Measuring and increasing the reach of security information through online media

Sruti Bhagavatula

CMU-ISR-21-106

September 2021

Institute for Software Research
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Lujo Bauer, Chair
Nicolas Christin
Timothy Libert (CMU & Google)
Apu Kapadia (Indiana University Bloomington)

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Societal Computing.

The research reported in this thesis has been supported in part by the Carnegie Mellon University CyLab Security and Privacy Institute. Parts of the datasets used throughout this thesis were created through work supported by the National Security Agency under Award No. H9823018D0008.

The views and conclusions contained in this document are those of the author, and should not be interpreted as representing the official policies, either expressed or implied, of any sponsoring institution, the U.S. government, or any other entity.

Keywords: measurement, social networks, information spread, security, privacy, security incidents, security behavior

For my loving family and friends

Abstract

With the growing number of technologies that have developed over the past several years and the similarly growing number of cyber attacks, people should ideally be aware of how to keep their information and systems safe. In general, awareness of security and privacy best practices is important for developing good security habits. Learning about real-world security incidents and data breaches can also alert people to the ways in which their information is vulnerable online, thus playing a significant role in encouraging safe security behavior online. In addition to awareness, it is important for people to take action to improve the security of their systems, particularly in the wake of a security incident or data breach. While prior work has been able to study problems about security awareness and incidents within a broad scope due to the affordances of self-reported methodologies, such studies largely relied on hypothetical or experimental scenarios.

In this thesis, we take steps towards (1) filling in the gap of a missing empirical understanding of engagement and action with security and privacy events through measurable behaviors, (2) understanding the effectiveness of social media as a platform for increasing the dissemination of security and privacy advice and for encouraging action, and (3) providing specific guidance for how security and privacy information may be shared on social media to encourage engagement and re-distribution.

Through measurements of real-world browsing and password data, we first show that online engagement with content related to large-scale security and privacy incidents is rare and that very few factors may encourage people to try to read more about incidents. We then show, by specifically analyzing password data, that people rarely take action after password breaches, much less action that is constructive, even when the breach definitely affected them. In understanding social media's effectiveness for disseminating security and privacy information, we find that discussions about security and privacy are scarce on Facebook and Twitter and that when these topics are discussed, they are often not discussed constructively. Finally, by analyzing Reddit posts about security and privacy, we identify and shed light on how security and privacy information may be shared on social media such that it garners wider spread and effectiveness.

Acknowledgments

The PhD has been a long, tough, and rewarding journey and has taught me about aspects of life that I may not otherwise have gotten experience with. This journey would not have been possible without the support of many people.

I am especially thankful to my advisor Lujo Bauer. From him, I have learned how to think like a scientist and to always critically analyze assumptions and approaches. He has always been willing to go deep into the nitty-gritty of research with me, from looking over code to annotating social media posts, which often made the labor-intensive tasks feel less solitary. In addition, he has always made time if I needed help or needed to talk and has been supportive and encouraging of my teaching interests, which has been vital to shaping my teaching experiences. I also thank my committee members Nicolas Christin, Timothy Libert, and Apu Kapadia. I am grateful to Nicolas for the fruitful discussions and feedback on research and for giving me the formative opportunity to build up my first teaching experience at CMU. I am grateful to Tim for making time to talk about my ideas, brainstorming research ideas with me, and being candid about his thoughts. Finally, I am grateful to Apu, for being a significant mentor to me for most of my PhD and for his positive and encouraging attitude during the highs and lows of research and the PhD.

I am fortunate to have worked with the following collaborators throughout my PhD, from whom I have learned about various aspects of research: Billy Melicher, Camille Cobb, Hana Habib, Joshua Tan, Kalil Anderson Garrett, Lorrie Cranor, Mahmood Sharif, Martin Degeling, Mattias Beckerle, Michael Reiter, Michelle Mazurek, Norman Sadeh, Pardis Emami-Naeni, Ranjita Bhagwan, and Saikat Guha. The work in this thesis would also not have been possible without the help of Jeremy Thomas and Sarah Pearman who spent hours with me teaching me about the SBO database and helping me figure out why my data collection code wasn't working.

I am indebted to the people who have helped me on my teaching journey. I was a TA for Nicolas Christin, Giulia Fanti, and my advisor, Lujo Bauer all of whom who have always given me multiple opportunities to teach and mentor to gain more teaching experience, experiences which have greatly shaped my teaching philosophy today. I am also grateful to Iliano Cervesato for being my co-instructor and mentor for my first experience as an instructor of a large core undergraduate course and for the valuable discussions and advice about teaching careers.

My foray into a PhD would not have been possible without the encouragement and support of my undergraduate mentors from UIC, Chris Kanich and Dale Reed. Working with Chris, I got introduced to a research experience in which I got to own a project from start to finish, making me realize how much I enjoyed research and my desire to pursue a PhD. From helping me decide where and how to apply for grad school to talking through my decisions with me, his guidance directly helped me get to where I am today. Dale has been a mentor from the start of my university computer science career and has always encouraged me in my teaching and leadership endeavors, which ultimately shaped my confidence in a major way. I am also grateful to my uncle, Prasad Calyam ("Babai"), for helping me with my grad school

applications and providing me with perspectives on the PhD and academia.

My time at CMU would not be what it was if not for the friends in my research group that have come and gone over the years: Andy Gallardo, Billy Melicher, Brian Singer, Camille Cobb, Clement Fung, Joshua Tan, Kalil Anderson Garrett, Keane Lucas, Mahmood Sharif, Michael Stroucken, Natalie Janosik, Nuno Sabino, Trevor Kann, and Weiran Lin. Special thanks to my cubemates at different points in time—Mahmood, Clement, and Keane—for making cube life so fun and for being so supportive. (One day we'll karaoke again, Clement!) The times we have all spent talking in the cubicles about life, sharing secrets of the trade about managing PhD stress, and more recently, our outdoor hangouts and that one game of chess (thanks, Trevor!) have been some of the fondest memories I have of being a PhD student.

Grad school has given me friends for life without whom my life in Pittsburgh and much of my 20s would not have been what they were. In particular, I'm thankful for Aaron Harlap, Abhilasha Jain, Aymeric Fromherz, Camille Cobb, Janos Szurdi, Joshua Tan, Mahmood Sharif, Orsi Kovacs, Pardis Emami-Naeni, and Rajat Kateja for always being there for me. Pardis has gone through the ups and downs of life with me both related to the PhD and much beyond and has always been my source of encouragement to be bold, sage advice, comfort, and sarcasm. Mahmood has been a mom and grandpa to me at the same time and somehow managed to survive my constant singing (remember, "Cheerleader"?) and annoying restaurant ordering habits. Aymeric has been my cheese night and "just one beverage" buddy and I could always count on him for much-needed wind-downs, heart-to-hearts, and listening to the "Wicked" soundtrack with me (we never did get to jam one last time!). Janos has always been the wise one in our group and always up for mischief and deep conversations. Orsi has always been an encouraging cheerleader and positive force in my life. Abhilasha was just a text away in the neighboring apartment and I'm thankful for our heart-to-hearts on a minute's notice, not to mention the long therapeutic tea breaks. I could always count on Rajat to stop by for a "quick" break or be a listener in a much-needed venting session or philosophical discussion from the apartment next door. And Camille has been a great pandemic buddy; our conversations about life and future plans have helped me survive the past year. I'm also thankful to other friends who have made my time at CMU and in Pittsburgh memorable: Akshay Gadre, Ankur Mallick, Antonis Manousis, Samarth Gupta, Sandeep D'souza, Sanghamitra Dutta, and Soo-jin Moon.

Other friends who are not at CMU have also been an important part of my life during this time. Aparna Varma has been one of my best friends and support systems even from almost 3,000 miles away. I'm not sure what I would do without our constant Hangouts messages, wonderings about Srubha, and totally sensical conversations. Sadhana Ravikumar, my roommate in Pittsburgh for the first two years of my PhD, has continued to be one of my closest friends and confidantes even after we ended up in different cities. I'm so thankful that she has always been just a text away to FaceTime to catch up or discuss life and marvel at how much we've grown. Sridivya ("Diva") Madem has always encouraged me to be strong and take leaps, for which I am very grateful. I am also thankful to my #lolwics buddies, Devina Dhawan

and Siham Hussein, for being some of my most ferocious cheerleaders and for always helping me put my achievements and failures into perspective. My roommate, Neeraja Gupta, my roommate for much of my PhD, has been a big presence in my life, always encouraging me to be bold, and willing to complain about grad school with me. Sai Vikneshwar has always made time to check in on how I'm doing, discuss Indian music with me, and chat about life and philosophy. Two of my closest friends from high school in Bengaluru, Aditi Bharatee and Maitri Yadav, have also continued to be important presences in my life, helping me to get through the past several years.

My piano teacher at CMU, Carla Larocca has been a source of encouragement and support since I first started with her five and a half years ago. Not only has she been a mentor and a friend to me, but my lessons with her have helped me learn several pieces which I'm excited to call another accomplishment from the past several years. I'm additionally thankful for the whole SCS Graduate Musical gang, for making each semester so enjoyable and for the karaoke sessions during which we belted out Idina Menzel and other ambitious songs from musicals. The CyLab staff has also been amazing throughout my time at CMU. Brigette Bernagozzi, Brittany Frost, and Karen Lindenfelser have always been especially kind and have gone out of their way to provide help when needed. Thanks also, Karen, for always encouraging my musical and drama interests.

My family has always consisted of my biggest cheerleaders. Without their wholehearted support, I would have never been able to embark upon the PhD journey so confidently. My Amma and Nanna have been the definition of supportive parents while also being my friends. They were always just a phone call away, whether I needed a listening ear or needed some parental guidance or whether my mom needed to recommend movies to me. I am grateful to my sister and brother, Swati ("Ka") and Swaroop ("Papps"), and my brother-in-law Hemant ("Jeej"), for always encouraging me and providing different perspectives on life, not to mention the nostalgia-inducing discussions of anime, games, and Spongebob that have kept me going. I am especially thankful for my almost two-year-old nephew Viraaj ("Babylu"), for always laughing when he sees me on video and updating me on his life in his own language, brightening up my day in a second. I've gotten through many exhausting days just by opening up videos of him and seeing his sweet and happy (while also naughty) face.

Finally, I am incredibly lucky to have had Krishna by my side for the better half of my PhD. He has been my rock and always made me feel empowered and I'm so happy I could lean on him through the happy and tough times. Additionally, our ridiculous inside jokes and "Concentration" games have increased my lifespan, which is a fortunate side effect.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Computers in all forms are becoming ubiquitous. Whether it is a mobile phone, a tablet, or a desktop or laptop, it is commonplace for a household to have at least one computer with 92% of households in the US having a computer as of 2018 [149]. However, with owning a computer comes great power but also great responsibility. Anyone who works with computers should ideally keep their computers safe from cyberattacks, their data protected, and their online data and accounts secure. Furthermore, with the growth of new technology and introduction of many new privacy-violating services, the number of dimensions across which security and privacy need to be maintained is also growing. An important step towards ensuring the security of people's digital devices and data is for system designers to relieve the burden of implementing security from the users of the system by prioritizing security and usability from the start [34, 197]. However, while it is ideal for users to be shielded from the nuts and bolts of implementing and maintaining security, the current state of the digital world requires security awareness on the part of the users [43]. For example, defending against social engineering attacks like phishing continues to rely on a degree of security awareness [65, 131]. Moreover, with security breaches becoming more rampant, it is especially important that people are aware of these incidents and how their occurrences apply to their own systems and data, taking constructive action when necessary [43, 115, 135]. Security awareness in general remains essential for people to have the motivation, knowledge, and agency to keep their systems and data secure and private [43, 115, 135].

Prior work found that there is an economic and demographic divide by which groups receive security and privacy advice [168, 189, 190]. It has also been found that people receive this advice from a variety of sources depending on their age group, where some sources may be more pervasive or accurate than others [76, 79, 168, 185, 189, 190] In fact, for specifically security incidents or data breaches, researchers found through surveys that people received varying types and amounts of advice depending on the sources of the information. They took different actions based on the source and only under certain circumstances [126, 235]. Even though many computer users do not receive the right information or advice, there is an abundance of security and privacy advice publicly available online [192]. Recent work has shown that when users in a study were presented with different advice articles available online, many of them perceived most of the advice they were shown to be somewhat actionable. However, often the articles provided multiple pieces of advice, leading participants to feel burdened with the decision on which pieces of advice to follow, serving as a bottleneck towards constructive action [192].

Social media has been shown to be an effective medium for discussion of security and privacy and encouraging constructive security behavior [77, 79, 81, 85]. For example, research has shown that Twitter has been useful for sharing security experiences and stories [85]. Other research found under experimental settings, that the practice of publicizing one's security practices to their friends in a large social network such as Facebook was correlated with their friends adopting similar practices [77, 79, 81].

This body of existing work is necessary to understand how users perceive security and privacy advice, what channels could be the most effective for promoting advice and behavior, who is and is not receiving this advice, and what kind of security-enhancing actions people take. Due to the self-reported and experimental methodologies used in prior work, these studies were able to study questions about engagement with security information within a large scope while also embodying the limitations of these methodologies [92, 111, 114, 194, 218, 224]. In addition, social media, though having been shown to be effective for encouraging security practices, has only been studied under experimental conditions with explicit social interventions [77, 81]. Studying real-world data would be a useful complement to prior studies that would provide us with an understanding of the actual extent of awareness and constructive action outside of hypothetical scenarios. Moreover, it seems promising to determine a path forward that combines and utilizes the findings of prior work in how to improve user safety through security awareness.

This thesis first takes a step towards developing an understanding of the actual extent of security awareness and constructive actions people take. It then aims to understand the ability of security and privacy information disseminated via social media to influence people to adopt better security practices. We first measure security awareness in terms of the extent of awareness about security incidents and the actions taken in the wake of password data breaches by studying browsing and password data collected from the computers of a large set of participants. We next study how security and privacy is talked about as part of social network activity (on Facebook and Twitter) for a set of participants, the prevalence of such content, along with how the consumption of such posts can be related to security behavior. Finally, to enable security and privacy information to reach more people and receive more engagement, we distill effective recommendations for how security and privacy information can be presented and distributed in social network.

1.1 Thesis statement

Real-world data shows that security and privacy information online is insufficiently disseminated to computer users and that they often do not engage further with this content even when they encounter it. Computer users also do not take security-enhancing action even when necessary for their security. However, security and privacy awareness can be increased through social media by sharing content exhibiting specific properties related to its text, tone, and visual elements.

We support this statement with the following four research studies:

Online engagement with security incident information[1] To study how much people engage with security information online, this project studies a specific type of security information: information about security incidents. Here, we empirically measured awareness and engagement with security incident information through real-world online browsing data. For a set of six security incidents, we found that only 16% of participants read about even one incident by visiting an incident-related page on the web and the people likely to have read about them were skewed towards older or more technically-savvy people. We further found that few people tried to learn more about an incident by reading more than one follow-up article about it. However, we found in the cases where people took the most action after reading about an incident, the articles were written with a positive sentiment. Very few other features were associated with trying to learn more about incidents.

Taking action after password breaches[2] In this study, we analyzed behavior surrounding a specific type of security incident: password breaches. We studied nine password breaches, and through real-world password data we examined how often people took action after these breaches and how effective these changes were in protecting affected people from further harm. We found that even when breaches definitely affected participants, they often did not change their password on the affected domains. For the few that did, their new passwords were often similar enough to their other account passwords that their new passwords on the breached domains could still be guessed.

The scarcity and ambiguity of security advice on social media[3] This project aimed to assess social media's effectiveness at sharing security information and encouraging healthy behavior through real social media data. We studied how often discussions about security and privacy came up on social media and found that for a set of about 300,000 posts, less than 0.09% of them were about security and privacy. We further studied what these discussions of security and privacy looked like and found that security and privacy was most often talked about either as not the main topic or jokingly and only rarely in a constructive manner. Exploratory statistical results further showed that the amount of security and privacy content people interacted with was not sufficient to consistently encourage healthy security behavior.

