work has studied how the designs of different security-related interfaces can influence compliance, e.g., in the context of browser warnings and indicators [15, 16], or the use of social proof to influence adoption of two-factor authentication [11].

There are many aspects of an interface's design that might impact adoption of security and privacy related behaviors or options. In this work, we focus specifically on the potential influence of the perceived formality of *text* in web security and privacy interfaces on intended compliance.

Several challenges hinder adoption of good security practices. The field of usable security and privacy concerns itself with these challenges, which may include fundamental tensions between different requirements or threat models, or the difficulty of designing usable tools or communicating with users. We stress that not all security practices are necessarily appropriate for all users or all contexts; for example, a user might reasonably choose not to use two-factor authentication because they share a low-value account with another person, but the account system has not been designed to allow for usable sharing (which prior work has shown is common in trusted contexts [32]).

2.3 Interface Personalization

Different groups of individuals may have different needs for security interfaces to satisfy. In fact, personalizing prompts to specific audiences has been shown to increase adoption [18, 43]. For example, tech savvy users might not need an explanation on two-factor authentication or why it is important, while others (such as older adults [38]) might appreciate an explanation on what security measures exists and how they can take them. Personalizing security prompts have been shown to increase their efficacy [20], suggesting that personalizing security prompt text is a potential method to increase compliance. However, it is unclear what features of security prompt text can impact compliance. Identifying these features are a first step to exploring how their variation can better suit different audiences. Our paper seeks to contribute to this research by identifying a new element of interface design for security prompts, language formality, and show its influence on users' intended compliance.



(a) Facebook's Password Change

(b) Google's New Account Creation

(c) Amazon's 2-Factor-Authentication

(d) Microsoft's Password Change

Fig. 1. Example screenshots of mobile security and privacy interfaces in our dataset.

3 DATASET OF SECURITY AND PRIVACY PROMPTS ACROSS 13 TECHNOLOGY PLATFORMS

Our first step was to create a dataset with prompts from the English versions of the web security and privacy interfaces of large technology platforms. While our primary goal was to analyze how platforms may vary in their approaches for communicating security and privacy requests with their users, we also make this dataset available to other researchers at https://www.labinthewild.org/data/.

3.1 Retrieving Language Strings

We selected 13 large technology platforms to represent platforms with a large number of users and across diverse application areas (see Table 1). We focused on security and privacy interfaces for the most common use cases seen by users: (1) new account setup, (2) password changes, (3) password resets, (4) two-factor authentication setup, (5) security and login settings, (6) notifications, and (7) account keys.

To retrieve the strings, we manually created accounts for each service using their app (downloaded from the Google Play store) and using their online platform through the Google Chrome Browser. This allowed us to retrieve both mobile and desktop language strings. For each platform, we created a new account (e.g., a Google account) using a unique email address and phone number. Following norms in computer security research, we created accounts under pseudonyms. Personal details were used from characters from the popular TV show, "X-Files" (e.g. name, birth date).

We then took screenshots of all privacy and security pages. Example screenshots are shown in Figure 1. From each page, we recorded all strings that were used to communicate with us. Strings were retrieved from both the mobile and the desktop interface. Each string was annotated with the use case (e.g., creating a password), whether the text was a header or in a paragraph, the text position on the page, text type (e.g., button, checkbox description), and device (mobile vs desktop). The total dataset includes 1,817 strings from 13 platforms. The number of strings per platform varies between 31 (for Reddit) and 363 (for Google).

Table 1. Number of strings collected from the web security and privacy interfaces of 13 platforms. Random samples of strings from the first eight platforms were used in the online study.

Platform	# of strings
Google	363
Microsoft	253
Facebook	222
Apple	184
Amazon	151
Yahoo	146
Twitter	142
Instagram	118
Netflix	64
Hulu	48
WhatsApp	48
Tinder	47
Reddit	31
Total	1817

3.2 String Coding

To uncover how different platforms communicate security and privacy recommendations and their relationship to formality, we developed a qualitative codebook for the complete dataset. We specified an initial set of codes related to our research questions based on randomly drawn strings from different platforms. Codes were also informed by prior work on nudges [1, 7] and formality [5, 28, 35, 36].

Four authors met five times to discuss the codes and iteratively refine them. Independent coding was then performed by a pair of authors on a subset of 140 strings. Inter-rater agreement had a Cohen's Kappa of 0.515, which was indicative of moderate agreement with some difference suggesting the difficulty and subjective nature of the codes [33]. After resolving disagreements, the codebook was updated and the authors each coded half of the dataset. Some strings did not contain enough text to be adequately coded (e.g., the text "Continue" on a button). These strings remain in the dataset, but do not have any associated codes.

The resulting codebook contains three high-level codes with subcategories (see Table 2). The first category, Tone, analyzes the characteristics of a prompt related to formality. The next code was the type of request a prompt makes (Request). Finally, we included a code for the presence of technical language in the prompt. We define technical language as being unfamiliar to an individual with little to no computer experience (e.g., "two-factor authentication").

None of these codes explicitly define a string as "formal" or "informal." Instead our goal was to see how platforms differ in their language across tones that are associated with formality (e.g., professional or casual tone). While automated measures exist for some of these codes (e.g., polite or professional language [10]), we chose to manually code strings because prior systems were trained on different text domains (e.g., blogs) and rarely generalize well to new domains. We hope that our new dataset can support future systems for identifying these tones in security and privacy interfaces. In Section 4.1 we introduce an explicit measure of formality based on crowdsourced ratings.

Table 2. Codebook for the language strings in web security and privacy interfaces, reflecting differences in the language tone, requests, and technical language used in security and privacy interfaces.