Improving the spread of security and privacy information in social networks This project studied, within Reddit, what makes posts garner more engagement and in turn spread further. We studied this by analyzing the properties a large set of Reddit posts about security and privacy. We found that posts that conveyed strong emotions or sentiment (e.g., opinions that conveyed anger, joy, or tentative emotions) contributed to a higher spread which we measured through the amount of engagement with these posts. Additionally, visual attributes on Reddit posts (e.g., in the form of emojis or "flair" which are stickers or icons next to posters and posts) were associated with higher spread. The length of the posted content was also important; longer posts were likely more

[1] Part of this work appeared at the IEEE Workshop on Technology and Consumer Protection (ConPro), 2021) [54] and the full study will appear at the European Symposium on Usable Security 2021 [55].

[2] This work appeared at the IEEE Workshop on Technology and Consumer Protection (ConPro), 2020) [53].

[3] This work is currently under submission [56].

thoughtfully written or well-written and hence, garnered more attention. However, engagement decreased when the content relied largely on links instead of text. Based on these findings, we provide recommendations and insight into how to distribute security and privacy information in social networks in the form of posts.

1.2 Outline

We first discuss background and work related to the topics of the above projects (Chapter 2). We next describe the study about online engagement with security incidents (Chapter 3) and actions after password breaches (Chapter 4). We then describe the work related to the scarcity and ambiguity of security information on social media (Chapter 5) and finally, we discuss how security and privacy information can be spread further in social networks (Chapter 6). We conclude with a summary of the contributions in this thesis and a discussion of future work (Chapter 7).

Chapter 2

Background and related work

In this chapter, we provide background and discussion of work related to the components of this thesis. Chapters 3 and 4tackle online awareness of security via security incident awareness. Therefore, we first discuss previous work regarding engagement with security incidents and data breaches. Next, a common theme of this thesis is the dissemination of security and privacy advice. Given this theme, the second topic for which we provide background is the dissemination of security and privacy information and notifications and their role in encouraging safe security behavior. We then discuss existing work about social influence in social networks and how social media has been used as an educational tool to support the motivation and background behind the final chapters in the thesis (Chapters 5 and 6), which investigate security and privacy discussions in social networks. We then discuss research on information spread in social networks in topic domains beyond security and privacy. Next, since Chapters 3, 4, and 5 measure and analyze security behavior, we provide background on security behaviors measured in prior work and approaches to measuring such behavior. Finally, we describe information campaigns in other fields, particularly public health, that can inform the design of future cybersecurity campaigns to increase public security awareness.

2.1 Engagement with security incidents

Much of the existing work on security incidents studies how people interact with incidents such as data breaches. Of particular interest is how people perceive data breaches and notifications and the risks involved [126, 235], what influences people to take action after a breach [126, 235], and how people are informed of data breaches [32, 79, 126, 235]. Prior work in this space highlighted that people are often not aware of breaches or their implications and that they may or may not take action depending on the source of their information.

Karunakaran et al. surveyed people about their comprehensions of the risks of data breaches and their sentiments towards remediation steps [126]. Out of those surveyed, 93% of participants indicated that they understood the implications of data breaches and that notifications were the most popular desired remediation step. In particular, 83% wanted affected companies to send immediate notifications to those affected. Participants were also asked about their comfort with several data breach scenarios in which different entities (e.g., security practitioners, researchers,

journalists) were investigating the exposed data. Respondents were largely comfortable with applications that used exposed data (e.g., for proactive password resetting) if the application provided a direct and tangible security benefit. Zou et al. similarly studied perceptions and actions after data breaches, specifically pertaining to the Equifax data breach [235]. They found that although many participants were aware of the breach, few knew if they were affected and even fewer took mitigating measures afterwards. The optimism of participants in believing that they were unlikely to be victims as well as a general tendency to delay action until harm has occurred were influential factors contributing to a lack of response. However, the source of advice about steps to take after the breach also affected participants' willingness to take action. Recent work by Mayer et al. further found that when people were asked about their awareness of and responses to data breaches that actually affected them, participants were aware of only 26% of the breaches they were asked about [154]. Redmiles also studied action in response to one-off security incidents individuals may experience in a social network, such as suspicious login incidents [188]. They found that when participants were notified, only a third of participants took protective measures and that the notifications were not as effective when the affected participants were uncertain about what caused the incident.

Researchers have also studied how consumer spending habits are affected after a breach announcement. Janakiraman et al. empirically examined how customers of a breached multi-channel retailer (a retailer that sells products across different platforms such as online or brick and mortar stores [1]) altered their shopping habits. They found that customers whose data was breached significantly reduced their spending by 32.45% [32]. Ablon et al. similarly measured how people's business with a company continued after a breach. While they found that customer attrition was 11%, consumers seemed to appreciate when companies responded and appeared to take responsibility for the breaches they suffered.

Previous work has also studied people's general awareness of breaches and how breach information arrives at the attention of internet users. Das et al. studied the types of security and privacy news events that people encounter [79]. They identified that financial data breaches, corporate personal data breaches, high sensitivity systems breached, and politicized/activist cybersecurity were the types of events people found the most salient. They further found that online news articles accounted for 70% of participants' news sources and that social media accounted for almost a third. The aforementioned study by Ablon et al. also found that when a company suffered a data breach, almost half of its customers learned about breaches from a source other than the affected company [32]. Mayer et al. similarly found that participants heard about a breach equally often from the breached organization and from third-party notification services [154]. In contrast to the study by Das et al., however, they found news media only accounted for 12% of participants' breach information sources.

Users can also be alerted about breaches that affect them by dedicated services such as HaveIBeenPwned [17], LifeLock [16], and Enzoic [13]. Additionally, password managers such as LastPass [15] and the password manager built into Firefox [14] alert users if their logins are found in data breaches. Thomas et al. recently created a privacy-preserving protocol to protect against credential-stuffing attacks (i.e., when an attacker uses lists of breached usernames and passwords to gain access on a large scale to several other websites) that notifies users about credential breaches without revealing the actual credentials [216]. They found that only 26% of these notifications caused participants to create passwords that were at least as strong as their

previous ones [216].

We draw inspiration from previous work that examines how people come across incident information and suggests that the source of information is important for influencing users to take protective action. Our work is also motivated by the low self-reported awareness of breaches found in previous studies. In particular, Chapter 3 focuses on incident information on web pages and we base our analyses on participants' real browsing behavior.

2.2 Security and privacy information dissemination and notifications

Related work has studied the mechanisms and sources by which people learn about security and privacy and has often found that informal stories and online media are common and suitable channels for this task. Rader et al. examined where non-experts in security get their information from and discovered that most people learned lessons from stories about security incidents informally told by friends or family [185]. Das et al. further examined what drove people to want to share security and privacy advice or experiences [76]. They observed that people were driven to have such conversations by a desire to warn or protect others from threats they have experienced or to try to gather information about others' experiences.

Prior work by has also studied themes in security and privacy advice across different news sources. Rader et al. studied these themes in three different sources—news articles, web pages with security advice, and informal stories from family or peers—and found that each source presented information in a uniquely constructive way [184]. Information from peers often focused more on who conducts attacks. Expert advice usually focused on how attacks are carried out, and news sources focused on the consequences of attacks. The researchers discuss that these differences in news sources may prevent users from understanding the implications of attacks sufficiently or from identifying adequate protective measures to take. Therefore, it appears that the source of information may have an impact on people's security behavior. Additionally, different demographic groups may be more likely to be exposed to different information sources and thus may receive inconsistently constructive advice. Specifically, recent qualitative work by Nicholson et al. found that older adults tend to rely less on internet sources and more on social resources such as advice from friends and family [168].

In support of the above work, other research has found that the sources of security and privacy advice are important factors for people's digital security habits (as also discussed in Section 2.1). Redmiles et al. examined information sources from which people received security advice and found that people's trustworthiness of the advice source was an important factor in whether they heeded the advice or not. The researchers also inferred that negative security incidents, when portrayed as part of a fictional narrative with relatable characters, could be effective teaching tools. Redmiles et al. later studied security advice sources and behavior on a larger census-representative scale and found that the source of information was still important. However, they also discovered that there is a "digital divide" in how security advice is distributed and received. In particular, people with higher socioeconomic status had different advice sources than those with fewer resources [189].

Prior work has also studied the potential of social avenues online for disseminating security information. For example, Dunphy et al. studied the potential of social media as a resource for understanding security experiences [85]. They assessed the potential of Twitter in providing insights into security practices related to passwords based on discussions of personal experiences. They found that people use Twitter to complain or share opinions regarding security incidents, although these discussions of security can also be carried out in non-constructive ways (e.g., playfully discussing approaches to bypassing password mechanisms). Researchers have also looked at the ways in which presenting people with security information in a connected network of computer users may help convince them to adopt good security practices. One such work by DiGioia et al. proposed interfaces for filesystems that show people how others implement security features in their own filesystems [81]. Das et al. similarly studied social influence within Facebook on a larger scale by showing 50,000 users in a network what security features their friends were using [77]. They found that by showing users the number of their friends who adopted a security-enhancing feature, users were more inclined to explore and subsequently adopt those features themselves.

Prior research has also studied security information in the form of general vulnerabilities or compromise notifications. For example, recent work surveyed people's reactions to notifications of password compromise. When advised or required to change their passwords by the notification, less than a third of respondents reported any intention to change their passwords [106]. Work studying notifications of system vulnerabilities found that presenting the information in a certain way (e.g., by presenting detailed information in the message itself) was more likely to result in system maintainers applying software patches; however, the majority of the organizations they alerted failed to respond to the notification [144]. Similarly, recent work has studied how to effectively notify owners of sites with misconfigured HTTPS configurations. They found that though notifications have a moderate impact on the likelihood of taking remediation action, it was not sufficient to incite action in the majority of owners they alerted and that public outreach and campaigns are a more promising approach [231].

This thesis is motivated by the findings that web-based media including social media are useful and suitable mechanisms for spreading computer security and privacy information. We are additionally motivated by the finding that the awareness of the general public is important for people to keep themselves secure. For example, in Chapters 3 and 4, we study online awareness of security incidents and actions after breaches regardless of whether people were impacted and notified directly. Furthermore, in Chapters 5 and 6, we focus on interactions with security or privacy content on social networks to increase the security awareness of the general public.

2.3 Social influence in security and privacy

Prior work has repeatedly shown that social influence both through person-to-person communication and social networks can play a role in the adoption of security-enhancing habits.

In Section 2.2, we described prior work in which people received security information through social connections. After receiving this information, people were inclined to adopt certain security behaviors, thus affected by social influence. For example, Rader et al. found that informal stories told by friends and family members were significant sources of information and encour-

aged people to advise others about security [185]. Das et al. found that people were inclined to change their security behavior as a result of indirect group or peer pressure [76]. DiGioia et al. further proposed that social navigation (a form of social computing in which "movement from one item to another is provoked as an artifact of the activity of another or a group of others") can be an effective model for building usably secure systems [81]. They demonstrated this phenomenon through an example involving users of a filesystem in which users are shown what security features other users of the filesystem are using.

Social influence has also been found to play a role in security and privacy decision making. Emami et al. studied what influenced people to make privacy-enhancing decisions when given the option to allow or deny Internet of Things (IoT) data collection [88]. When making security decisions, participants were more influenced to take defensive measures if they were told several of their friends made that decision [88]. In particular, they found that participants were most influenced by friends when those friends denied data collection but were influenced by experts when the experts allowed data collection to occur.

Prior work has also studied social influence in adopting healthy security practices. These influences are implemented via the structures of a social network. Das et al. implemented "social announcements" within Facebook to study their effect on social network users' security behavior [77]. These "social announcements" told users how many people in their network were using certain security features. The researchers experimented with showing users in the network seven announcements containing social proof (evidence of social connections adopting a behavior) and one without social proof. They found that 37% more users explored security features when exposed to the social announcements when compared to adoption after the non-social announcement. The same researchers further studied how information about security features can be diffused through a social network and be adopted by users [78]. Though they found that users in the network increased their adoption of security features, this increase depended on several factors. These factors were the perceived visibility of the feature, its adoption rate by a user's friends in the network, and the number of distinct social groups within the network surrounding the users that shared the information.

Perceived benefit or visibility has been found to be an important factor in people changing their security behavior despite social pressure [43, 118, 235]. Social networks have also served as effective platforms for people to engage in discussions about security and privacy experiences, advice, and complaints [85]. Overall, prior work has shown that social influence is an effective tool for encouraging adoption of certain security-enhancing behaviors and decisions. However, such social influence has also been seen to have a negative effect on people's security practices. Gaw et al. found that people viewed security-conscious internet users as paranoid or as exhibiting undesirable behavior [102]. Therefore, these people were less inclined to implement practices taken up by security-aware people (e.g., using encrypted mail).

Inspired by existing work, parts of this thesis (Chapters 5 and 6) measure—using empirical, real-world data—how social media may be used as a platform to talk about security and privacy.

2.4 Social media as an educational tool

In the work described in Section 2.3, social network interventions were used to influence other users to adopt security features. While this could be considered educational concerning other people's behaviors, social media has also been successfully used as a purely educational tool for spreading awareness and education in areas not related to security and privacy across various scientific domains.

Hamid et al. discuss how social media can be used to disseminate environmental sustainability awareness but that it is not currently effectively used for that purpose [113]. Kaur et al. further examined factors in social media that were correlated with people's environmental issue awareness. Fundamentally, they found that people often use social media for the purpose of absorbing new information, making social media particularly promising as an educational tool [129]. They also found that factors such as the persuasive power and people's perceived trust of social media were important factors in changing their attitudes about environmental issues.

In a similar vein, researchers have also studied education through messaging social networks. For example, Wu et al. analyzed how WeChat can popularize content about wildlife conservation measures and concerns [227]. They found that higher readership counts were correlated with more pictures and fewer words. They discuss that messaging social network applications such as WeChat can be used to increase general awareness of wildlife conservation measures and to reduce misunderstandings of popular topics such as policy changes and recommendations by scientific experts.

Other work has highlighted that scientific information should be published in spaces other than traditional scientific journals or legacy media. Mueller et al. proposed spreading scientific information through social media so that the information is more obtainable outside the scientific community [164]. The researchers reported that in addition to simply using social media to spread awareness, the way the posters of scientific information used social media were determining factors in spreading issue awareness. For example, the heterogeneity of the network (the degree of demographic variation in a network) in which they were disseminating information was important for increasing awareness.

Social media has also been extensively used in other areas. For example, social media has been used to launch mental health awareness campaigns and to increase geographic awareness across the world. Saha et al. examined how mental health is discussed on the social network Twitter [195]. They found that there was an abundance of discourse surrounding this topic and that often the content was inspirational, provided clinical tips, or contained useful resources. Ye et al. discuss the promise of social media in increasing the digital footprint across the world and how this can contribute to understanding human dynamics in terms of how people communicate with each other [229].

It is clear from existing research that social media harnesses a great ability to share accurate information on a variety of topics. The work in this space motivates our study of how often and how security and privacy is discussed on social media in the wild. In particular, this work supports our goal of understanding how to improve the dissemination and presentation of security and privacy advice on social media such that it can be used as an educational tool to reach more people.

2.5 Information spread in social networks

Social networks are useful educational tools because of the important role they play in information diffusion [46, 109]. Bakshy et al. showed this by examining how the social structures in a social network contribute to the propagation of information [46]. They randomized exposure to signals about what friends in a network are sharing and found that those exposed to these signals were more likely to spread information themselves. They also studied how weak and strong ties (when users interact infrequently or frequently with each other) within a network contributed to the propagation of information. They found that stronger ties between users in a network were more influential but that receiving information from weaker ties resulted in more information propagation. Given the important role of social networks in information diffusion, researchers have found that social media has the potential for information spread and popularizing content related to various topic domains in social networks such as Facebook, Twitter, Instagram, and Reddit.

Mazloom et al. analyzed what makes certain posts popular and others ignored in social networks [155]. They specifically studied the popularity of brand-related content posted on social media in relation to several engagement parameters. These engagement parameters encompassed several features including whether an indication of the brand logo was present in the post, the number of detected faces in the image, the presence of a product image in the image, the sentiment of the content, image aesthetics, and the number of followers of the poster (as a measure of the poster's popularity). In Chapter 6, we analyze properties of posts on Reddit based on features computed by Mazloom et al. (i.e., we analyze the sentiment of the post content and the popularity of the poster).

Laor et al. examined the same problem but specifically concerning posts about radio programs and how to popularize these radio programs through Facebook interactions [139]. They studied several properties of posts including the type of media present in the post, the number of times it was shared, and the number of likes and comments. They also examined several language-related properties such as whether the post had emotional or rational language, formal or informal language, or whether it presented information. One important finding was that incorporating different types of media (e.g., videos or images) into a post can generate more popularity. Chapter 6 again computes and analyzes features similar to those used by Laor et al. In particular, we analyze whether the Reddit posts we analyze have links or images present and whether the posts were crossposted on Reddit (i.e., shared before). Inspired by the analysis of tones in the text about radio programs, we also determine the presence of different tones (e.g., anger, joy) in the post content in relation to the post's spread or popularity.

Hong et al. predicted the popularity of messages in Twitter without considering content pertaining to a specific topic [120]. They studied over 10 million posts and predicted post popularity with high accuracy using features such as the topics of the content computed through topic models, the popularity of the posts made by the user, the number of posts made by the user, and whether the post had been retweeted before. As with the aforementioned studies, Chapter 6 uses features similar to work by Hong et al. in particular, the popularity of Reddit posts made by the poster, the number of Reddit posts made by the poster, and whether the post had been crossposted.

Other research has studied the popularity of image posts by studying features of the posted

images. Peng et al. examined what makes image posts of politicians more popular [177] and Zailskaitė-Jakštė et al. studied how the colors used in posts are related to their popularity [230].

The goal of this thesis is motivated by the effectiveness of social media as a vehicle for information sharing as evidenced by prior work. Furthermore, Chapter 6 analyzes properties of posts related to their spread using several features inspired by this prior work.

2.6 Measuring security behavior

There has been an abundance of prior work that has measured and studied security behavior for a variety of research questions. Security behaviors have been measured in a couple different ways and have spanned behaviors from updating security settings and clicking on unsafe links to password reuse habits. We discuss these studies, behaviors, and measurement approaches below.

Forget et al. examined whether there was a relationship between people's self-reported engagement with computer security and maintenance and their actual security behavior [98]. They interviewed participants about their engagement and measured and analyzed several security behaviors based on participants' computer usage logs, finding that higher engagement does not necessarily mean better security behavior. Canfield et al. studied whether security behaviors of individuals were correlated with their ability to detect phishing attacks [67] but found no evidence that detection abilities were related to healthy or detrimental security behaviors. In both studies, the security behaviors measured for each participant included how often they updated their computer systems, what security settings they had on their systems, and whether their systems were infected with malware. Wash et al. measured security behaviors for the purpose of examining whether users can self-report their security behavior accurately [224]. They examined this by surveying participants to obtain the self-reported behaviors and collecting their system logs to identify correlations. The researchers identified only few correlations between self-reported and real behaviors, indicating that people were often not able to accurately self-report their security behavior. They additionally observed that participants were able to self-report behaviors more accurately when the associated action was done proactively or in response to a prompt (e.g., installing an ad-blocking extension) in contrast to behaviors that were less visible (e.g., automatic Chrome updates). Wash et al. also measured how often participants updated their computer systems in addition to other behaviors including the presence of third-party applications on their systems and whether they installed security-enhancing extensions.

Sharif et al. have studied how to predict the occurrence of a specific negative security behavior: clicking on unsafe links [200] (also measured by Canfield et al.). In particular, they used data about user's past browsing visits to predict the likelihood of the next link a user clicks on being malicious according to Google Safe Browsing [9]. Habib et al. have again studied a specific type of security behavior: using private browsing mode in browsers [111]. They analyzed what people used private browsing for and whether it mitigated participants' browsing-specific concerns, finding it to be the case. They also found that participants overestimated the privacy that private browsing mode afforded them in terms of being tracked. In this thesis, Chapter 3 focuses on security behaviors related to web usage such as visiting websites related to security incidents. Meanwhile, Chapter 5 measures security behaviors similar to prior work involving behaviors about protecting computer systems and filesystems and security settings.

Chapter 4 measures behaviors specific to passwords. One major behavior we study is password reuse. Several large-scale password studies have shown that password reuse is rampant. Florencio et al. were the first to conduct a large-scale password analysis of its kind in 2007 in which they analyzed the password data of half a million users over a three-month period [96]. In addition to characterizing the passwords entered and their strengths, they found that people reused their passwords widely, with each client whose passwords they analyzed sharing each of their passwords with 3.9 other sites on average. Pearman et al. later analyzed real-world data about passwords entered by almost 200 participants across all domains they visited (over an average of 147 days each) to understand more current patterns in password security [174]. Das et al. similarly studied password reuse across several hundred thousand leaked passwords from eleven websites [75]. Both studies found that on average people reused over half their passwords on other sites.

One approach internet users can use to maintain strong and unique passwords is using password managers. However, password managers are only used by a small fraction of the population [38]. Furthermore, existing studies by Stobert et al., Alkaldi et al., and Pearman et al. have found that users are often not aware of the purpose of password managers and struggle to set them up [38, 175, 211], contributing to password reuse [212]. Other password-related studies analyzed behaviors in response to password compromise. For instance, as described in Section 2.2, Golla et al. surveyed people's reactions to notifications that their password was compromised or was being reused on other sites and observed scarce action in response [106]. Thomas et al. notified people about real occurrences of their passwords being breached and found that in only a fourth of the cases people changed their passwords to be stronger [216].

Two main approaches have been used to measure the security behavior described above: collecting self-reported data through surveys, interviews, or controlled experiments (e.g., people's behaviors when exposed to internet attacks [171], password updating habits [112], and willingness to take remediation measures after a breach [126, 235]) and instrumenting users' computers to observe security behaviors (e.g., measuring password reuse [75, 96, 174] or the presence of malware on people's computers [98]). Since self-reported data can be prone to biases and may not be representative of the reality of peoples' security and privacy [42, 92, 111, 114, 194, 218, 224], we focus on empirical measurement of actual behavior for much of this thesis.

Previous work that extracted security behaviors from real data has collected data in multiple ways. Sharif et al. partnered with an internet service provider that recorded all HTTP traffic of consenting participants [200]. Wash et al. asked study participants to install a tool that collected their system logs [224]. Researchers have measured password-related behaviors in a variety of ways (e.g., by asking participants to install password-logging tools [96, 224] and analyzing breached passwords from publicly posted lists [31, 75] or privately collected datasets [156]).

Much of this thesis (Chapters 3, 4, and 5) leverages a data-collection infrastructure called the Security Behavior Observatory (SBO) (first described in Chapter 3), which captures detailed, real-world behavioral data of home computer users that was collected through instrumenting participants' operating systems and browsers [97, 98]. The SBO has been used to study password reuse [174], private browsing [111], and people's maintenance of their systems for security [67, 98].

2.7 Information campaigns in public health

Section 2.2 described work that emphasized the need for the general public to have security awareness for people to take security-enhancing actions. Information campaigns have been useful in promoting awareness in other domains. One of the most prominent domains in which these campaigns have been successful is public health. Two well-known examples are campaigns for anti-smoking and HIV-prevention. Health-focused campaigns, in general, have been conducted over centuries and have been executed on multiple platforms. In this section, we describe research about what makes these public health campaigns effective. Research in this space often found that making fear appeals, tailoring content to different groups of people, and using different forms of media increased the likelihood of these campaigns' success. We also discuss how findings from studies about these health-related campaigns can inform the design of cybersecurity campaigns.

2.7.1 Fear appeals

Several studies and position papers have attributed the success of health campaigns to fear appeals [71, 91, 108, 214, 226]. Fear appeals are persuasive messages that arouse fear in the audience. For example, in health awareness campaigns, fear appeals attempt to scare people about the consequences of bad health. Fairchild et al. presented a case study on how fear appeals have been successful in three different public health campaigns in New York City: tobacco use, obesity, and HIV infection [91]. They found that depicting the damage in a way that elicits disgust from a viewer was effective and the city saw marked declines in tobacco usage as a result. Strong fear appeals related to tobacco were not directly applicable to obesity and HIV infections as they had the potential to create controversy or socially stigmatize different groups by showing examples of groups of people more likely to be affected. Fear appeals were also often accompanied by other regulations (e.g., increasing the price of cigarettes or banning their use in certain areas), which made their effects stronger. Overall, this study shows that fear appeals that do not target or profile specific demographics may be effective in spreading awareness and comprehension and that eliciting an emotional response from viewers is key to a successful campaign.

In the realm of HIV and AIDS infections, Green et al. analyzed how fear appeals affected HIV infection rates and how the effectiveness differed between American and African health campaigns [108]. They discuss that American AIDS professionals have rejected fear appeals. However, Uganda implemented fear appeals and subsequently saw large declines in infection rates until they adopted the American style of messaging without invoking fear. The researchers' findings imply that fear appeals are effective but may not be socially accepted by everyone including academics and therefore, may face resistance in practice.

Witte et al. evaluated the general effectiveness of fear appeals by conducting a meta-analysis of studies about fear appeal campaigns [226]. They emphasize that fear appeals' success is often measured across three broad dimensions: perceived fear, perceived threat, and perceived efficacy. They also discuss the differences between the effectiveness of weak and strong fear appeals. Weak appeals contained an efficacy message which focused on reinforcing that people have the ability to implement changes (self-efficacy) or that certain behaviors will produce results (response-efficacy) [214], while strong appeals contained an efficacy message. One of their key

findings across related work was that strong appeals were shown to have a higher impact on fear, perceived threat, and perceived efficacy. Based on the findings of their meta-analysis, they provided several recommendations for practitioners to promote public health. These recommendations include referencing the severity of the threat clearly, explaining the likelihood of being vulnerable to the threat, emphasizing health issues that are most likely to be relevant, and including strong efficacy messages.

Cho et al. examined the effectiveness of fear appeals on people at different stages of change regarding problematic or risky behaviors [71, 183]. The stage of behavioral change relevant to this study is "precontemplation." People in this stage have no intention to stop a behavior in the foreseeable future. The researchers exposed people at various stages of change to high-threat messages and low-threat messages (i.e., they only provide general facts about health issues) and found that people in the precontemplation stage of change were the most likely to exhibit effects unintended by the messaging and to react to messaging in a maladaptive and counterproductive manner. This finding suggests that health education messaging may not be sufficiently effective for the people who need it most. People in the precontemplation stage are the least aware of risky behaviors and their consequences and hence require education.

Findings from the above studies can directly inform how to approach the design of cybersecurity campaigns that use fear appeals. Messages about security could contain efficacy messages about the ease of taking protective actions and could clearly outline the consequences of threats so that the messages have the most impact. Additionally, governmental regulations alongside security awareness would significantly help the improvement of public security. For example, current data breach laws do not require companies to force their users to take security-enhancing actions. By mandating security-enhancing actions, users would be required to protect their security. Finally, messaging could be altered to target different populations at different stages of change. People who are exposed to security behavior advice and plan to take action are likely to respond differently to fear appeals than those who have never been exposed to such advice.

2.7.2 Factors contributing to successful campaigns

Basu et al. examined how branding can play a role in successful public health campaigns [47]. Branding in this case refers to assigning an identity to a campaign, for example, treating public health as a "product." Branding allows people to engage in a relationship with the campaign rather than simply consume or be educated by its information. Branding can help with capturing consumer attention, focus on educating as well as forging a relationship with consumers, help with ensuring long-term commitment to the cause, and help connect with people based on their lifestyle and values. One component of branding is presenting a direct benefit to the consumers in addition to education. Another larger component is ensuring that the campaign is backed up by an organization that is willing to sustain the campaign and use all communication tools at their disposal as advertisers do.

Rofail et al. studied the successes and failures in public health campaigns, specifically related to promoting the consumption of folic acid by women before and during pregnancy to prevent neural tube defects [193]. They identified several factors that contributed to the success and failures of these campaigns. One important factor was that the campaigns' reach was incomplete and the media they used to promote the campaign did not reach all of the intended population

nor may have been understood by all. People who had prior knowledge of the promoted health concept were more likely to perceive the benefits of folic acid, which influenced consumption. Other demographics such as socio-economic status and age were also important factors in the uptake of folic acid. On the note of demographics, Petersen et al. discussed how geodemographics can be useful for targeting different geographical locations effectively [181]. In the space of health campaigns, they find that by targeting neighborhoods with fewer resources and in more need of health education, they can more effectively provide public health education.

Much of the findings from the work described here can be relevant to cybersecurity campaigns. For example, focusing on providing a brand identity to cybersecurity campaigns such that people remember it and involving the public in the design of campaigns can be effective. The last two studies suggest that campaigns should be designed differently for different target populations, which is in line with findings from work described in Section 2.7.1. For example, campaigns could be made up of multiple versions where each targets a group. These groups could be based on whether people in the group have prior knowledge about the cybersecurity concepts or whether they're novices in addition to age and other demographics.

2.7.3 Type of media used in campaigns

Public health campaigns have used a variety of media to carry out their messages. Two of the most prevalent forms of media are mass media (e.g., television, magazines, newspaper) and interactive media (e.g., internet, telephone).

Randolph et al. examined the success of mass media in promoting public health and distilled recommendations for future campaigns [186]. In addition to discussing success factors in line with the above related work, the paper highlights how mass media (e.g., billboard, television, and radio advertisements) are useful for increasing the amount of information about public health. The study found that campaigns that were most successful bought prime advertising time and space on media such as television or radio. However, often organizations that organize public health campaigns do not have sufficient funds to purchase this type of media space and often rely on these media channels to donate their time for public service announcements (PSAs), which may often not correspond to prime time slots. To increase their campaigns' effectiveness, these organizations have supplemented their campaigns with billboard ads, printed ads, and labeled promotional items such as t-shirts and other worn items.

Hughes-Hallett et al. analyzed a specific mass media campaign in England called "Be clear on cancer: Blood in pee" which was meant to encourage people to visit their doctor to get checked for bladder or kidney cancer [123]. The campaign was conducted through mass media and saw a promising increase in cancer referrals. However, the results were short-lived and the rates of referrals reduced after a while which emphasized the need for an approach with sustaining effects.

Snyder et al. synthesized multiple meta-analyses of mediated (i.e., using mass media, interactive media, and little media) public health campaigns' success and identified additional factors that played a role in this success [205]. They emphasized that messages that presented new information rather than repackaging old information were more successful. They also discuss that tailoring messages based on feedback, source similarity (the degree to which a message source and the recipient are alike with respect to certain attributes), and clearly describing the rewards

for successful behavior played roles in the success of messages in public health campaigns.

Finally, Ahmed et al. analyzed how social media has been used for interactive media public health campaigns, specifically by studying its effectiveness related to World Autism Awareness Day [36]. Their study found evidence that the volume of content about autism awareness was increased on Twitter in addition to the positive sentiment of this content. However, they state that further work is needed to assess their effectiveness on individuals' behavior. In particular, these studies point to the effectiveness of conducting information campaigns using various forms of media, which can be applied to cybersecurity campaigns. Social media seem to be wide-reaching and effective channels for putting out more information about cybersecurity but other media such as traditional news and print media may be effective supplemental media through which to carry out a campaign. As with prior studies, cybersecurity campaigns would also benefit from doing targeted research on effective messages for different groups and which messages should be shared through which channels.

Chapter 3

What breach? Measuring awareness and action after security incidents by studying real-world browsing behavior

3.1 Introduction

With the rise of security incidents and data breaches, security awareness is crucial for people to have the tools and know-how for keeping their computers and online data safe [115]. High-profile incidents and breaches in the past decade such as WannaCry, Heartbleed, Petya, and NotPetya have compromised over 300,000 systems worldwide [99, 122, 187]. The data compromised has ranged from passwords, names, and email addresses to credit card and social security numbers. People affected by these incidents typically need to become aware of them before they can take remedial action. More generally, awareness of the extent and effects of security incidents increases the adoption of better security practices [115, 207].

To this end, research about awareness of security incidents (completed using surveys and interviews) has found that people learn about breaches from a variety of sources and that some breaches are more likely to be discussed than others [79]. One survey found that almost half of respondents heard about a breach from a source other than the breached company [32]. People's reported willingness to take action was shown to be correlated with the source of information [235] and if they perceived a tangible security benefit [126]. Overall, these studies provide an important step towards understanding how people learn about and react to incidents.

In this chapter, we take a significant step toward a more detailed understanding of how people learn about and take action after incidents, specifically through online browsing. For six national-scale security incidents of potentially varying relevance to people, we use *longitudinal, real-world browsing data* to examine to what extent people may become aware of these incidents and the subsequent actions they may take (e.g., to learn more about the incident or generally about security). With the underlying goal of improving the spread of incident information through online media, we specifically study these problems in the context of *online browsing* without considering other channels by which this information may be shared. Our dataset was collected from the home computers of 303 participants between October 2014 and August 2018

and includes all URLs visited and passwords used to log onto online services from participants' home computers. As users could also read about security incidents on devices from which we do not collect data, we conducted a follow-up survey of 109 participants to confirm our results (Section 3.5).