Code		Description	Example
Tone	Casual	Text with colloquial language, similar to how friends may speak together [28]. Casual language is a major element of informal language, but informal language can also include smaller textual changes, such as irregular, or alternative punctuation (e.g., !!) and capitalization [35].	"My best pic is:" vs. "Upload your profile picture"
	Authoritative	Text with a demanding tone, indicating a power imbalance [5]. More authoritative language is often associated with lower formality [36], which is thought to be due to a higher power difference between a speaker and listener [5].	"Finish signing up" vs. "Done signing up?"
	Professional	Polite text used in professional contexts. Polite and professional text is often associated with higher formality [36].	"It may take a few minutes to arrive" vs. "It'll be here in a jiffie"
	Dialog	Neutral text present in dialogue boxes. This code was added based on observations in our dataset that dialogue boxes had a unique, neutral tone to them.	"Trying to sign in from another computer?"
Request	Command	Requests that give the impression that the user must complete an action.	"Re-type new password"
	UserRequest	Requests that ask (rather than command) a user to do something.	"Read our privacy policy"
	Optional	Requests that explicitly present options for the user to choose from.	"Select your gender or decide not to say."
Technical		Text that includes technical terms.	"A security key is a physical device (like a USB security key)"

3.3 Codes Across Platforms

Our dataset is the first to allow for an exploration into the variations in how platforms communicate with their users. Below we illustrate some examples focused on how the top eight platforms with the most prompts varied across three attributes of tone that are related to language formality: Casual, Authoritative, and Professional.

We found that platforms strongly differed in the number of language strings that were coded as casual tone, from none for Apple to over 18% for Google, Instagram and Twitter (see first column in Table 3). For example, when opening

Table 3. Percentage of strings with various tones across the top eight platforms with the most prompts. Each percentage is out of total number of strings coded with a tone of Casual, Authoritative, or Professional, and therefore sum to 100% for each platform.

% Casual	% Authoritative	% Professional	% Dialog
Google (19.6%)	Apple (20.0%)	Apple (70.0%)	Yahoo (18.1%)
Twitter (18.9%)	Twitter (17.6%)	Google (65.5%)	Microsoft (16.9%)
Instagram (18.2%)	Microsoft (14.6%)	Instagram (63.6%)	Twitter (16.2%)
Yahoo (15.6%)	Facebook (13.5%)	Amazon (63.1%)	Facebook (15.2%))
Facebook (13.4%)	Amazon (11.9%)	Yahoo (59.7%)	Amazon (14.3%)
Microsoft (11.5%)	Instagram (10.9%)	Facebook (58.0%)	Apple (10.0%)
Amazon (10.7%)	Google (6.9%)	Twitter (47.3%)	Google (8.1%)
Apple (0.0%)	Yahoo (6.5%)	Microsoft (56.9%)	Instagram (7.3%)

a new account with Twitter, users are prompted with "Don't think too hard, just have fun with it." *[Twitter, new account creation]*, while Apple users are prompted with "This will be your new Apple ID." *[Apple, new account creation]*. The finding indicates that Twitter more commonly uses casual language when communicating security suggestions.

While both authoritative and casual language were associated with lower formality in prior work, we see that the same platform does not often use both. Twitter and Apple used more authoritative strings (>17%) than platforms like Google and Yahoo (<7%). The majority of platforms that had high rates of casual prompts (e.g., Google, Instagram and Yahoo) also had low rates of authoritative prompts. Similarly, while Apple had the lowest rate of casual prompts (0%), it had the highest rate of authoritative prompts (20.0%). Twitter is the one outlier in this trend, having a high rate of casual prompts (18.9%) as well as a high rate of authoritative prompts (17.6%), though this is mostly due to short prompts demanding some action (e.g., "Change password" *[Twitter, Password reset]*). This suggests that platforms may adopt different tones within informal or formal communication that represent their particular style of communication.

Platforms using authoritative strings frequently directed users to complete an action (e.g., "Finish signing up" *[Facebook, new account creation]*) or warned them about an action (e.g., "Apple will not be able to reset your password on your behalf." *[Apple, Security phrase]*, "Before you enable two-step verification, you must agree to the following conditions:" *[Apple, 2SV sign-up]*). In contrast, platforms like Google or Yahoo instead opted for highlighting the benefit of actions for users (e.g., "If you ever have trouble signing in, your up-to-date recovery email and mobile number will help you get to your account." *[Yahoo, Password Change]*) or allowing users to request more information (e.g., "After leaving the app, how long until a device PIN / pattern / fingerprint is required for re-opening?" *[Yahoo, Security phrase]*).

We also observed a variation in the use of professional language. Google had a higher rate of professional prompts than Twitter (65.5% and 47.3%, respectively). Looking at Google's prompts compared to Twitter's, Google's was often professional when requesting a user action (e.g., creating a stronger password) while Twitter often used a more authoritative tone. For example, Google's way of telling a user to provide a stronger password was, "Please choose a stronger password. Try a mix of letters, numbers, and symbols." *[Google, password change]*. In comparison, Twitter provided more explicit and authoritative guidance: "Your password must be at least 6 characters." *[Twitter, password change]*. This is also reflected in Twitter's higher rate of authoritative codes (17.6%) compared to Google's (6.9%).

Our results show that the use of language in security and privacy interfaces strongly varies across platforms. While some platforms attempt to nudge users with authoritative and professional tones (e.g., Apple), others use noticeably more casual tones (e.g., Google) or use casual and authoritative communication interchangeably (e.g., Twitter).

These differences highlight some of the ways that platform language differs across tones related to formality. In our online study we further explore how these codes are associated with users' perception of formality across platforms.

4 ONLINE STUDY

We designed an online experiment to answer our two research questions (Section 1). We have the following hypotheses based on our online study and codebook:

H.1: People perceive the language formality of prompts as different across platforms.

H.1.1: Perception of formality is positively associated with professional language and negatively associated with casual and authoritative language.

H.2: Intention to comply with a security and privacy prompt is positively associated with the prompt's formality.

H.2.1: Intention to comply is positively associated with professional language and negatively associated with casual and authoritative language.

4.1 Methods

We designed the experiment to run on the online study platform LabintheWild to recruit diverse volunteer participants who are intrinsically motivated to provide reliable subjective responses [41]. The study was advertised with the slogan "Amazon, Apple, Facebook, Google: Can you tell a difference?" with an expected duration of 10 minutes. The study was labeled as "only available in English." We explicitly did not mention security or privacy in the advertisement of the study to avoid a selection bias.

Materials. Starting from our original dataset of 1,817 strings from 13 platforms, we selected strings from the 8 platforms that had at least 100 strings (see Table 1). Two researchers then manually selected 135 strings representing a balanced variety of length, formality, and types of prompts. While we included all 8 platforms for this analysis, 5 platforms had the majority of strings (126 out of 135) being rated: Amazon, Apple, Facebook, Google, and Microsoft. Only a small sample of strings from Instagram, Twitter, and Yahoo were included (1, 3, and 5 respectively). No string presented in the study mentioned platform names.

Participants. In total, we had 818 participants begin the study. In order to ensure data quality, we removed participants who did not pass an attention check in the study (explained further in this section), participants who did not finish the survey, and participants who answered the same formality or compliance value throughout the entire survey. We were left with data from 512 participants. The participants primarily identified as female (n=267), 185 identified as male, and 60 identified as nonbinary or did not list their gender. While participants reported coming from 51 countries, the majority came from English speaking countries: United States (n=243), United Kingdom (n=26), Canada (n=24), India (n=16), Australia (n=14), China (n=12), Germany (n=11), South Korea (n=10). Other countries had fewer than 10 participants who completed the survey. Participants were on average 28.85 years of age (sd=14.80) and reported an average number of 15.01 years of education (sd=3.94).

Procedure. Participants provided informed consent and their demographic information, including age, gender, education level, and country. They were then shown a page explaining that the study consisted of three parts.

The first part of the study aimed at testing H.1 by gathering participants' impressions of the formality of security and privacy prompts. Each participant was given a random selection of 12 prompts from the 135 included in the study. Participants were asked to rate the 12 prompts on a 5-point scale with 1 labeled as "Very informal" and 5 labeled as "Very formal." Prompts were shown on separate pages in the study (see Figure 2).

The second part of the study aimed to confirm that participants could not reliably identify the platform that a prompt came from. If participants could identify the platform, this could bias compliance ratings (due to participants' attitudes

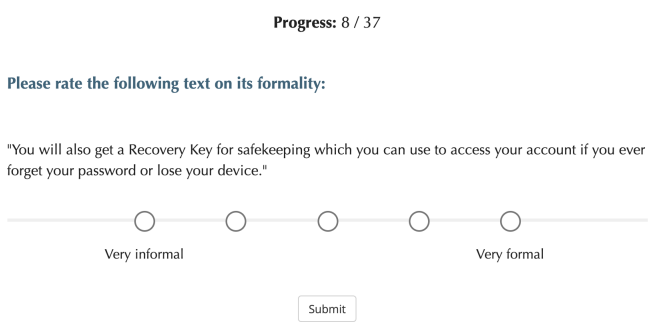


Fig. 2. Presentation of security and privacy prompts to participants in our online study.

towards those platforms in particular). For this part we included the prompts of four platforms with the most prompts: Amazon, Apple, Facebook, and Google. Participants were randomly shown 12 of 104 prompts (26 prompts for each platform), and asked to assign each to one of the four platforms by clicking on a button showing the platform's logo with a text label of the platform name. This was the only part of the study where platform was explicitly mentioned. In this part we included an attention check used to filter participants. The check was a prompt that explicitly mentioned Google (no other prompt mentioned platform name). The participant was expected to guess Google for this prompt.

The third part was designed to test H.2 by evaluating participants' intention to comply with specific prompts. Participants were given a random sample of twelve prompts, drawn from a set of 42 prompts that were selected by two authors as having a clear request (e.g., signing up for two factor authentication) across the 8 platforms used in the study. This set of prompts was a subset of all prompts coded as a request (Section 3.2). Each prompt presented was given its own page in the study, containing the question "How likely would you be to follow this prompt, based on its tone?" followed by the prompt and a 5-point scale with 1 labeled as "Very unlikely" and 5 as "Very likely."

At the end of the study, participants were presented with a personalized results page specifying their accuracy at guessing the source of prompts from the second part of the study, as well as how their average rating of perceived formality compared to that of other participants. To minimize the influence of presentation and font type on perceived formality, all parts of the study presented the prompts using the same font type and size in quotation marks (Figure 2).

Analysis. We analyzed whether people perceived the formality of the prompts across platforms as different (H.1) by constructing a mixed-effects linear regression model relating formality as a continuous dependent variable to the platform of origin as an independent variable. To evaluate the effect that our codes had on rated formality (H.1.1), we also included the prompts' tone code (e.g., casual), request type code (e.g., command, request, optional) and presence of technical language as independent variables in the model. Our model was trained on 6,456 datapoints, where each datapoint was a single rating of the level of formality in a string by a single participant.

Next, to analyze the relationship between the formality of a prompt and the likelihood that people comply with that prompt (H.2), we ran a second model with the compliance ratings of a prompt as the dependent variable and the average formality rating of that prompt as an independent variable. We also included the prompts' tone, request type, and presence of technical language as independent variables in the model to evaluate how these variables influence intended compliance (H.2.1). We controlled for confounding variables that might affect perceived formality and compliance by including participants' age, gender, and prompt length as additional independent variables. Participant ID was modeled

as a random effect. We used the average formality rating in this model because participants did not necessarily rate the same strings on formality and compliance (to avoid a familiarity effect). Our model was trained on 5,470 datapoints, where each datapoint was a single rating of compliance by a single participant.

Different actions requested by a prompt might have different levels of intended compliance (e.g., signing up for two factor authentication versus using a stronger password). These differences could impact our model by biasing compliance ratings. In order to control for the possible effect of different requested actions, we included the action, which we refer to as prompt type, in our compliance model (H.2) as an additional independent variable. Prompt type is defined as the action a prompt is requesting. We included 5 prompt types based on the requested actions in the prompts used for the study: Two-Step Verification, Password Strength, Unique Password, Physical Password, and Other (e.g., adding a profile picture). We control for prompt type in the compliance model but not in the formality model because the majority of the prompts used in the formality rating portion of the study did not have a requested action (e.g., "It may take a few minutes to arrive"). We built our models using the `statsmodels` and `pymer4` toolkits [27, 45].

4.2 Online Experiment Results

H.1: People perceive the formality of security prompts as different across platforms. Our results show a significant main effect of platform on formality rating ($F = 25.859, p < .001$) suggesting that people perceive the language formality used to communicate in their security and privacy interfaces differently across platforms. This confirms H.1.

The regression's beta coefficients in Table 4 show that, in comparison to Google (used as a baseline), Facebook, Instagram and Yahoo have significantly lower levels of perceived formality (see also Fig. 3). Microsoft, Apple, and Amazon had higher levels of perceived formality. Prompt length also had a significant effect on formality ($F = 94.01, p < 0.001$), suggesting that the longer the prompt is, the more formal it was perceived to be.