We explore two main topics in this chapter: First, we examine *how often* participants read about incidents on the web and whether the likelihood of reading about incidents varies by demographics, browsing habits, or self-reported security behaviors. Second, we seek to understand *how participants came to read* about incidents, *how they reacted* to reading about them, and how different ways of finding out about incidents affect the action they take. For example, we examined whether the type of web content (e.g., news vs. social media) on which we first observed participants reading about incidents affected whether they took constructive action, such as further investigating an incident.

We found that only 16% of the 303 participants visited an incident-related web page about any of six major security incidents between 2014 and 2017. For example, only 15 of 59 likely Equifax credit-report holders read about the breach online in our dataset. These number remain alarmingly low even after accounting for mobile browsing not captured by our dataset. Overall, we found that older and more tech-savvy participants as well as those who were more proactively aware about security [86] were more likely to read about security incidents on the internet.

Of the participants whom we observed reading about an incident, 73% subsequently visited additional web pages with information about the incident or about security and privacy in general. Reasonably, the higher the severity of the data compromised, the more likely participants were to visit related web pages. Participants' likelihood of taking action was higher if the content through which they found out about the incident had a positive sentiment; no other property of the incident-related content seemed to be associated with taking action, even though our power analyses showed we had a sufficient sample size to show medium-sized effects.

Overall, our results suggest remarkably low awareness of, or inclination to follow up on, security incidents. The implications of these results are two-fold: first, our results suggest that people may not sufficiently engage with information about security incidents and that for those who do engage, the presentation of the information can play a role in inducing action. Second, the extremely low rates of engagement may also indicate that increasing awareness is not the most effective avenue for keeping users safe. Further research should study other approaches to improving user security while systems that people use should take steps to keep their users safe without requiring them to maintain their own security.

We first describe our dataset (Section 3.2). We then describe the methodology for and results of investigating *how many* and *which* people read about incidents (Section 3.3) and *how* people learn about incidents online and how this affects their actions (Section 3.4). We also describe a follow-up study that substantiates our results using self-reported data (Section 3.5). Finally, we discuss the limitations and implications of our work (Sections 3.6 and 3.7).

3.2 Data collection and dataset

3.2.1 Data collection

We obtained data collected as part of the Security Behavior Observatory (SBO) project, a longitudinal study of the security behaviors of Windows computer users [97] from October 2014 to July 2019. Data collected by the SBO includes information about system configuration, system events, operating system updates, installed software, and browser-related data such as browsing history, settings, and the presence of browser extensions. To collect this information, participants' home computers were instrumented with software that collects data via system-level processes and browser extensions. Data related to passwords entered into web pages was collected starting January 2017 and only in the Google Chrome and Mozilla Firefox browsers. Participants were compensated by the SBO project with $30 for enrolling and an additional $10 each month they were enrolled in the study.

The SBO and its use for our work were approved by the ethics review board at Carnegie Mellon University. Data collected by the SBO has been used to study, for example, private-browsing habits [111], people's ability to detect phishing attacks [67], people's maintenance of their systems for security [67, 98], and password reuse habits [52, 174]. The SBO dataset contains data about a broad range of people across multiple demographics. (described in Section 3.3.2).

Our study is based on longitudinal datasets that were collected by the browser extensions. In particular, we use the following two sets of data.

Browsing history: The browsing data we analyze spans a subset of the whole SBO dataset from October 2014 to July 2018, encompassing 303 participants who were active in the study at the time of when at least one of several security incidents was publicly announced (see Section 3.3). Participants enrolled in the SBO study on different dates and for different durations. The dataset covers participants' browsing using Google Chrome, Mozilla Firefox. The average duration for which the 303 participants were enrolled was 505 days. This dataset includes information about every URL visited in the web browser, along with page titles and timestamps.

Password data: This dataset spans from January 2017 to August 2018 and includes 233 of the 303 participants. The data includes information about every entry made into a password field in a web page, as determined by a browser extension, including: a salted one-way hash of the password and the URL of the form in which the password was submitted. We filtered this dataset to exclude passwords used during failed login attempts or entered by a user other than the main computer user as described below.

The browsing data was retrieved from participants' main computers. We assessed the accuracy of our results in the context of participants' overall browsing across multiple devices through a follow-up study of 109 SBO participants (see Section 3.5) which appeared to support our main findings. We further discuss the limitations of this dataset in Section 3.6.

Filtering passwords

The SBO browser extension collected every entry made into an HTML password field. This captured both the entry of correct passwords as well as attempted logins that failed because an

incorrect password was entered. The recorded passwords may occasionally have been entered by other users on the participant's computer. A single participant could also have multiple accounts and passwords on the same domain.

We needed to eliminate any failed login attempts from this dataset and any passwords that did not belong to the participant's main account. We combined collected password entries across multiple browsers on each participant's machine and extracted the "correct" passwords for a participant by applying heuristics inspired by Pearman et al. [174] and Wash et al. [223], as follows.

We first compiled all password entries on each domain in chronological order. For each domain, starting from the participant's first password entry on that domain in our dataset, we divided the entries into clusters where the differences between timestamps within one cluster was less than 15 minutes. We considered the last entry in this ordinal cluster to be the "correct" password of a cluster, i.e., signaling that the user probably logged in correctly and will not attempt to log into that domain again for a while. We then further filtered these clusters to remove occasional non-participant logins and each participant's secondary accounts, if they had multiple accounts. If the "correct" password of a cluster reappeared in a later cluster, we assumed that the passwords entered between the two occurrences could have been due to intermittent logins either not by the main user or for less-used accounts. We only did not consider the entries to be due to intermittent logins when any of the passwords entered between the two occurrences occurred more frequently than the re-appearing password for the participant or if the password was submitted over more days in the case of frequency ties. We do not consider the re-occurrence of an older password to mean the participant changed their password back to an old password since domains typically do not allow users to change their password to a previously used password.

This process left us with a set of "correct" password entries, which is the final dataset we use for password-related analyses.

3.3 Who reads about security incidents

We examine how many and which people visit security-incident-related web pages and what factors are associated with their likelihood of doing so. We focus on selected security incidents (Section 3.3.1) and model participants by their demographics and technical backgrounds, self-reported security intentions, and internet browsing behavior (Section 3.3.1). We report on the relationship between these features and the likelihood that participants visited pages related to security incidents (Section 3.3.2).

3.3.1 Methodology

Identifying who read about security incidents

We examined six security incidents that occurred between 2013 and 2017 [40, 140, 232] for which we expect most people to have read about *at least one*. We selected these incidents because they (1) were large-scale incidents (not affecting only a local population), (2) spanned a variety of incident types (from personal financial data losses and company document leaks to cy-

ber attacks on home computers), (3) are well-known, and (4) were represented in our browsing history dataset. We selected well-known incidents because people's awareness and engagement are likely stronger and easier to observe for such incidents. In particular, we studied:

- **Equifax breach:** September 2017 breach of the credit reporting site that compromised the personal information of almost 150 million customers [60].
- **Uber breach:** Late 2016 breach that compromised the personal information of 57 million Uber users [141].
- **Ashley Madison breach:** Data breach on the affair-centric dating site in July 2015 and compromised around 33 million users' private information [147].
- **Panama Papers:** April 2016 breach of 11.5 million files from the database of the world's fourth largest offshore law firm, Mossack Fonseca [48, 116].
- **WannaCry:** Ransomware attack in May 2015 that initially affected over 70,000 computers in 99 countries [49, 187].
- **Yahoo! breaches:** Two breaches: one in late 2014 affecting over 500 million user accounts and another in 2013 affecting over one billion user accounts [104, 179]. It was later revealed that all user accounts were compromised [142].

Each incident we studied may have been relevant to users in different ways. They could have been affected by it, they could be users of the compromised service and may want to be more cautious in the future, or they could learn about general security and privacy dangers. For example, although Panama Papers may not be directly relevant to most users, we included it because awareness about it could indirectly encourage users to be cautious about their own private records (e.g., medical records) and maybe be selective in trusting institutions with their data.

To study who reads about these incidents, we studied the 303 SBO participants who were active in the study before the incident became public and for three months after.

For each incident, we identified participants who visited an incident-related page (henceforth, we may call this *reading about an incident*). This page visit could have been the first exposure to the incident or an attempt to learn more about the incident online. Since we seek to study how often people actually signal their intent to learn more about an incident rather than simply "hearing about it," we did not consider a participant to have read about an incident if they may have seen it on some web page (e.g., social media) but did not click on the article.

To determine whether a participant read about an incident, we performed a keyword search over the URLs and titles of all the pages in their browsing history. For each incident, we manually selected a set of keywords that we believed would identify web pages that focus on that incident. For example, we searched for various combinations of "Yahoo" and one of the following: "compromise", "attack", "breach", and "hack". To confirm that our keyword lists were inclusive enough and that our identification process was robust, we also performed multiple Google searches using a variety of search terms to find web pages about the incidents and then confirmed that each of the top 100 Google search results about each incident would be identified by our keyword lists. We then manually verified that each page visit that matched a keyword actually corresponded to a page about the incident. For example, a page on `yahoo.com` with the path containing the word "hack" referring to a page about life hacks would not be considered

an incident-related page.

Equifax and Yahoo! users To provide further context for our observations of how many participants read about an incident, we observed, for people who were likely to have been *affected* by an incident, how many of them read about the incident as part of our analysis. Equifax and Yahoo! are the two breaches for which we were able to relatively accurately estimate how many participants were *actually affected* by examining whether they logged in to certain web sites. In both cases, the number of affected people was all or almost all users or consumers [4, 60, 142].

We determined which participants were likely to have had an Equifax credit report by observing who had entered a password on `equifax.com` or on one of the popular credit-report sites that report Equifax scores [68] (`identityforce.com`, `identityguard.com`, `annualcreditreport.com`, `creditsesame.com`, `creditkarma.com`, and `quizzle.com`) before Sep. 7, 2017, when the breach became public. While most Americans were likely to have been affected [4, 60] regardless of whether they had an account with a credit-reporting site, for this analysis, we considered this set of participants that were *very likely* to have been affected according to the above criteria.

Similarly, we determined which participants had a Yahoo! account by searching for participants who had entered a password on the `yahoo.com` domain before each of February 15, 2017—when the breach had first become public—and October 3, 2017—the time of the second breach announcement.

Studying which people read about incidents

After determining which participants read about an incident, we studied what participant characteristics correlated with visiting pages related to the security incidents. We modeled participants and their behavior using three feature sets and then performed a logistic regression for each feature set, where the binary outcome variable in each regression indicates whether a participant read about an incident.

Feature set 1: Demographic characteristics Based on findings from prior work showing that demographics were correlated with how people share security and privacy news and their comfort with uses of breached data [79, 126], we hypothesized that certain demographics would also be correlated with whether participants read about a security incident. Therefore, the first feature set contains the following demographic information: age, gender, income, highest education level, student status, whether the participant's primary profession involves programming, and whether the participant knows at least one programming language. The last two demographics are included to serve as measures of technical savviness.

Feature set 2: Self-reported security intentions Prior work found self-reported security intentions (as measured by the SeBIS scale [86]) to be correlated with how people heard about and

shared security and privacy news [79]. Hence, our second feature set is comprised of four continuous feature values of the SeBIS scale (values in $[1, 5]$), which participants optionally filled out upon enrollment in the SBO. The four values represent the extent to which participants (1) secure their devices, (2) generate strong and varied passwords across accounts, (3) demonstrate proactive awareness of security issues or safety of websites and links, and (4) update the software on their computers.

Feature set 3: Participants' observed internet behavior We hypothesized that the types of web pages people browse are correlated with their likelihood to encounter information about a security incident (e.g., people who browse more technology-related news articles are more likely to come across web pages about security incidents). To test this hypothesis, we examined the kinds of topics of web pages that participants typically visited and the amount of their web browsing that involved visiting web pages on technical topics. We describe each of these next.

Characterization of browsing behavior: We used the Non-negative Matrix Factorization (NMF) topic-modeling algorithm [143] to generate a set of topics that categorized participants' browsing. NMF has been used in prior work for mining browsing behavior patterns [124, 203].

We built this model by first looking up the category of the domain of every web-page visit in Alexa Web Information Services (AWIS) [6]. We then created one document per participant, each consisting of the tokens of the AWIS categories of the participant's web-page visits. For example, the category for `google.com` is `Top/Computers/internet/Searching/-Search_Engines/Google`.

For each participant, we tokenized the AWIS categories of the domain of each page visit and discarded the "Top" token to create a multiset of tokenized categories. If a domain appeared multiple times in a participant's browsing history, the tokenized AWIS categories appeared an equal number of times. We then computed the term frequency-inverse document frequency (TF-IDF) score [150] for each token to produce the document-token matrix to be used as input by the NMF algorithm.

We applied the Non-negative Matrix Factorization (NMF) topic modeling algorithm [143] to the matrix. We varied the number of topics from two to 10 and identified the optimal number of topics by observing when the most-frequently occurring tokens in a topic were on average most similar to each other [178, 209]. To determine this similarity, we computed the average of the pairwise cosine similarities of the top 20 tokens within each topic, using a Word2Vec model [162] trained on the same documents used to train the topic model, and then averaged these average similarities. The average within-topic cosine similarity was highest when the number of topics was two.

We examined the top 20 tokens in each topic to determine the themes of the topics. The words in one topic seemed to represent more leisure-oriented browsing ("Social_Networking", "Shopping", etc.) and the other professional-oriented ("Education", "Business", "E-mail", etc.).

NMF also outputs a value that describes how much of each participant's web browsing matches each topic. We use these values as the features that characterize participants' browsing behavior.

We also experimented with Latent Dirichlet Allocation (LDA) [59] but we observed the re-

sulting topic clusters to not be as coherent as the ones derived with NMF.

Amount of technical content browsed: We also characterized people's browsing by how many of the web pages they visited were technical or technology-related. We again used NMF to build a topic model with two distinct topics: (1) technical or technology-related content and 2) all other content.

We trained this model using a 1% sample of each of the participants' browsing histories and Alexa categories described in above. Here the input only contains two documents: one for the technical webpages and one for the non-technical pages. We built each document to be a representation of the content that is categorized as technical or non-technical by its domain's AWIS category, i.e., the technical document contains content about web pages that have the word "Technology" or "Technical" in its AWIS category and the non-technical document contains all other content. We downloaded the content of each web page in the sample with the newspaper library [172], tokenized each page's content, and concatenated the tokens from all technical web pages to construct the technical document and from all other web pages to construct the non-technical document (from a sample of web pages of the same size as the set of pages in the technical category). When computing the TF-IDF scores for tokens in each document, we only included tokens that appeared in one document but not the other. This way, we could construct topics with tokens that were unique to either the technical or non-technical category.

After training the two-topic topic model, we determined the index of the column in the resulting document-topic matrix that corresponded to the technical document and therefore, to the technical topic. We applied the trained model on the sample of each participant's browsing history (a multiset of tokens of the page content of web pages with a defined AWIS category). The model computed two weights for each participant, of which we used the weight of the technical topic as the feature characterizing the amount of a participant's browsing related to technical content.

3.3.2 Results

The 303 participants we studied spanned a broad range of demographics. Table 3.1 shows the demographic distribution of the participants. We were surprised to discover that only 48 of the 303 (16%) read about[1] *any* of the six incidents[2]. In three additional instances, participants searched for incident-related keywords but did not visit any of the search results. Table 3.2 shows how many participants visited a page about each incident. This was computed based on browsing history from participants' home computers, which our confirmatory study suggests accounts for the majority of participants' browsing (see Section 3.5).

We also examined a subset of participants that we hypothesized were particularly likely to have been affected by the Equifax or Yahoo! breaches (see Section 3.3.1). Table 3.3 shows that very few likely affected participants for both incidents read about each incident. For example, of the 59 participants with likely Equifax reports, only 15 read about the incident in our dataset. We substantiate these low numbers through our follow-up study (Section 3.5).

[1]We say that these participants "read about the incident," even though we cannot confirm they understood the content of the pages they visited.

[2]While not all participants may be interested in *every* incident, the incidents we study were chosen so that the majority of participants was *affected by* one or more incidents.