Both participant age and education had a significant effect on formality ratings ($F = 4.41, p = 0.036$ and $F = 4.41, p = 0.024$ respectively). This suggests that age and education level both negatively influence perceived formality: The older and more educated a participant is, the less formal they perceived the presented strings.

H.1.1: Strings with an authoritative or casual tone have lower perceived formality than strings with a professional tone. The codes from our qualitative codebook also had a significant effect on perceived formality. Our results show a significant main effect of tone on perceived formality ($F = 26.600, p < 0.001$.) We observe that when compared to authoritative strings, strings with a casual tone have lower perceived formality ($\beta = -0.305, p < 0.001$) while strings with a professional tone have higher perceived formality ($\beta = 0.352, p < 0.001$). This confirms hypothesis H.1.1, though with the caveat that participants seem to find casual prompts the least formal, while authoritative prompts are more formal than casual prompts but less formal than professional prompts. There were no prompts with the dialog code in the subset of prompts in this part of the study.

We also observed that prompts requesting similar actions could still receive markedly different formality ratings. Table 5 shows example prompts with their different average formality ratings. We see that prompts varied in how formal their language was, even when requesting the same action (e.g., "do not use dictionary words" versus "your password is too easy to guess").

H.2: Formality impacts people's intention to comply with security and privacy prompts. Table 6 reports the main effects of the model's independent variables on intended compliance. Formality had a significant effect on intended compliance ($F = 65.97, p < .001$), confirming H.2. The higher the average formality rating for a prompt was, the higher participants'

Table 4. Final regression model for predicting formality (H.1). Adjusted $R^2 = 0.437$. For tone, authoritative was used as a comparison point. For request, command was used as a comparison point. For platform, Google was used as comparison point. Participant ID is coded as a random variable.

Variable	β-coefficient	p-value
Constant	3.165	<.001
Tone: Casual	-0.319	<.001
Tone: Professional	0.341	<.001
Request: Optional	-0.072	<.001
Request: UserRequest	0.222	.154
Technical	-0.103	0.014
Prompt Length (characters)	0.006	.021
Platform: Microsoft	0.038	0.682
Platform: Facebook	-0.301	<.001
Platform: Apple	0.167	0.001
Platform: Amazon	0.401	<.001
Platform: Instagram	-0.796	<.001
Platform: Twitter	0.014	0.878
Platform: Yahoo	-0.341	<.001
Age	-0.004	0.036
Education	-0.014	0.024
Gender (Female)	-0.063	0.166

Table 5. Example prompts from the the dataset that request the same action but had different average ratings of formality.

Reported Information	Prompt	Formality
Password Strength	"Do not use dictionary words, your name, e-mail address, mobile phone number or other personal information that can be easily obtained."	3.86
	"Use at least 8 characters."	2.32
	"Your password is too easy to guess, try making it longer."	2.00
Two-Step Verification	"As long as the One Time Password (OTP) suppression cookie is present, a Sign-In from that browser or application will only require a password. (Note This option is enabled separately for each browser that you use.)"	4.47
	"Protect your account by enabling an additional security step using your personal device."	3.88
	"You can still use your password if your phone isn't handy."	1.78

reported intention to comply. We also found that prompt length, prompt type, platform of origin and participant age had a significant effect on reported compliance.

Participant age was the only demographic variable to have a significant association with compliance. The positive coefficient suggests that older participants were more likely to rate their compliance higher. Older adults usually struggle more with adopting security best practices like 2 factor authentication [19]. Our results might suggest that with age participants are more willing to comply with prompts. Alternatively, older adults might rate their compliance higher because they are basing their rating off of other people their age (i.e., a reference group effect).

Table 7 details the respective β coefficients for each variable. For example, formality's β coefficient of 0.553 suggests that for every point higher (on a scale from 1 to 5) the average formality rating was, intended compliance ratings were

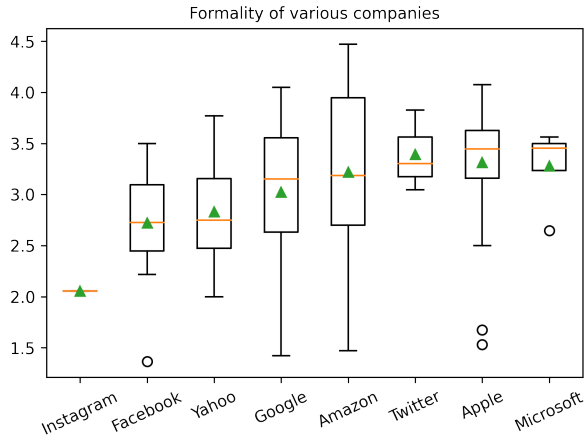


Fig. 3. The distribution of average formality ratings for the set of prompts from each of the eight platforms included in the online study. The y-axis shows the ratings on a scale from 1 = very informal to 5 = very formal.

Table 6. H.2: Main effects for each independent variable in the regression models predicting compliance. Prompt length and native English speaker are reported as control variables. Participant ID is coded as a random variable.

	F-statistic	***p-value***
Avg. Formality	65.97	<.001
Prompt Length	52.81	<.001
Prompt Type	77.35	<.001
Platform	11.98	<.001
Age	14.77	<.001
Tone	7.44	0.008
Technical Language	4.51	0.80
Request Style	0.843	0.379
Education	2.11	0.147
Gender	0.331	0.564

estimated to go up by 0.553 points (on a 1 to 5 scale). For platform, Yahoo, Google, Amazon and Facebook have prompts with higher levels of compliance compared to Apple, while Microsoft, Instagram and Twitter have lower.

The type of request a prompt made (i.e., prompt type) also had a significant effect on compliance, with password strength requests (e.g., not using dictionary words) having the highest compliance ($\beta = 0.558$).[1] Table 8 also includes samples of paired high and low formality prompts for the same prompt type, and their average formality and compliance ratings. We can see that even within the same request type (e.g., reusing a password), higher formality ratings are associated with higher compliance ratings.

This model has adjusted R^2 of 0.275, suggesting that while formality, prompt type, and length all play significant roles in self reported likelihood of compliance, these factors are not the complete story in what encourages users to comply with account security suggestions.