Category	*Distribution*
Age	Range: 20 to 83 Mean age: 36 Median age: 29 Standard deviation: 16.28
Gender	Female: 59% Male: 41% Did not provide: <0.5%
Education	High-school: 8% Associates degree or some other college: 29% Bachelor's degree: 40% Advanced degree: 23%
Income	$\leq$ 50k: 50% 50k-100k: 24% 100k-200k: 10% $\geq$ 200k: 2% Did not provide: 14%
Is_student	Students: 48%

Table 3.1: Demographic distribution of the 303 participants.

We analyzed the relationship between the binary outcome of whether a participant read about any of the incidents and each of the three feature sets described in Section 3.3.1 by computing three logistic regression models. When interpreting results, we used a significance level (p-value) of 0.05.

First, we computed a model exploring the effect of demographic characteristics over the 303 participants. Table 3.4 shows the results of the model (see Appendix A for model assumptions). We found that participants' ages and whether they knew a programming language were significant factors. Specifically, older and more technology-savvy participants were more likely to read about incidents. The odds of reading about an incident increased by $3.216\times$ ($p = 0.017$) if the participant was 50 years old or older and the odds of reading about an incident increased by $3.917\times$ ($p = 0.002$) if a participant knew a programming language.

Incident	*# users*	*% users*
Equifax	26	54%
Yahoo!	6	13%
Uber	4	8%
Ashley Madison	6	13%
WannaCry	14	29%
Panama Papers	10	21%

Table 3.2: Number of participants who read about each security incident; some read about multiple incidents.

Incident	*# users likely affected*	*% users that read*
Equifax	59	25% (15)
Yahoo!	48	2% (1)

Table 3.3: Number of participants who were likely affected by the Equifax or Yahoo! breaches and of those, the number of participants who read about the incident.

feature_name	*baseline*	*coef.*	*exp(coef.)*	*std. err.*	*z*	*p*
(Intercept)		**-2.293**	**0.101**	**0.492**	**-4.661**	**<0.01**
age: ≥ 50	**<50**	**1.168**	**3.216**	**0.491**	**2.380**	**0.017**
gender: male	female	0.011	1.011	0.346	0.032	0.975
gender: not provided	female	-11.663	<0.01	882.744	-0.013	0.989
education: ≥ ugrad	<ugrad	-0.080	0.923	0.349	-0.229	0.819
income: >$50k	<$50k	-0.637	0.529	0.378	-1.684	0.092
income: declined to answer	<$50k	-0.530	0.588	0.544	-0.975	0.329
knows_prog_lang: yes	**no**	**1.365**	**3.917**	**0.441**	**3.093**	**<0.01**
is_programmer: yes	no	0.090	1.094	0.423	0.212	0.832
is_student: yes	no	-0.075	0.927	0.480	-0.157	0.875

Table 3.4: Logistic regression model describing the relationship between whether a participant read about an incident and characteristics of the participant including their demographics. "Ugrad" denotes that the participant indicated achieving a Bachelor's degree.

Our second model examines the relationship between whether participants read about an incident and their self-reported SeBIS scale values. Table 3.5 shows the results of the model (see Appendix A for model assumptions). This model was computed over 247 participants who provided SeBIS data to the SBO at the time of enrollment. Only one of the four SeBIS scale values was statistically significant, which we modeled by its Z-score for easier interpretation; the odds of reading about an incident were increased by a factor of 1.594 ($p < 0.01$) for each standard deviation increase in the SeBIS proactive awareness score of a participant.

feature_name	*coef.*	*exp(coef.)*	*std. err.*	*t*	*p*
(Intercept)	**-2.828**	**0.059**	**1.344**	**-2.105**	**0.035**
device_securement	0.044	1.045	0.151	0.290	0.772
password_generalization	0.386	1.471	0.345	1.120	0.263
Z(proactive_awareness)	**0.467**	**1.594**	**0.169**	**2.768**	**<0.01**
updating	0.167	1.182	0.200	0.836	0.403

Table 3.5: Logistic regression model describing the relationship between whether a participant read about an incident and the SeBIS scale values they provided. The proactive_awareness feature was represented by its Z-score.

Our third model examines the relationship between reading about an incident and participants' internet browsing behavior (i.e., browsing topics and amount of technical browsing; see

Section 3.3.1). Table 3.6 shows the results of the model (see Appendix A for model assumptions). This model was computed over 302 participants who had enough browsing data from which a sample sufficient for computing the technical browsing descriptor could be drawn (see Section 3.3.1). Of the factors examined by this model, only the amount of technical or technology-related browsing was a significant factor (again modeled by its Z-score). The odds of reading about an incident were increased by a factor of 3.315 ($p < 0.01$) for every standard deviation increase in the technical browsing score.

feature_name	*coef.*	*exp(coef.)*	*std. err.*	*t*	*p*
(Intercept)	**-4.773**	**0.008**	**1.760**	**-2.712**	**<0.01**
Z(browsing_technical)	**1.199**	**3.315**	**0.464**	**2.582**	**<0.01**
sq(browsing_leisure+1)	1.519	4.566	1.253	1.212	0.226
browsing_professional	4.859	1.289	3.187	1.525	0.127

Table 3.6: Logistic regression model describing the relationship between whether a participant read about an incident and characteristics of their internet browsing behavior. The browsing_-leisure feature was applied a square transformation to meet the linearity assumptions of logistic regression and browsing_technical was represented by its Z-score.

We built a fourth logistic regression model in which the features were the four significant features from the above three regression models. Table 3.7 shows the results of this model (see Appendix A for model assumptions) in which all features were again found to be significant and to increase the likelihood of reading about an incident. For example, with this model, participants over the age of 50 were $4.021\times$ more likely to read about an incident than their younger counterparts.

feature_name	*baseline*	*coef.*	*exp(coef.)*	*std. err.*	*z*	*p*
(Intercept)		**-5.655**	**0.004**	**1.108**	**-5.104**	**<0.01**
age: $\geq$ 50	**<50**	**1.391**	**4.021**	**0.463**	**3.005**	**<0.01**
knows_prog_lang: yes	**no**	**1.201**	**3.322**	**0.414**	**2.898**	**<0.01**
sebis_proactive_awareness		**0.911**	**2.487**	**0.366**	**2.489**	**0.013**
browsing_technical		**1.465**	**4.326**	**0.458**	**3.196**	**<0.01**

Table 3.7: Logistic regression model describing the relationship between whether a participant read about an incident and the four significant features from Tables 3.4, 3.5, and 3.6.

3.3.3 Summary of findings

Overall, participants who were older, more proactively aware about computer security, and who were more technology-inclined were more likely to come across information about security incidents online. This indicates a potential imbalance in the dissemination of important security information.

3.4 How people learn about incidents and take action

We now study how the 48 participants who read about security incidents came to visit incident-related web pages and what behavior they exhibited in response. We first explain how we characterize reading about (discovery) and taking action after an incident (Section 3.4.1). We then examine how the characteristics of discovery or of the incident relate to participants' reactions (Section 3.4.2).

3.4.1 Methodology

We defined features that characterize the process of discovery of web pages about incidents (Section 3.4.1) and we characterized participants' actions after discovery (Section 3.4.1).

Learning about incidents

We examined the *browsing trajectories* (sequences of page visits that surround the visit of an incident-related web page) of each participant for each incident. We then measured the characteristics of the page visits that were part of the trajectory before the visit to an incident-related web page, and, separately, the characteristics of page visits after the first visit to the incident-related web page. We analyzed how the actions people take—as observed by examining the part of the trajectory after first visiting the incident-related web page—were related to characteristics of the incident and of the browsing path up to reading about the incident, participants' demographics, and browsing behavior.

We constructed browsing trajectories as follows: we first identified each participant's visits (if any) to web pages related to any of the incidents from Section 3.3.1. For each *first occurrence*—the first visit to any incident-related page about a specific incident—we defined a trajectory to be composed of the 20 page visits immediately preceding this first visit, the actual first visit, and the 20 page visits that immediately followed. In this manner, we constructed one trajectory per incident for each participant who visited any page about that incident.

To study how people read about incidents through browsing and their subsequent actions, we defined and analytically examined several features:

Precursor web page type (`precursor_type`) This feature describes the type of web page on which the participant clicked on a link that took them to the first occurrence of a page about the incident. This is commonly called the "referrer" page; we call it the precursor page because we identified these pages manually instead of via referrer headers, which are often not available. To create this feature, we manually categorized all the precursor pages as follows:

- **Social media:** A social media site (e.g., Facebook) page or home page.
- **Message boards:** A message forum such as 4chan message boards or Reddit.
- **News page:** A web page on a news website.
- **Purposeful:** Search engine results about the incident.
- **General browsing:** The page did not fall into one of the above categories but contained a link to the first-occurrence page (e.g., a stackexchange page with a sidebar link to an incident-related page).

- **Unknown:** No pages in the trajectory preceding the first occurrence of an incident-related page appeared to have a link to that page (e.g., the participant entered the URL manually or clicked on a link in an external tool).

Whether the precursor page was a home page (`precursor_is_homepage`) This feature captures whether the precursor page was a home page or whether the participant had to have browsed more deeply into a website before encountering the precursor page. We examined this feature to determine whether the link to the first-occurrence page was easily visible to anyone (i.e., on a home page) or would be seen only by some visitors to that site (i.e., those who navigated to a specific section).

First-occurrence page type (`1st_occur_type`) This feature categorizes the first-occurrence page according to whether it was specifically about the incident, and, if so, whether it was descriptive or prescriptive. We used the newspaper library [172] to extract the main content of each page. We then manually examined the content and developed the following three categories, using which we then classified each page: (1) general information about this specific incident (e.g., what caused it); (2) advice about this specific incident (e.g., what to do in response or how to find out if one is affected) and; (3) not specifically about this incident, but mentions it (e.g., a political article that mentions the incident).

First-occurrence page sentiment (`1st_occur_sentiment`) Inspired by research on how the sentiment of social media posts influences the poster's followers [44], we hypothesized that people's reactions to web pages about incidents might be related to the sentiment of the pages. In particular, we hypothesized that a positive sentiment might correlate with more constructive action. We first computed the sentiment for the main content of each first-occurrence page (collected as described above) using the NLTK Vader library [103]. The library returned a score in $[-1, 1]$; lower values indicated more negative sentiment, higher more positive sentiment, and values closer to zero neutral sentiment. The feature we defined consists of three categories depending on this score: "positive," "neutral," and "negative". Scores greater than or equal to 0.1 fell into the "positive" category; scores less than or equal to -0.1 fell into the "negative" category; and all other scores were classified into the "neutral" category.

Incidents previously read about (`incident_num`) We hypothesized that people react to incidents differently depending on how many incidents they have come across through web browsing. Hence, for each incident that a participant read about, we counted the number of trajectories previously constructed for this participant for other incidents and exposed this as a feature in our analyses.

Type of data compromised (`data_compromised`) This feature represents the type of data compromised in the incident. We broadly grouped the data types and incidents as follows: **PII:** names, phone numbers, partial credit card numbers, email or physical addresses (Ashley Madison[3], Uber); **PII++:** PII with credit card information or social security numbers (Equifax); **passwords** (Yahoo!); and **miscellaneous** (WannaCry, Panama Papers);

[3]We categorized the Ashley Madison breach as only including PII and not passwords since passwords were cracked and leaked months after the original leak became public.

Actions after reading about incidents

We manually examined the 20 page visits in each trajectory immediately following the first occurrence as well as any visits to incident-related web pages *after* the first visit. We called one of these page visits an *action* taken in response to reading about the incident if it fell into at least one of the following categories:

- **Educating themselves about the incident:** e.g., reading additional articles about how the incident occurred, who was responsible, or implications of the incident.
- **Educating themselves about general security:** e.g., reading articles about how to secure their network or whether using personal emails for work is safe.
- **Taking action to make themselves more secure:** e.g., attempting to freeze their credit reports, visiting a website to download patches after a cyber attack, or reading "what you need to do" articles.

We then counted the number of actions after reading about an incident. We use this raw count of actions as the outcome variable in our analyses (see Section 3.4.2). For example, if a participant visited two more pages that discussed the incident as well as one page with a "what you need to do" article, this would count as having taken three actions after reading about the incident.

So as to treat incidents uniformly, we did not consider actions tailored to any specific incident (e.g., changing passwords after a password breach).

3.4.2 Results

In Section 3.4.2, we describe how participants came to read about incidents and their subsequent actions. In Section 3.4.2, we report on the amount of action participants took in relation to the type of incident, how they came across incident-related pages, participant characteristics, and their browsing behaviors.

Descriptive results

Using the methodology described in Section 3.4.1, we identified 66 distinct trajectories across the 48 participants (out of 303) who visited an incident-related web page. About twice as many trajectories described a participant reading about a PII++ breach (26) than a PII breach (10), and about four times as many described a PII++ breach than a password breach (6).

The types of web pages that led participants to visit the first incident-related page (`precursor_type`) were relatively evenly distributed across social media (11), message boards (9), news pages (13), searching for the incident (9), and general browsing (10). For 14 trajectories we could not identify the precursor page. For the other 52, approximately half (24) the precursor pages were home pages (`precursor_is_homepage`); the others (28) were pages deeper in a website.

When we categorized the first incident-related page visits according to their content (`1st_occur_type`), we found that 10 were advice articles, 35 were pages with general information about the incident, and 21 had content related to the incident (e.g., a story about a woman's identity stolen 15 times after the Equifax breach) without specific information about the incident. The

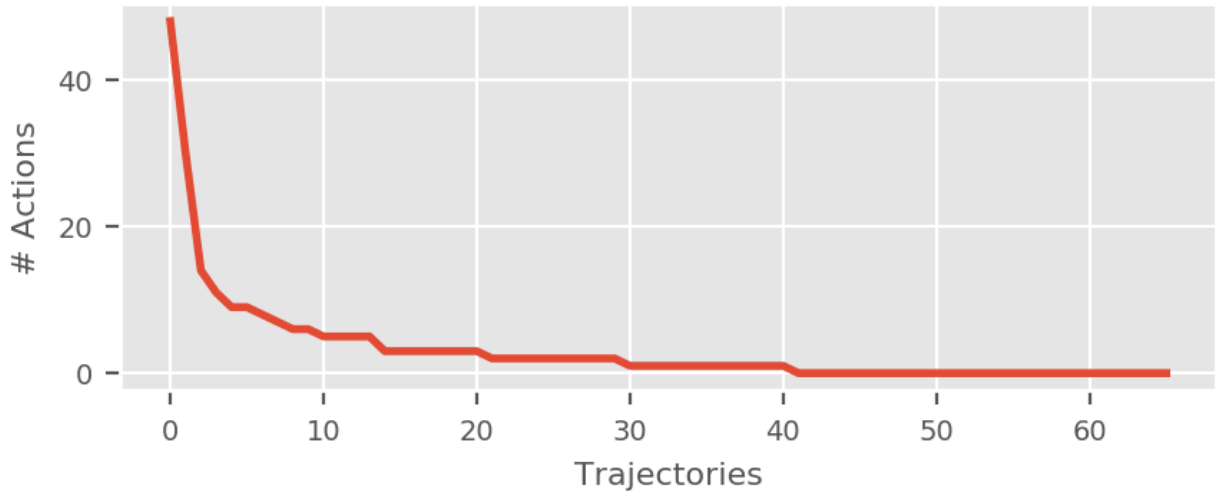


Figure 3.1: Number of actions taken per trajectory.

sentiments of the first-visited incident-related pages (`1st_occur_sentiment`) were slightly positively skewed, with 31 trajectories categorized as "positive," 20 as "negative," and 15 as "neutral".

Most participants (71%) visited pages about only one of the six incidents. Only 10 visited pages about two incidents, two about three incidents, and two visited pages about four distinct incidents.

Most participants (73%) who visited an incident-related web page afterward took at least one action (i.e., visited another page about the incident, a page about security in general, or a page describing how to react to an incident). The mean number of actions taken across the 66 trajectories was 3 with a standard deviation of 7.19, the median 1, and the maximum 48. Figure 3.1 shows how the number of actions is distributed across the trajectories.

Relating actions to features

We now examine the relationships between how much action participants took and four feature groups: features relating to the trajectory, the participant's demographics, the type of incident, and the participant's internet browsing behavior.

Three of the analyses are over the 48 participants (and 66 trajectories) who read about an incident. The analysis of participants' reactions relative to what led them to visit an incident-related web page is over 52 trajectories, since we removed trajectories for which we could not determine what led the participant to visit an incident-related page (i.e., `precursor_type`was "Unknown"; see Section 3.4.1).

Actions in relation to trajectories The first analysis studies the number of actions participants take related to the following features of the browsing trajectories: `precursor_type`, `precursor_is_homepage`, `1st_occur_type`, `incident_num`, and `1st_occur_sentiment`.

We modeled this relationship by a quantile regression model, a non-parametric linear model suitable when the assumptions of linear regression are not satisfied [132]. In particular, we built quantile regression models to predict the conditional 0.25, 0.5, 0.75, and 0.9 quantiles of the number of actions outcome [62]. In this analysis, each trajectory (not each participant) is one data item and we used the raw action count as the outcome variable. We modeled each of the categorical features (with n levels) with $n-1$ indicator variables compared to a baseline level.

The only factor that was correlated with the number of actions participants took after visiting an incident-related page was the sentiment of the content of this first visited page. This feature was only significant for the trajectories with many actions in the 0.9 quantile (90th percentile), where the number of actions in the 0.9 quantile increased by seven if the article's sentiment was positive instead of negative.

Table 3.8 shows the results of the quantile regression model for the 0.9 quantile.

feature_name	***baseline***	***coef.***	***std. err.***	***t***	***p***
(Intercept)		2.000	3.600	0.556	0.581
`precursor_is_homepage`: yes	no	-1.000	1.886	-0.530	0.599
`1st_occur_type`: info	advice	5.000	3.152	1.586	0.120
`1st_occur_type`: related	advice	1.000	3.403	0.294	0.770
`incident_num`: not first	first	3.000	2.035	1.474	0.147
`1st_occur_sentiment`: pos	**neg**	**7.000**	**2.485**	**2.816**	**<0.01**
`1st_occur_sentiment`: neu	neg	-1.000	2.875	-0.348	0.730

Table 3.8: Quantile regression model for the 0.9 quantile of the relationship between actions participants took and the trajectories that led them to reading about incidents. We grouped the values of `incident_num`into two buckets: "first" and "not first," where the second bucket means that there was at least one incident previously read about. The model excludes `precursor_type`due to its correlation with `precursor_is_homepage`.

Actions in relation to participant demographics The second analysis examines the number of actions participants took relative to their demographics, via a linear regression (see Appendix B for model assumptions). Table 3.9 shows the detailed results. If a participant read pages about multiple incidents with multiple trajectories, we averaged their actions across the trajectories.

No participant descriptors were statistically significant in relation to the actions participants took. We conducted a power analysis following previous work [37, 70, 87, 153, 167, 217] aiming for an experimental power of 0.8, a p-value (α) of 0.05, and a medium effect size. We calculated that a minimum sample size of 36 was necessary to see medium-sized effects with our model, a criterion which our sample of 48 participants exceeded. This suggests that if our model did not show a factor to be statistically significant, that factor was likely to not have had a "medium" or greater effect.

Actions in relation to the type of incident We examine the relationship between the number of actions taken per trajectory and the type of data compromised (`data_compromised`) using

feature_name	*baseline*	*coef.*	*std. err.*	*t*	*p*
(Intercept)		0.782	0.533	1.467	0.150
(1/age)		6.932	19.085	0.363	0.718
gender: male	female	-0.271	0.299	-0.906	0.371
knows_prog_lang: yes	no	0.143	0.301	0.477	0.636
is_programmer: yes	no	-0.536	0.366	-1.465	0.151
is_student: yes	no	-0.028	0.439	-0.063	0.950
education: $\geq$ ugrad	<ugrad	0.507	0.284	1.785	0.082
income: >$50k	<$50k	-0.424	0.291	-1.453	0.154
income: declined to answer	<$50k	-0.142	0.420	-0.337	0.738

Table 3.9: Linear regression model of the relationship between actions participants took and their demographics. The outcome variable is log(actions taken) + 1. The age feature was transformed to its reciprocal for the model to meet the assumptions of linear regression.

the Kruskal-Wallis one-way test of variance [137].

The amount of action differed significantly between categories of the data compromised ($\chi^2 = 19.843$, $df = 3$, $p < 0.001$). To understand which groups were statistically different from each other and in what direction, we conducted a post-hoc analysis with pairwise comparisons using Dunn's test between each group, applying the Bonferroni correction for each comparison [58, 84].

Participants took an average of 5.35 actions after a PII++-compromised incident, 3.04 after a miscellaneous incident, but only 0.5 and 0.3 after reading about a passwords or PII incident, respectively. The greater number of actions taken for PII++ was significantly higher than for passwords ($Z = 3.002, p = 0.02$) or just PII ($Z = 3.85, p < 0.001$). Figure 3.2 shows a ranking of the average number of actions taken for the trajectories of each category.

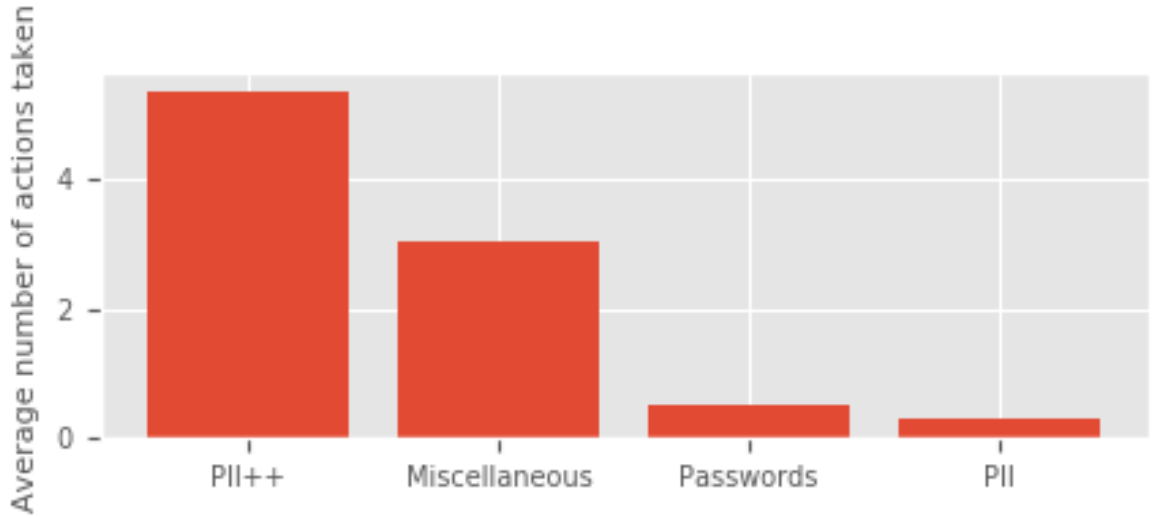


Figure 3.2: Number of actions taken on average for each incident type.

Actions in relation to participants' browsing behavior Finally, we tested for relationships between the number of actions people took and the types of pages they visited on the web and the amount of technical browsing (as described in Section 3.3.1). The linear regression model, though suitable for this analysis (see Appendix B for model assumptions), did not reveal any statistically significant relationships, although, as before, a power analysis showed that we had sufficient power to see medium-sized or larger effects. Table 3.10 shows the results of this model.

feature_name	*coef.*	*std. err.*	*t*	*p*
(Intercept)	**1.304**	**0.419**	**3.111**	**<0.01**
browsing_technical	0.159	0.324	0.491	0.626
browsing_leisure	-3.413	2.190	-1.558	0.126
browsing_professional	-0.671	1.399	-0.479	0.634

Table 3.10: Linear regression of the relationship between actions participants took and characteristics of their internet browsing behavior. The outcome variable is log(actions taken) + 1.

3.4.3 Summary of findings

In summary, participants came across pages about security incidents through a variety of media in similar proportions. Most of the times participants came across the page about the incident after browsing deeper through a website, suggesting that such pages about incidents are not easily accessible (e.g., from a homepage). Participants were likely to read more about the incident or take an action when the page exhibited a positive sentiment but no other features were correlated with taking action, implying that the lack of action was nearly universal across our dataset.

3.5 Confirming dataset validity

Our findings (Sections 3.3–3.4) are based on the browsing activity collected from one home computer of each participant. However, participants could have read about incidents or taken action on other devices, data about which is not captured in our dataset.

To shed light on how representative our dataset is of participants' overall browsing behavior, we collected additional self-reported data. We conducted a survey of 109 SBO participants who were active in May 2019, in which we asked about their familiarity with, and any actions after, several security incidents, as well as about how much web browsing they perform on which devices (see Appendix C). This study was approved by the review board at Carnegie Mellon University. The survey took between one and five minutes and participants received $5. Many participants in our main dataset were not active when we conducted this follow-up survey and vice versa; 84 of the 109 survey participants were in our original SBO dataset. Hence, we use this survey as *a measure of the self-reported behavior of SBO participants in general*, rather than of individuals who were in both datasets. (Results for the 109 participants described in this section are consistent with results computed over the overlapping 84 participants only.) Since the 109

participants may not have been active in the SBO around at least one of the incidents we studied, we did not ask participants about each of those incidents. Instead, we asked about a variety of events and security incidents, and additionally about incidents with a wide impact.

Participants reported that they conducted, on average, 59% of their web browsing on their SBO computers. As the amount of browsing on desktop and laptop computers may have decreased over time in favor of browsing on mobile devices [158], this 59% is likely a lower bound. Participants earlier in the study likely performed more of their overall browsing on their SBO computers.

We also asked participants how often (on a 5-point Likert scale [220]) they read about security incidents on (1) their SBO computer or (2) any other devices. We found no significant difference between the two distributions (Kolmogorov-Smirnov [133]: $D = 0.056, p = 0.997$). We also found no significant indication that the distribution of browsing on SBO computers vs. other devices varied by participant age (Spearman's correlation test: $S = 180990, p = 0.155$). Finally, to gauge the accuracy of the self-reported data, we asked how familiar participants were (on a 5-point Likert scale) with five security incidents, four non-incident-related events, and one fictitious security incident (an Airbnb social security number breach). 8% indicated moderate or extreme familiarity with the fake Airbnb breach, suggesting that the self-reported results may exaggerate actual familiarity.

Our results from Section 3.3 indicate that 25% of the participants in our main dataset who were likely affected by the Equifax breach read about the breach, a surprisingly small percentage. If we assume that this percentage is computed based on 59% of all browsing, then the actual percentage of people who read about the breach—if our data included 100% of all browsing—could be as high as 42%, which is still low considering the significance of the breach. However, when asked what action they took following the Equifax breach, 41 of the 86 survey respondents who indicated at least slight familiarity with the breach (48%) responded that they read about the breach online or visited the Equifax website, with the majority of the rest answering "didn't do much/didn't do anything". Five of the 41 respondents additionally replied that they "can't remember" and/or "didn't do much/didn't do anything," implying that the actual number of participants who read about the incident online or visited the website might be even lower than reported.

Similarly, our results from Section 3.3 indicate that 2% of participants in our main dataset who had a compromised Yahoo! password [142] read about the breach online. Self-reported data also suggests low awareness: when asked about reading and reacting to the Yahoo! breach, only nine of the 72 participants who indicated at least slight familiarity with the breach (13%) answered that they read about the breach online. Two respondents answered "can't remember" and/or "didn't do much/didn't do anything" in addition to reading online, again indicating that a lower number of participants than self-reported may have actually read about the incident.

Overall, our results suggest that the browsing data that we used for our analyses (Sections 3.3–3.4) covers the majority (with a lower bound of approximately 59%) of the browsing performed by the participants. While the additional browsing participants performed on non-SBO devices may dilute some of our findings, the self-reported data supports the big picture: a surprisingly small subset of users may read about incidents and try to learn more about the incidents.

3.6 Limitations

Although our work provides valuable insights into how, and the extent to which, people attempt to make themselves secure after an incident, it is subject to a few limitations.

Our dataset contains data about (relatively) few participants due to the difficulty of recruiting participants to the SBO. However, the SBO data is a tradeoff: it offers rich browsing and password data that is typically infeasible to obtain, at the cost of a limited participant pool and concerns about generalizability. We believe it is the big picture revealed by our results that matters – that a very small fraction of people seem to engage with information about security incidents – rather than the specific percentages involved. Similarly, the browsing history may not be representative of all the browsing users do. Hence the confirmatory study, which suggests that the high-level results of the original SBO analysis hold: participants were rarely familiar with or read about major security incidents regardless of the devices on which they browsed the internet.

Although we considered each monitored SBO computer to correspond to one participant, some of the computers were used by multiple users throughout the duration of the SBO study. We attempted to remove password entries made by users that were not the main SBO computer users (see Section 3.2.1). However, the findings pertaining to browsing data of one participant may actually be based on multiple users' browsing. We found few instances of the monitored SBO computers being used by multiple users in practice (16% of the 303). Furthermore, results computed based on each computer being used by only the main user are likely to correspond to lower numbers than reported, which still supports our conclusions.

Since browsing history was represented via URLs and page titles, we could not include analyses that depended on the content of dynamic pages (e.g., social network pages). We also could not distinguish between content that participants consumed and content they loaded but did not read. Finally, since participants might have opened multiple pages in parallel and click on links in pages opened earlier, we could not always accurately determine the precursors to the first incident-related pages participants visited. In practice we found only a few instances where this was a problem.

The data we analyzed was collected only from Windows computer users. Windows is the dominant OS for personal computers [61], but users of non-Windows operating systems might exhibit behaviors different from the behaviors of the participants in our dataset.

Although data from SBO participants has been used for several security- or privacy-related studies [67, 98, 111, 174], the SBO participants may have been biased towards less privacy- and security-aware people, given the nature of the SBO data collection infrastructure.

Finally, the subsets of participants we used in specific analyses were of sufficient size to make uncovering medium-sized effects likely, but not so large as to reliably discover small-sized effects.

3.7 Discussion

Our dataset allows a comprehensive view of the actual browsing behaviors of 303 participants across 44 months. Although our sample is small, our results substantiated by our confirmatory study suggest a potentially bleak picture for security engagement in general.

Our findings show that for the people in our dataset, the consumption of security and privacy incident information was not as prevalent in people's online activities as was ideal. Even when information was presented and consumed, participants often did not attempt to learn more about the incident or show further interest in reading about it. On the one hand, further research is needed to study how this information can be disseminated more widely and be studied in the context of a general population. To elicit and encourage interest, websites should better highlight problems, the implications and risks, and suggestions for staying secure or maintaining privacy in light of the incident [225, 234, 236]. On the other hand, further research is needed to understand how companies can keep their users safe without requiring them to have security awareness and take action.

3.7.1 Improving dissemination of security incident information

Our results highlight the challenge of increasing awareness of security incidents. Although the Equifax breach affected more than half of adult US residents, and hence, likely the majority of the participants, only 25% of likely affected participants visited an article related to the breach. Without adequate awareness of such incidents, people are unlikely to understand the importance of safe security behavior or that the implications of the incident may be relevant to them even if they were not directly affected [207]. We confirmed through a Google search that each of the three most popular news sites [30] published multiple articles describing each of the incidents during the three months after each became public. That is, there appeared to be sufficient publicity about each incident. However, although these incidents were highly publicized, one might wonder why people are unlikely to read or learn about them. Perhaps "security fatigue" made people reluctant to learn more about them [208]. Additionally, while it is expected that people will be more engaged when they have more at stake (as supported by our findings that the most severe incidents were associated with more action), challenges remain with improving their engagement in the context of everyday services [202].

Recent work examined what kinds of "stories" are more likely to make an impact on people's security behaviors or to be shared [39, 185]. We found that a number of the incident-related articles participants first discovered were stories about the impacts of the incident, and not about the incident itself. Additionally, articles with a more positive sentiment were associated with people exhibiting security-enhancing behaviors in the form of trying to learn more about the incident. For instance, posts that had a high positive sentiment score had titles such as "The One Move to Make After Equifax Breach" or "'WannaCry' on Linux systems: How do you protect yourself?", which suggest that the articles contain constructive advice and information. On the other hand, an example of an article title with a strong negative sentiment was "The next ransomware attack will be worse than WannaCry," which seems largely about warning people to the perils of the incident. News organizations reporting on security incidents could encourage action by presenting constructive advice. In general, these organizations could benefit from research on what kinds of stories and content are most likely to influence security behaviors and the further sharing of such content.