[1]We also ran a model that used prompt type as a random effect, another common method with mixed effects models for controlling for variation within a group of responses [9] (e.g., modeling reader response to different versions of a paper [22]). This explored if formality had an effect on compliance independent of the variations between prompt type. We found results similar to those reported in Table 7.

Table 7. Final regression model for predicting compliance (H.2). For platform, Apple was used as comparison point. For tone, authoritative was used as a comparison point. For request, command was used as a comparison point. For prompt type, 2 Step Verification was used as a comparison point. Adjusted $R^2 = 0.275$

Variable	β-coefficient	p-value
Avg. Formality (Likert-rating)	0.553	<.001
Prompt Length (characters)	-0.005	<.001
Tone: Casual	0.240	.001
Tone: Professional	-0.203	0.012
Request: Optional	-0.194	.149
Request: UserRequest	0.046	.507
Technical	0.084	.080
Prompt Type: Strength of Pass.	0.558	0.001
Prompt Type: Avoid Pass. reuse	-0.235	0.195
Prompt Type: Physical Pass. records	0.072	0.698
Prompt Type: Other	0.302	0.04
Platform: Microsoft	-0.235	0.006
Platform: Instagram	-0.131	0.400
Platform: Twitter	-0.033	0.746
Platform: Amazon	0.138	0.097
Platform: Yahoo	0.237	0.007
Platform: Google	0.282	<.001
Platform: Facebook	0.489	<.001
Age	0.007	<.001
Education	-0.011	0.147
Gender	-0.028	0.565

Table 8. Example strings from the rated dataset, including their average rating on formality and compliance.

Prompt Type	Prompt	Formality	Compliance
Unique Password	"Do not use the same password you have used with us previously."	3.55	3.43
	"Don't use a password you've used for other accounts or websites."	2.97	2.72
Physical Password	"We strongly recommend that you don't store your recovery code on a device."	3.48	3.32
	"Do not save it on your computer."	2.50	3.03
Password Strength	"To help keep your account safe, choose a strong password that's at least 8 characters."	3.04	3.97
	"Use at least 8 characters."	2.32	3.71

H.2.1: Strings with an authoritative or casual tone have lower intended compliance than strings with a professional tone. We additionally found that prompt tone codes had a significant impact on intended compliance ($F = 7.065, p < 0.001$). Compared to strings labeled as authoritative, strings labeled as casual had higher levels of intended compliance ($\beta = 0.185, p = 0.011$), and strings labelled as professional had lower levels of intended compliance ($\beta = -0.279, p = 0.012$.) This refutes hypothesis H.2.1, suggesting that although formality is associated with increased intended compliance, certain tones have less clear associations (e.g., professional language might not increase compliance even though it is often more formal than a casual tone). Furthermore, we see that using the form of the request (e.g. command vs presented as optional) does not have a significant influence on compliance ($F = 1.59, p = 0.203$).

Participant ability to identify prompt platform. One risk of this study was that participants might rate formality or their intention to comply differently depending on which platform they thought the prompt came from, even though the prompts were anonymized. We ruled out this risk by analyzing the data from the second part of our experiment in which participants had to assign prompts to one of four platforms: Amazon, Apple, Facebook, and Google. We found that out of 5,835 responses, 1,645 were correct, or a 28.19% success rate, where random guessing would expect a success rate of 25%. Furthermore, a logistic regression analysis with correct guesses as the binary dependent variable showed that platform does not have a significant main effect on correct guesses ($F = 1.017, p = 0.384$). Therefore, we conclude that the platform source is *not* a significant influence on the ability of participants to recognise a string's source, and that they are only slightly better than chance in knowing which platform a string comes from.

5 DISCUSSION

In this work we set out to understand how people perceive the formality of the language used in account security and privacy strings and how it influences their intention to comply with security prompts. Our user study confirmed our hypotheses that platforms use different levels of formality in their communication with users (H.1), and that these different levels of formality impact users' intentions to comply with suggested security and privacy practices (H.2). These results extend prior work by identifying a new context (security and privacy interfaces) where formal language can be more effective at encouraging compliance than informal language. Informal language is often used in more relaxed, conversational settings [24], such as online communities [35], or social media [42] and is usually less precise [24]. Formal language, on the other hand is usually perceived as more trustworthy [42], and can increase participant attention in online studies [3]. Our findings are in line with this prior work in that formal language may be perceived as more appropriate in one-way communication where users are asked to perform specific actions.

Our work also extends prior work on encouraging users to comply with security and privacy prompts by providing more precise information about security actions. More detailed prompts have been found to increase compliance [37], and users often perceive risks communicated concretely (e.g., a clear example of a risk) as more severe. Formal language is associated with more precise, concrete language [24], and in our study we found that longer prompts were usually rated as more formal, suggesting that one reason formality might increase compliance is because of the detail it can provide in prompts. At the same time, we also observed that prompt length was negatively associated with rated compliance. Considering that formality was positively associated with compliance, it might be that longer prompts that do not provide more detailed information instead discouraged participants from complying.

We also found that prompts from different platforms had widely varying language style, as reflected in differences across our manual codes and crowdsourced ratings. Instagram, the platform with the lowest rated formality, used a relatively high rate of casual strings (such as those that address users by mentioning their account names). In contrast, Amazon, the platform with one of the highest formality ratings, also used some of the most professional prompts. This suggests that addressing users personally contributes to the perception of low formality, while communication that includes professional language seems to be perceived as more formal.

Platforms can encourage users to follow security and privacy practices through security prompts, but understanding what language is best at clarifying security practices and persuading adoption is difficult. Should a platform strike a friendly, encouraging tone, or is an authoritative tone more compelling? Our study revealed that for the security and privacy prompts we explored, more formal language can lead to higher intended compliance. Platforms can use this information to design their security prompts with more formal language. One way platforms can do this is by manually changing the language in their security and privacy interfaces, such as by following the examples in our codebook and

results that describe which strings are perceived as high formality and which ones users are most likely to comply with. Platforms can also automatically rewrite sentences from informal to formal based on common edits found in style transfer datasets on formality [35, 39]. These approaches usually include edits such as expanding contractions (*"don't"* to *"do not"*), changing punctuation (*"!!!!"* to *"!"*), and paraphrasing (*"awesome"* to *"very nice"*). Platforms can use these common edits as a simple way of rewriting security and privacy interfaces to more formal language (e.g., changing "It'll be here in a jiffie!" to " It will be here in one minute.").