3.7.2 Demographic factors related to dissemination and action

Most of the demographic factors that we examined were not significantly associated with the likelihood to come across incident-related articles or with the number of actions. However, older and more technology savvy participants were much more likely to have read about incidents. The latter may be because information about incidents is disseminated more towards technical audiences, perhaps because of the challenges of disseminating incident information (which may be seen as more technical) on non-technical outlets. Prior work found that security information is disseminated unevenly based on socio-economic status [189], which could be linked to technical savviness. Another potential explanation is that technologically savvy people are more receptive to such information, and so it remains an open challenge to convince less technologically savvy people about the importance of security incidents and effectively communicate online risks to them. Recent work found that video communication can raise the saliency of risk for people and concluded that it might be a more effective way to reach such populations compared to text [101]. Thus, in addition to exploring what types of "stories" are more effective, further work is needed to explore the medium of delivering such stories for different populations through different information channels.

3.7.3 Improving users' security without increasing awareness

We show that engagement with security information is low and needs to be improved to ensure users' security. However, given the very low rates of engagement, our findings also strongly suggest that relying on consumer awareness to ensure their security may not be viable.

Our analysis yielded several negative results, both in terms of what was correlated with coming across incident information and with following up on an incident. The negative results suggest that there may be a deeper issue in terms of how concerned about incidents people are in the first place. Reading articles about incidents may be too high of a burden if users do not have interest in the topics. Placing the burden of awareness on users also raises additional issues, such as requiring users to differentiate between correct and incorrect information they encounter online. For this to be effective, they need to already be educated about security practices. Users may also not always be inclined to act on security advice they come across if they find the insecure systems useful and serving of their purpose [107]. In fact, users are likely to reject security advice and make the decision to not improve their security if they perceive the cost of the associated effort to outweigh the potential harms [118].

Instead of relying on users to become aware of and seek help after incidents on their own, companies and organizations should inform their consumers of an incident directly with instructions on immediate remediation if they were affected. While breach notification is a legal requirement in certain countries such as the US and countries in Europe [5, 23], not all companies that suffered a breach or incident that we studied notified their customers. For example, Yahoo! did not notify their customers about their data breaches [161]. Companies should additionally provide detailed guidance on what consumers need to do to protect their data and accounts beyond on the affected site (e.g., identifying theft insurance, changing reused passwords on other domains). When possible, companies should also take immediate steps to protect their users on their behalf after a security incident. After password breaches, for example, companies can force

a reset on all passwords. For widespread cyberattacks, a patch can be automatically deployed to all computers on the affected platforms. To help users stay secure in general, system designers should consider from the start how to remove the responsibility of security decision-making from users [34, 118, 197].

3.8 Conclusion

Using the actual browsing histories of 303 participants over four years, we measured how often participants read web pages about security incidents, what actions they took after reading, and what factors were associated with how likely they are to read about an incident or take action. Our findings are bleak: only a small minority (16%) of participants visited an incident-related web page about at least one of six large-scale security incidents. Furthermore, few participants who read about an incident showed further interest in the incident or took some action by reading more about it.

Our results highlight the challenges of increasing awareness of security incidents and of disseminating information about them. Even when an incident was highly publicized and participants were likely to have been affected, few showed engagement with or awareness of the incident (e.g., only 25% read about the Equifax breach). Without adequate awareness, it is unlikely that people will act to improve their security. We found the low rate of discovery, and of constructive action after discovery, to be nearly universal across participants. Participants that were older, exhibited a higher proactive security awareness, or had an affinity for technology were more likely to read about incidents; but other factors including the remaining demographics that we explored had no impact. When reading web pages that spoke about the incident in a constructive way, participants were more likely to try and learn more about the incident or take action. However, no other factors correlated with taking action. Even though our results are based on a relatively small population, our results highlight the need for wide and effective dissemination of security incident information and advice and for exploring alternative avenues to ensure user safety that do not rely on user awareness and education.

To increase the dissemination of security advice, future work should examine how incident communications are featured in media frequented by more technology-savvy people and how their appearance in media with a more general audience could be modified to improve uptake. Future work should also study content-sharing platforms used by different demographic groups (e.g., social media) and understand how such platforms can improve the spread of incident news.

Chapter 4

(How) Do people change their passwords after a breach?

4.1 Introduction

As mentioned in Chapter 3, security incidents are rampant and we found that online engagement with incidents was generally low. In this chapter, we focus on a specific type of security incident: password data breaches. Password breaches have been on the rise, affecting mainstream companies such as Yahoo! and gaming sites such as League of Legends and Neopets among others [17]. Stolen passwords have been largely exposed in insecure forms such as in plain text or by weak hashes (often unsalted or easily guessed through dictionary attacks) such as MD5 and SHA$\bar{1}$ hashes, leaving users vulnerable unless they change their passwords on the affected sites [17]. Additionally, when a company suffers a breach involving passwords, rarely are the users affected solely on the compromised domain [75]. Previous work has shown that, on average, a user exactly or partially reuses their passwords on over 50% of their accounts [75, 96, 174]. In such cases, when a person's password on one domain is compromised, they incur the risk that an attacker will be able to gain access to their other accounts that use similar or the same passwords. In order to make informed recommendations to companies on best risk mitigation practices after a breach, it is instructive to examine people's current password-changing behavior after breaches.

Prior work has explored problems related to data breaches and changing passwords (e.g., how people comprehend data breaches [126, 235], what factors make them more inclined to take action after breaches [126, 235]), and how people change passwords in response to reuse notifications [106]. Researchers found that people were more likely to heed advice about actions after security breaches based on who was giving the advice and often underestimated the harm that could be incurred as a result of a compromise [126, 235]. Related to password changes, researchers found that very few of their participants in an online study reported intentions to change passwords after being notified that their passwords were compromised or reused, including because they believed in the "invincibility" of their passwords [106]. These studies are important to understand how to better inform people about the impact of data breaches and to understand people's mental models when it comes to taking action to protect themselves. However, we still lack an understanding of the actual extent—empirically measured—to which actions taken by

companies to inform their users after a breach are effective.

We make a significant effort towards developing this understanding. We analyze longitudinal, real-world password data over two years to understand whether people change their passwords after a breach and the quality of these password changes. Specifically, we examine: (1) whether people with an account on a breached domain changed their passwords after the breach and how constructive these changes were; (2) the extent to which people changed similar passwords on domains other than the breached domains; and (3) how password changes related to breaches compare to *all other* password changes.

Our dataset was collected from the home computers of 249 participants between January 2017 and December 2018 and includes *all* passwords used to log onto online services. Of the 249 participants, 63 had accounts on one of the breached domains we studied and were active in the study at the time of the breach announcement and for three months after. We found that only 21 of the 63 participants changed their password after a breach announcement and only 15 did so within three months of the announcement. The majority of these changes were after a high-risk breach (i.e., the Yahoo! breach). We also found that only a minority of password changes were to stronger or less reused passwords and that new and old passwords shared a substring on average almost half the length of the longer of the two passwords.

Participants who changed passwords on the breached domains had on average 30 accounts with similar passwords. Of the 21 participants who changed passwords, 14 changed at least one similar password within a month of changing their password on the breached domain. These 14 changed, on average, only four similar passwords within that month.

As a baseline for the quality of password changes, we looked at all password changes made by the 249 participants over the two-year period. A large fraction (70%) of the password changes resulted in weaker or equal-strength passwords. Similarly, a large fraction (68%) resulted in passwords that were more or equally reused across participants' other internet accounts. Old and new passwords on average shared a substring 85% the length of the longer of the pair. Overall, the properties of password changes on breached domains were roughly similar to the properties of the baseline password changes, though on average resulted in more dissimilar passwords.

Our results suggest that current breach notifications may not be effective, in that most users who were affected did not react sufficiently to mitigate their risk either on the breached domain or on others. Our results clearly indicate that more should be done—through breach notifications or other means—to induce users to change passwords both on the affected domain and especially on other domains, which users generally ignore. Similarly, additional means are needed to educate and encourage users to make their new passwords both strong and different from their existing passwords.

We first discuss the dataset we use in this chapter (Section 4.1.1). We then discuss our methodology for measuring password changes and their quality (Section 4.2) and the results of these measurements (Section 4.3). We conclude with limitations (Section 4.4) and a discussion of the implications of our work (Section 4.5).

4.1.1 Dataset

As in Chapter 3, we use data collected as part of the Security Behavior Observatory (SBO) project. However, this chapter only analyzes the password data collected by the SBO. As in the

previous chapter, the SBO and the work in this chapter were approved by the review board at Carnegie Mellon University.

The password data we analyze in this chapter spans from January 2017 and December 2018 and includes 249 participants who participated in the SBO study for at least 90 days during that period. To collect this password data in particular, the SBO browser extensions recorded every entry into an HTML input field at the time of browser events such as clicks, key presses, form submissions, and page loads. The dataset we examine includes information about every entry made into a password field in a web page, as determined by the browser extension, including: a salted one-way hash of the password, the URL of the form in which the password was submitted, the strength of the password (represented as the approximate number of guesses a sophisticated attacker would need to guess that password [159]), and hashes of all three-character-or-longer substrings of each password. Substring hashes are particularly useful for analyses related to partial password reuse (e.g., as used by Pearman et al. [174]). Password guess numbers less than 10 are rounded to 10 for easier comparison when $\log_{10}$-transformed. Throughout this chapter, we represent password strength by its $\log_{10}$-transform (see Section 4.3).

We further filtered this raw password dataset to remove failed login attempts and passwords entered by users other than the main SBO computer user employing the same approach as in Section 3.2.1.

4.2 Methodology

We study how participants changed their passwords after nine data breaches that became public in 2017 and 2018. We selected these breaches based on two broad criteria.

We started with a list of breaches comprised of:

- Identity Force's list of biggest breaches in 2017 [73] and Digital Information World's list of biggest breaches in 2018 [196]; and
- breached domains listed on `haveibeenpwned.com` (HIBP) for which breached data included passwords [17]. HIBP is a website that keeps track of sites that have been compromised and a service that people can query to find out whether their personal data has been compromised in a breach.

We then selected only those breaches that met the following criteria:

1. The breach *announcement* date overlapped with the time interval for which we had SBO password data.
2. At least one participant in our dataset entered a password on the breached domain before the breach announcement and remained active in the study for 90 days afterward.

This yielded the following nine breached domains, for which we studied participants' password-change behavior: Imgur (breach announced Nov. 2017) [146], Deloitte (Sep. 2017) [134], Disqus (Oct. 2017) [228], and Yahoo! (Feb. and Oct. 2017) [138, 142], MyFitnessPal (Mar. 2018) [93], Chegg (Sep. 2018) [69], CashCrate (Jun. 2017) [173], FLVS (Mar. 2018) [119], and Ancestry (Dec. 2017) [166].

For each of these breaches, we first identified participants who entered passwords on one of these domains, implying that they had an account on the domain and therefore were *potentially*

affected. We identified these participants as those who entered a password on at least one of the breached domains before the breach announcement date and were active in the study for at least 90 days after the announcement. We then checked whether identified participants changed their password on the affected domain. If they did, we checked whether the new password was stronger than the old one, how similar the new and old passwords were, whether they also changed similar passwords on other sites, and whether the password change caused less reuse between the password on their breached account and other passwords. We next describe the process of identifying password changes.

4.2.1 Identifying password changes

For each participant who had an account on at least one breached domain, we extracted the last password that they entered on the domain before the breach announcement date. We then looked for the first new password (i.e., different from the last one entered before the breach announcement) successfully entered on the breached domain after the breach announcement. If no new password was found, we concluded that the participant had not changed their password.

We also identified whether participants who changed their passwords on the breached domains changed any similar passwords on other domains. We considered two passwords *similar* if they shared a substring that was at least as long as half the length of the longer password. For example, the passwords "iluvDONUTS90" and "ih8DONUTS90" are similar since they share the substring "DONUTS90" that is at least half as long as the longer password, "iluvDONUTS90". We measured similarity by examining passwords similar to the last passwords entered on any domain before the breach announcement. If a participant changed their password on a breached domain, we examined whether they changed any of their similar passwords in the month that followed.

Even though our dataset directly captures passwords only when they are entered on participants' home computers, we are able to capture *password changes made from other devices too*, because we observed the new (or unchanged, if they have not been changed) passwords on the next login from participants' home computers. Many sites cache authentication credentials and do not require users to type in their password on every login. However, we studied people's behavior over a long enough period that authentication credentials, if properly implemented, would have timed out and participants would have had to eventually use their passwords to log in.

4.2.2 Measuring the effect of password changes

When participants changed their passwords on a breached domain, we computed how much stronger (or weaker) the new passwords were (as described in Section 4.1.1), the similarity between their old and new passwords, and whether the new password was more unique compared to passwords used on other accounts.

We computed the similarity between old and new passwords using a normalized similarity metric: the length of the longest common substring (of length ≥ 3) between two passwords divided by the length of the longer password. If two passwords did not share a substring longer than two characters, we considered them completely dissimilar [174].

To examine the relative uniqueness of the old and new passwords, we computed the difference in the amount of (exact or partial) reuse across a participant's passwords after and before they changed their password on the breached domain (results described in Section 4.3). We calculated the extent of reuse of the old password at the time of the latest entry of the old password, and the extent of reuse of the new password a month after the password change, i.e., a month after the first entry of the new password on the breached domain. We calculated this reuse after a month to allow time for the similar passwords on other domains to be changed. If a participant changed passwords on more than one breached domain, we computed the average.

Computing password reuse: To quantify password reuse, we build on the concepts of *exact* and *partial* reuse as defined in previous work on password reuse [174]. A password for a particular account is *exactly* reused if the same participant uses the same password on another account. A password is *partially* reused if it shares at least a three-character substring with another of that participant's passwords [174]. An *exactly-or-partially* reused password is one that satisfies either of these definitions.

Given a password on a domain, we computed its reuse score as the fraction of that participant's *other* passwords that exactly or partially reuse the password in question. We measured reuse based on the latest password entered by the participant on each distinct domain before a given point in time.

4.2.3 Computing baseline password-change statistics

To provide a baseline against which to compare breach-related password changes, we computed password-change statistics for all password changes by all 249 participants over the two years spanned by the dataset. For every instance of a new password per participant—ignoring the first occurrence of a password since those may have been created prior to the start of data collection—we captured the ratio of the strength of the new password to the old. We also captured the difference in reuse for every new password as the reuse score (computed as described above) of the old password subtracted from the reuse score of the new password in addition to the length of the substring (of at least three characters) shared by the new and old password. Finally, to have a baseline for how strong participants' passwords were overall, we computed the average strength of all of each participant's unique passwords entered per domain during the time period spanned by the dataset. If a participant had three unique passwords on `google.com` and five on `yahoo.com`, we computed the average strength of those eight passwords even if some of the `yahoo.com` passwords were exactly reused on `google.com`. We similarly computed a baseline for participants' overall reuse of each of their passwords where the reuse was calculated at the time of that password's first entry.

4.3 Results

4.3.1 Participants

Of the 249 participants, 60% identified as female, 39% as male, and the rest did not provide their gender. Ages ranged from 20 to 81 years with a mean of 34.1. A majority (57%) were students,

and 28% had professions that involved programming. Of the 249 participants, 63 had passwords on one or more of the nine domains involved in a password breach. Table 4.1 shows the number of participants who had an account on each breached domain.

Table 4.1: Number of participants who had an account on each breached domain; some had accounts on more than one of the domains.

Breached domain	***Number of participants***
yahoo.com	49
myfitnesspal.com	9
chegg.com	1
disqus.com	1
cashcrate.com	2
flvs.net	1
ancestry.com	7
imgur.com	6
deloitte.com	1
Total	63

4.3.2 Changed passwords

Only 21 of the 63 participants with passwords on a breached domain changed a password on the domain after the breach announcement. In total, 23 passwords were changed on these domains. Of the 21 participants, 18 were Yahoo! users; the remaining 31 Yahoo! users (out of 49) did not change their passwords although all were affected by the breach according to the breach announcement [142]. Two participants changed their Yahoo! passwords twice, once after each breach announcement. Two participants changed their password on the breached domain within one month of the breach announcement, a total of five within two months, and eight within three months.

4.3.3 Quality of new passwords

For each changed password, we measured the similarity between the old and the new password, the strength of the old and the new passwords, and the extent of password reuse before and after the password change (see Section 4.2). If a participant changed more than one password, we report the average results over all the participant's password changes.