Language style is an important part of the overall image of a platform, with style guides and new tools to make such style consistent across many platform writers. In our analyses we saw evidence of this both in the tone of prompts in our dataset (e.g., Apple used more professional prompts than Twitter) and in formality ratings (e.g., Apple's prompts were on average rated as more formal than Twitter's). Our findings suggest that even in security prompts, such style differences are noticeable and can lead to downstream effects in compliance. Platforms can use our methodology of crowdsourcing formality ratings as a tool for making their overall image consist in these interfaces. Furthermore, researchers or platform writers interested in collecting language style ratings can use our methodology as a template.

6 LIMITATIONS AND FUTURE WORK

Most people would agree that it is good practice to comply with security and privacy measures when prompted, whether or not they use them in practice. For this reason, our participants might have over-reported their likelihood to comply with these prompts compared to their real life behavior. However, because this effect would be present across all compliance ratings, we believe it is unlikely to have impacted understanding the *relationship* between formality and compliance. Nevertheless, future research and validation through A/B tests are needed to confirm that a higher formality leads to actual compliance (as opposed to self-reported intentions).

Our study is also limited by the fact that we do not have information on whether or not participants comply with any of the given security and privacy prompts in "real-world" applications. For example, they may have rated a prompt suggesting they sign-up for two-step verification as "very unlikely to follow" because they currently do not use two-step verification. While our instructions for participants emphasized that they should focus on the tone of the prompt (rather than the action itself), this could have still influenced the results. This behavior could also disproportionately affect prompts based of the type of request, as previous research has shown that some security practices such as two factor authentication are significantly less common than other practices, such as using strong or unique passwords [26]. By including the type of prompt in our models, we are able to expose the formality-compliance link independent of these prompt types. However to solidify this, future studies are needed to control for people's previous choices to comply with security and privacy prompts.

Additionally, self reported behavior may differ from real world behavior because participants may face additional factors influencing their real world decisions, such as the desire to use the account or service that they are signing up for or logging on to use [8, 13, 21]. Prior work has analyzed the validity of self reported studies in reflecting real world behavior, finding that self reported intentions can provide valuable insight into real world behavior [14, 34]; however, Wash et al. [50] finds that self-reported password strength and uniqueness statistics can be inaccurate.

Finally, one potential limitation that our study faces is the possibility of priming effects in the study. Because participants are initially exposed to questions about formality, they have likely been primed to notice differences in formality within prompts. This may have an effect on their compliance ratings, as they are more likely to be mindful of the formality of the prompt types. In addition, the second part of the study where participants tried to identify the platform that a prompt came from might have primed participants to apply their perception of the platforms they were

guessing with intended compliance of security prompts. Our results showed that participants were only able to guess platforms correctly at a rate close to random chance, mitigating any bias on compliance based on platform.

We also hope that others will build on our work to investigate whether our findings are generalizable across demographic and geographic groups of people. Because languages can differ greatly in level of formality [23], one could expect speakers of highly formal languages to be more likely to comply with formal prompts and vice versa. Similarly, elderly people may be more likely to comply with formal language than younger people and this may be further impacted by native language [29]. To answer these questions, future efforts are needed to collect strings from security and privacy interfaces in other languages and to recruit a larger and more diverse sample.

Finally, language is only one of many design aspects of an interface that may influence perceived formality and compliance. For example, the choice of colors can influence how people perceive a design [47]. Even the typography may play a role [30]. Hence, we are excited to see our work being extended to investigate the influence of other design choices on formality and compliance.

7 CONCLUSION

In this paper, we examined how the language formality of security prompts varies across major technology platforms, and how these variations in formality impact the likelihood of an individual to comply with these prompts. We find that platforms present security prompts with significantly different levels of formality. We also find that increased formality in security prompts is associated with an increase in self-reported intention to comply with those prompts. This suggests that formality is an important factor in the design of security and privacy prompts. In addition to these findings, we contribute a dataset containing 1,817 strings in security and privacy interfaces across 13 different platforms, along participant ratings of compliance and formality on 135 prompts, providing sources for further research to examine the text that platforms use for their security interfaces.

ACKNOWLEDGMENTS

This work was partially funded by Google and the National Science Foundation, proposal #2006104. We thank the LabintheWild participants who drive this research. We also thank the reviewers for their time and input.

DATASETS

We make available two datasets for download at https://www.labinthewild.org/data/: (1) the coded string dataset, and (2) the dataset with participant ratings on formality and compliance.

REFERENCES

[1] Alessandro Acquisti, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, Norman Sadeh, Florian Schaub, Manya Sleeper, Yang Wang, and Shomir Wilson. 2017. Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online. *ACM Comput. Surv.* 50, 3, Article 44 (2017), 41 pages. https://doi.org/10.1145/3054926

[2] Tal August, Dallas Card, Gary Hsieh, Noah A. Smith, and Katharina Reinecke. 2020. Explain like I Am a Scientist: The Linguistic Barriers of Entry to r/Science. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA). Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376524

[3] Tal August and Katharina Reinecke. 2019. Pay Attention, Please: Formal Language Improves Attention in Volunteer and Paid Online Experiments. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3290605.3300478

[4] Cristian Bravo-Lillo, Lorrie Faith Cranor, Julie Downs, and Saranga Komanduri. 2011. Bridging the Gap in Computer Security Warnings: A Mental Model Approach. *IEEE Security and Privacy* 9, 2 (2011), 18–26. https://doi.org/10.1109/MSP.2010.198

[5] Penelope E. Brown and S. Levinson. 1987. Politeness: some universals in language usage.

[6] Moira Burke and Robert Kraut. 2008. Mind Your Ps and Qs: The Impact of Politeness and Rudeness in Online Communities. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work (CSCW '08)*. Association for Computing Machinery, New York, NY, USA, 281–284. https://doi.org/10.1145/1460563.1460609

[7] Ana Caraban, Evangelos Karapanos, Daniel Gonçalves, and Pedro Campos. 2019. 23 Ways to Nudge: A Review of Technology-Mediated Nudging in Human-Computer Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk). Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3290605.3300733

[8] Jessica Colnago, Summer Devlin, Maggie Oates, Chelse Swoopes, Lujo Bauer, Lorrie Cranor, and Nicolas Christin. 2018. "It's Not Actually That Horrible": Exploring Adoption of Two-Factor Authentication at a University. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada). Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3173574.3174030

[9] Robert Cudeck. 1996. Mixed-effects Models in the Study of Individual Differences with Repeated Measures Data. *Multivariate behavioral research* 31 3 (1996), 371–403.

[10] Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. A computational approach to politeness with application to social factors. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Sofia, Bulgaria, 250–259. https://www.aclweb.org/anthology/P13-1025

[11] Sauvik Das, Adam D.I. Kramer, Laura A. Dabbish, and Jason I. Hong. 2014. Increasing Security Sensitivity With Social Proof: A Large-Scale Experimental Confirmation. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security* (Scottsdale, Arizona, USA). Association for Computing Machinery, New York, NY, USA, 739–749. https://doi.org/10.1145/2660267.2660271

[12] James Price Dillard and Michael Pfau. 2002. The Persuasion Handbook: Developments in Theory and Practice.

[13] Jonathan Dutson, Danny Allen, Dennis Eggett, and Kent Seamons. 2019. Don't Punish all of us: Measuring User Attitudes about Two-Factor Authentication. In *2019 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, IEEE, EuroS, 119–128.

[14] Cori Faklaris, Laura Dabbish, and Jason I. Hong. 2019. A Self-Report Measure of End-User Security Attitudes (SA-6). In *Proceedings of the Fifteenth USENIX Conference on Usable Privacy and Security* (Santa Clara, CA, USA) *(SOUPS'19)*. USENIX Association, USA, 61–77.

[15] Adrienne Porter Felt, Alex Ainslie, Robert W. Reeder, Sunny Consolvo, Somas Thyagaraja, Alan Bettes, Helen Harris, and Jeff Grimes. 2015. Improving SSL Warnings: Comprehension and Adherence. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 2893–2902. https://doi.org/10.1145/2702123.2702442

[16] Adrienne Porter Felt, Robert W. Reeder, Alex Ainslie, Helen Harris, Max Walker, Christopher Thompson, Mustafa Embre Acer, Elisabeth Morant, and Sunny Consolvo. 2016. Rethinking Connection Security Indicators. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*. USENIX Association, Denver, CO, 1–14. https://www.usenix.org/conference/soups2016/technical-sessions/presentation/porter-felt

[17] S. M. Furnell, P. Bryant, and A. D. Phippen. 2007. Assessing the Security Perceptions of Personal Internet Users. *Comput. Secur.* 26, 5 (2007), 410–417. https://doi.org/10.1016/j.cose.2007.03.001

[18] Maximilian Golla, Grant Ho, Marika Lohmus, Monica Pulluri, and Elissa M. Redmiles. 2021. Driving 2FA Adoption at Scale: Optimizing Two-Factor Authentication Notification Design Patterns. In *30th USENIX Security Symposium (USENIX Security 21)*. USENIX Association, usenix, 109–126. https://www.usenix.org/conference/usenixsecurity21/presentation/golla

[19] Maximilian Golla, Grant Ho, Marika Lohmus, Monica Pulluri, and Elissa M Redmiles. 2021. Driving 2FA Adoption at Scale: Optimizing Two-Factor Authentication Notification Design Patterns. In *30th {USENIX} Security Symposium ({USENIX} Security 21)*. USENIX Association, usenix, 109–126.

[20] Marian Harbach, Markus Hettig, Susanne Weber, and Matthew Smith. 2014. Using Personal Examples to Improve Risk Communication for Security and Privacy Decisions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 2647–2656. https://doi.org/10.1145/2556288.2556978

[21] Marian Harbach, Emanuel Von Zezschwitz, Andreas Fichtner, Alexander De Luca, and Matthew Smith. 2014. It's a Hard Lock Life: A Field Study of Smartphone (Un)Locking Behavior and Risk Perception. In *Proceedings of the Tenth USENIX Conference on Usable Privacy and Security* (Menlo Park, CA). USENIX Association, USA, 213–230.

[22] Andrew Head, Kyle Lo, Dongyeop Kang, Raymond Fok, Sam Skjonsberg, Daniel S. Weld, and Marti A. Hearst. 2021. Augmenting Scientific Papers with Just-in-Time, Position-Sensitive Definitions of Terms and Symbols. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* 1 (2021), 1 pages.

[23] Francis Heylighen and Jean-Marc Dewaele. 1999. Formality of language: definition, measurement and behavioral determinants. *Interner Bericht, Center "Leo Apostel", Vrije Universiteit Brüssel* 4 (1999), 1.

[24] Francis Heylighen and Jean-Marc Dewaele. 2002. Variation in the contextuality of language: An empirical measure. *Foundations of science* 7, 3 (2002), 293–340.

[25] Mitch Holmes and Jacques Ophoff. 2019. Online security behaviour: factors influencing intention to adopt two-factor authentication. In *Proceedings of the 14th International conference on cyber warfare and security, ICCWS 2019 (Proceedings)*, Noelle van der Waag-Cowling and Louise Leenen (Eds.). Academic Conferences and Publishing International Limited, ACPIL, 123–132. 14th International Conference on Cyber Warfare and Security, ICCWS 2019 ; Conference date: 28-02-2019 Through 01-03-2019.

[26] Iulia Ion, Rob Reeder, and Sunny Consolvo. 2015. "...No One Can Hack My Mind": Comparing Expert and Non-Expert Security Practices. In *Proceedings of the Eleventh USENIX Conference on Usable Privacy and Security* (Ottawa, Canada) *(SOUPS '15)*. USENIX Association, USA, 327–346.

[27] Eshin Jolly. 2018. Pymer4: Connecting R and Python for Linear Mixed Modeling. *Journal of Open Source Software* 3, 31 (2018), 862. https://doi.org/10.21105/joss.00862

[28] Shibamouli Lahiri. 2015. SQUINKY! A Corpus of Sentence-level Formality, Informativeness, and Implicature. *arXiv* 1 (2015), 1. arXiv:1506.02306 https://arxiv.org/pdf/1506.02306.pdfhttp://arxiv.org/abs/1506.02306

[29] Shizuka Lauwereyns. 2002. Hedges in Japanese conversation: The influence of age, sex, and formality. *Language Variation and Change* 14, 2 (2002), 239–259.

[30] Jo Mackiewicz and Rachel Moeller. 2004. Why people perceive typefaces to have different personalities. In *International Professional Communication Conference, 2004. IPCC 2004. Proceedings.* IEEE, IPCC, 304–313.

[31] J. C. Marquié, L. Jourdan-Boddaert, and N. Huet. 2002. Do older adults underestimate their actual computer knowledge? *Behaviour & Information Technology* 21, 4 (2002), 273–280.

[32] Tara Matthews, Kerwell Liao, Anna Turner, Marianne Berkovich, Robert Reeder, and Sunny Consolvo. 2016. "She'll Just Grab Any Device That's Closer": A Study of Everyday Device and Account Sharing in Households. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5921–5932. https://doi.org/10.1145/2858036.2858051

[33] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-Rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 72 (nov 2019), 23 pages. https://doi.org/10.1145/3359174

[34] Rand DG Mosleh M, Pennycook G. 2020. Self-reported willingness to share political news articles in online surveys correlates with actual sharing on Twitter. *PLoS ONE* 1 (2020), 1. https://doi.org/10.1371/journal.pone.0228882

[35] Ellie Pavlick and Joel Tetreault. 2016. An Empirical Analysis of Formality in Online Communication. *Transactions of the Association for Computational Linguistics* 4 (2016), 61–74. https://doi.org/10.1162/tacl_a_00083

[36] Kelly Peterson, Matt Hohensee, and Fei Xia. 2011. Email Formality in the Workplace: A Case Study on the Enron Corpus. In *Proceedings of the Workshop on Language in Social Media (LSM 2011)*. Association for Computational Linguistics, Portland, Oregon, 86–95. https://www.aclweb.org/anthology/W11-0711

[37] Sarah Prange, Niklas Thiem, Michael Fröhlich, and Florian Alt. 2022. "Secure settings are quick and easy!" – Motivating End-Users to Choose Secure Smart Home Configurations. *Proceedings of the 2022 International Conference on Advanced Visual Interfaces* 1 (2022), 1.

[38] Anabel Quan-Haase and Dennis Ho. 2020. Online privacy concerns and privacy protection strategies among older adults in East York, Canada. *Journal of the Association for Information Science and Technology* 71 (2020), 1. https://doi.org/10.1002/asi.24364

[39] Sudha Rao and Joel Tetreault. 2018. Dear Sir or Madam, May I introduce the YAFC Corpus: Corpus, Benchmarks and Metrics for Formality Style Transfer. *CoRR* abs/1803.06535 (2018), 1. arXiv:1803.06535 http://arxiv.org/abs/1803.06535

[40] Elissa M. Redmiles, Michelle L. Mazurek, and John P. Dickerson. 2018. Dancing Pigs or Externalities? Measuring the Rationality of Security Decisions. In *Proceedings of the 2018 ACM Conference on Economics and Computation* (Ithaca, NY, USA). Association for Computing Machinery, New York, NY, USA, 215–232. https://doi.org/10.1145/3219166.3219185

[41] Katharina Reinecke and Krzysztof Z. Gajos. 2015. LabintheWild: Conducting Large-Scale Online Experiments With Uncompensated Samples. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 1364–1378. https://doi.org/10.1145/2675133.2675246

[42] Kristina M Rennekamp and Patrick D Witz. 2020. Linguistic Formality and Audience Engagement: Investors' Reactions to Characteristics of Social Media Disclosures. *Contemporary Accounting Research* 1 (2020), 1.

[43] Navdeep S. Sahni, S. Wheeler, and Pradeep Chintagunta. 2018. Personalization in Email Marketing: The Role of Noninformative Advertising Content. *Mark. Sci.* 37 (2018), 236–258.

[44] Yukiko Sawaya, Mahmood Sharif, Nicolas Christin, Ayumu Kubota, Akihiro Nakarai, and Akira Yamada. 2017. Self-Confidence Trumps Knowledge: A Cross-Cultural Study of Security Behavior. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA). Association for Computing Machinery, New York, NY, USA, 2202–2214. https://doi.org/10.1145/3025453.3025926

[45] Skipper Seabold and Josef Perktold. 2010. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, Stéfan van der Walt and Jarrod Millman (Eds.). scipy, North Carolina Chapel Hill, 92 – 96. https://doi.org/10.25080/Majora-92bf1922-011

[46] Brian Stanton, Mary Theofanos, Sandra Prettyman, and Susanne Furman. 2016. Security Fatigue. *IT Professional* 18 (2016), 26–32. https://doi.org/10.1109/MITP.2016.84

[47] Patricia Valdez and Albert Mehrabian. 1994. Effects of color on emotions. *Journal of experimental psychology: General* 123, 4 (1994), 394.

[48] Paul van Schaik, Debora Jeske, Joseph Onibokun, Lynne Coventry, Jurjen Jansen, and Petko Kusev. 2017. Risk perceptions of cyber-security and precautionary behaviour. *Computers in Human Behavior* 75 (2017), 547–559.

[49] Zhao Wang and Aron Culotta. 2018. When Do Words Matter? Understanding the Impact of Lexical Choice on Audience Perception Using Individual Treatment Effect Estimation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. AAAI, AAAI, 7233–7240.

[50] Rick Wash, Emilee Rader, Ruthie Berman, and Zac Wellmer. 2016. Understanding Password Choices: How Frequently Entered Passwords Are Re-Used across Websites. In *Proceedings of the Twelfth USENIX Conference on Usable Privacy and Security (SOUPS '16)*. USENIX Association, USA, 175–188.

[51] Alma Whitten and J Doug Tygar. 1998. *Usability of security: A case study.* Technical Report. CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF COMPUTER SCIENCE.

[52] Jiaheng Xie, Bin Zhang, Susan Brown, and Daniel Zeng. 2021. Write Like a Pro or an Amateur? Effect of Medical Language Formality. *ACM Trans. Manage. Inf. Syst.* 12, 3, Article 24 (2021), 25 pages. https://doi.org/10.1145/3458752

[53] Yixin Zou, Kevin Roundy, Acar Tamersoy, Saurabh Shintre, Johann Roturier, and Florian Schaub. 2020. Examining the Adoption and Abandonment of Security, Privacy, and Identity Theft Protection Practices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA). Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376570