Of the 21 participants who changed their passwords, nine created stronger (see Section 4.1.1) passwords and 12 created weaker passwords or ones of equal strength. On average, participants created new passwords that were $1.3\times$ stronger than their old passwords after transforming strength on the $\log_{10}$ scale (henceforth, all such comparisons are on $\log_{10}$-transformed strengths). Seven of the 21 participants who changed their password created a new password that shared at least a three-character substring with their old password; for all participants who changed a pass-

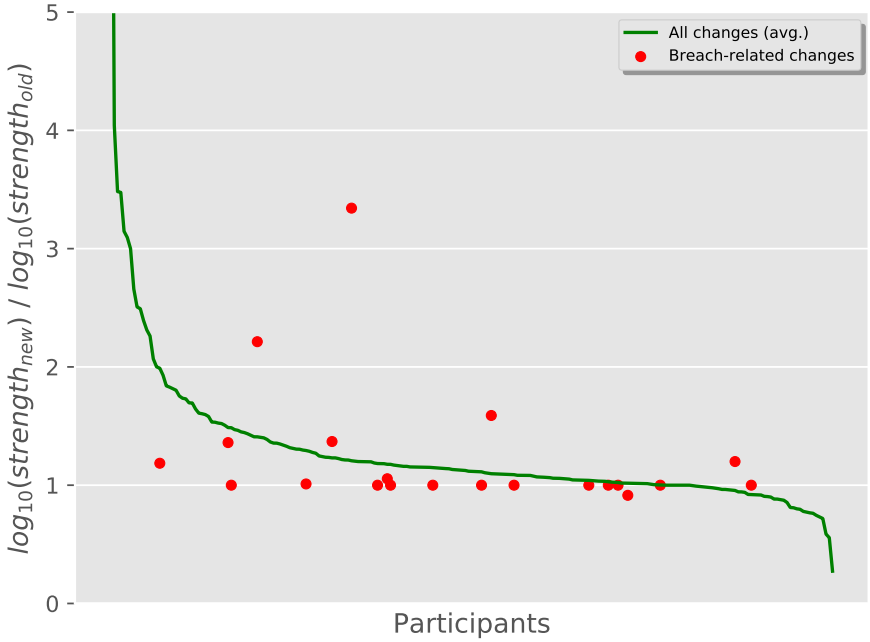


Figure 4.1: Change in password strength across each password change, per participant. Participants (x axis) are sorted by the average amount of improvement in password strength when they changed passwords. Y-axis values below one indicate that passwords became weaker.

word, new and old passwords shared a substring that was on average 41% as long as the longer of the two passwords.

The 21 participants who changed a password on a breached domain had, on average, 30 passwords similar to their older breached password (where similar passwords are those that share a substring of at least half the length of the longer password). 14 of these participants changed, on average, only four of these similar passwords on other sites within the month after changing their password on the breached site. These 14 participants changed their similar passwords to be on average $1.10\times$ stronger than their original password on the breached domain and $1.18\times$ stronger than the password being changed. However, the majority (63%) of the changes resulted in weaker or equal-strength passwords. Nine participants changed to a password that shared a substring of three or more characters with their old password; these nine participants' new passwords on average shared a substring 44% the length of the longer password with their older counterparts.

Overall, participants changed very few passwords on breached domains and even fewer similar passwords on other domains. Even when they did change a password, the change was often not constructive.

4.3.4 Password reuse

The passwords changed by participants were roughly evenly divided between being less reused and more or equally reused. We examined the change in password reuse for each participant who changed a password on a breached domain, comparing the reuse before the password change and a month after it. For nine participants the new password on the breached domain was more reused, for ten it was less reused, and for two it was equally reused.

In other words, while participants' new passwords were slightly stronger and often substantially different from their old passwords on the same domain, the new passwords on breached sites were still often similar to passwords on other domains.

4.3.5 Comparison to baseline password changes

Looking at *all* password changes by the 249 participants over the two year period, we observed 223 participants making a total of 3041 password changes, including the changes on the breached domains. 70% of these password changes resulted in weaker or equally strong passwords. However, new passwords were on average $1.23\times$ stronger than older passwords (again $\log_{10}$-transformed) and the median change in password strength was neutral (i.e., the old and new

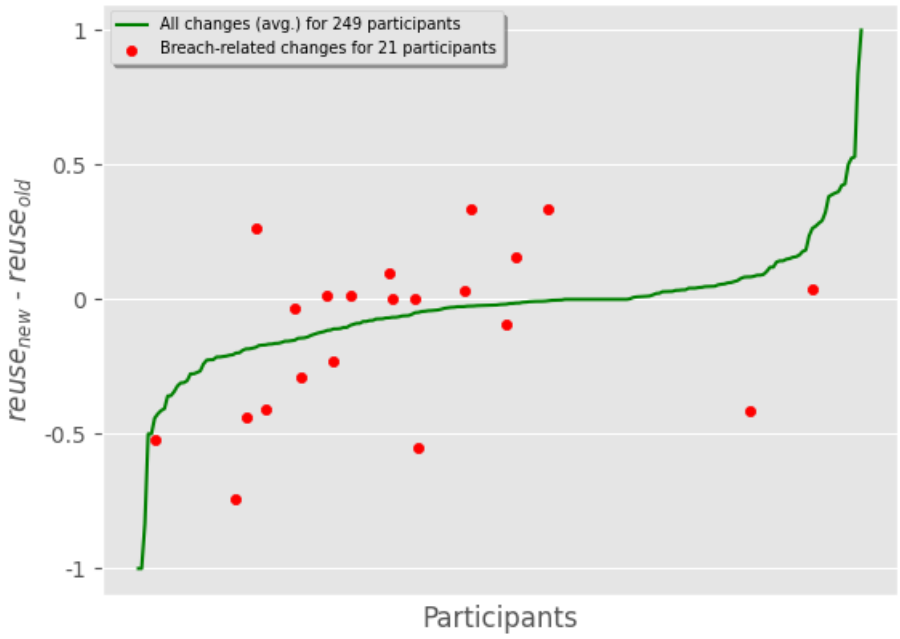


Figure 4.2: Change in password reuse across each password change, per participant. Participants (x axis) are sorted by how much more reused their changed passwords were on average than their old passwords when they changed passwords. Y-axis values below zero indicate that passwords became less reused which is more desirable.

passwords were equally strong). 68% of these changes resulted in equally or more reused passwords and all 223 participants who changed passwords made at least one password change that involved carrying over a substring of least three characters. In such cases, old and new passwords shared a substring, on average, 85% the length of the longer of the two.

Figure 4.1 shows, per participant, how changes in password strength for passwords on breached domains compared to changes in strength of other changed passwords. The green line on the graph shows the average increase in strength after a password change for each of the 223 participants over all their password changes. The red dots show password changes on a breached domain. Most participants' changes on breached domains resulted in slightly weaker passwords (red dots below the green line) and a minority resulted in substantially stronger passwords (red dots above the green line), compared to the average changes in password strength. Figure 4.3 shows the average strength of all of each participant's unique passwords entered per domain, computed as described in Section 4.2.3.

Similarly, Figure 4.2 shows, per participant, how much participants' new passwords were reused across their internet accounts compared to their old passwords. In contrast with the previously described graph, in this figure, red dots above the green line indicate that a participant's breached password changes resulted in lower quality passwords, i.e., more reuse. More than half of the participants changed their passwords on a breached domain to be more reused across their other accounts than their old breached domain password, compared to their average changes in password reuse. Figure 4.4 shows the average amount of reuse of all of each participant's unique passwords entered per domain, computed as described in Section 4.2.3.

Overall, password changes showed relatively similar changes in strength and reuse, regardless of whether they were on breached domains; however, breach-related password changes resulted in more dissimilar new passwords.

4.4 Limitations

Although our work provides valuable insights into the effectiveness of post-breach regulations through actions people take after password breaches, it is subject to a few limitations, including those due to the nature of the data collection.

The participants whose behavior we study are not representative of the larger population; for example, a quarter had jobs that involve programming and many were students. Hence, we make no claims with respect to generalizability. We also did not have data about the relative importance of each breach to the data subjects. However, for the 49 participants with Yahoo! accounts, we observed (by examining their web browsing history) that almost a fourth visited a Yahoo! mail page multiple times a day and another fourth visited such a page at least once every four days. This suggests that a large fraction of these participants were using their Yahoo! passwords to protect email accounts, and hence they should have been concerned about the breach.

We do not have data about whether participants were explicitly notified about a breach. Furthermore, we could not determine whether password changes were made directly in response to a breach; rather, we study changes within a window of time after a public breach announcement.

Our analysis of passwords was limited in its precision because passwords were represented by the hashes of three-character and longer substrings instead of in plaintext. This type of infor-

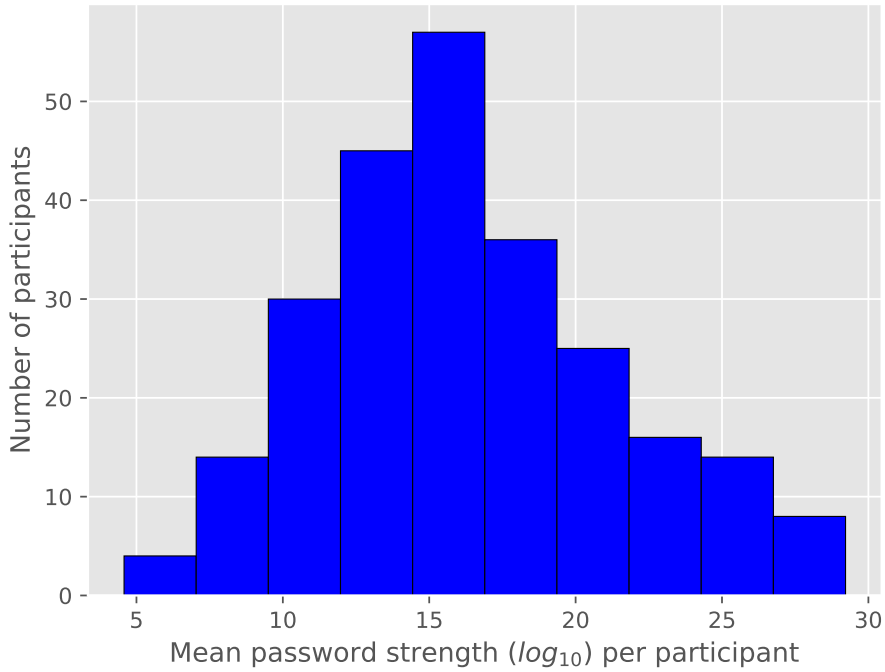


Figure 4.3: The average strength of all of each participant's unique passwords entered per domain.

mation about passwords has been used previously to study password reuse [174] and is sufficient to reveal substantial reuse in our application.

As in Chapter 3, the data we analyzed was collected from Windows computer users and limited to passwords entered on Google Chrome and Mozilla Firefox. Users of non-Windows operating systems may exhibit behaviors different than the participants in our dataset. The participants whose password data we analyzed used Internet Explorer (IE) on average for only 2.86% of all their browsing and largely to visit websites that would likely not require them to log in. Given that IE usage was low among the participants in our dataset and that Windows is the dominant OS for personal computers [61], we do not believe that the unavailability of data about people using non-Windows machines and of password data from other browsers is likely to fundamentally affect our findings.

Again, as in the previous chapter, the participants enrolled in the SBO study may have been biased towards less privacy- and security-aware people.

4.5 Discussion and Conclusions

Out of 63 participants with an account on a breached account, only 21 changed a password on the breached domain, and only eight did so within three months. Participants on average had 30 passwords similar to their password on the breached domain, but on average changed only

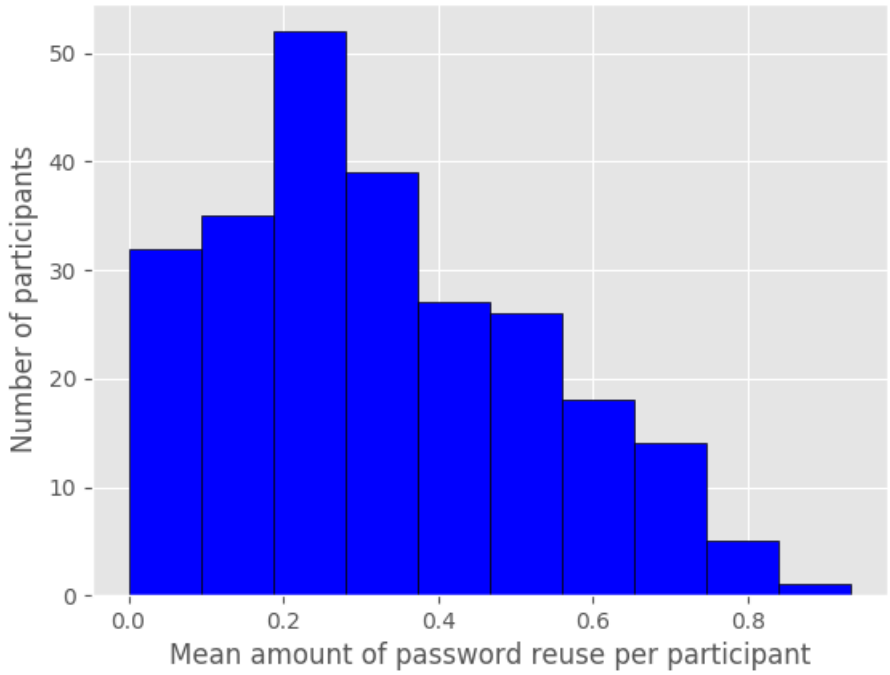


Figure 4.4: The average amount of reuse of all of each participant's unique passwords entered per domain.

four of these within a month after changing their password on the breached domain. Even when they changed their password on a breached domain, most participants changed them to *weaker* or *equally strong* passwords. And, regardless of whether participants changed their similar passwords within a month of the first change, their new passwords on the breached domains were on average *more* similar to their remaining passwords.

Some facets of good password maintenance behavior may be difficult for an average user to grasp [42, 111, 114, 218, 224]. For instance, the affinity towards changing to weaker or equal-strength passwords could be because when people feel compelled to choose new passwords they have poor awareness of password strength or the additional memory burden leads them to pick weaker passwords [83, 218]; for example, they might change just enough characters to satisfy system requirements. Related to partial password reuse, people may find it difficult to understand how their "different" password is still similar to other passwords, i.e., they might be unintentionally partially reusing passwords. Potential mitigating efforts could be to integrate password-reuse trackers within tools that people may already use and trust to store their passwords. Some password managers, such as 1Password, already warn users if one of their saved passwords is reused. Password managers, including those built into web browsers, could go further and more actively discourage password reuse.

Overall, our findings suggest that password breach notifications are failing dramatically, both at causing users to take action and at causing users to take *constructive* action. Regulators should take note of the ineffectiveness or absence of breach notifications and impose requirements on

companies to implement better practices [106, 219, 225, 234, 236]. In particular, they should encourage companies to send repeat notifications until they have positive confirmation that the notifications have been understood and that any instructions have been followed. Regulators should also require that companies force password resets after a breach and provide actionable instructions on how to create "strong" passwords, describe the risks of password reuse, and strongly suggest to users that they change passwords beyond the affected domain. From a preventative standpoint, regulators could incentivize companies to use an authentication method other than passwords or to require their users to use two-factor authentication. Companies should also be required to hash and salt their passwords to avoid credential-stuffing and rainbow-table attacks on plaintext or weakly hashed passwords [170, 216]. Regulators could also require services to subscribe to HIBP and to force users to change their passwords when they encounter a matching hash.

Chapter 5

"Adulthood is trying each of the same six passwords that you use for everything": The scarcity and ambiguity of security advice on social media

5.1 Introduction

Chapters 3 and 4 highlighted low security awareness of and engagement with security incidents. In this chapter, we start to explore how social media can be used to better spread security awareness and encourage healthy security behavior.

Prior work has showed that social media are some of the most prevalent channels for discussing security and privacy [85] and that in experimental settings, they are effective platforms through which friends can be encouraged to adopt security-enhancing behavior [77, 78, 81]. Inspired by these findings and the prevalent use of social media among adults [180], we explore the promising idea of improving user security through social networks. In particular, in this chapter we examine, using a dataset of real-world Facebook and Twitter behavior, how prevalent such sharing may be in practice and what discussions about security and privacy may typically convey (e.g., educational discussions of security and privacy). More specifically, we focus on two research questions: (1) how common is sharing and consuming security and privacy content on social media?; and (2) what are the different ways in which security and privacy are talked about in social media posts? To answer these questions, we study the Facebook and Twitter logs of 38 participants who were enrolled in the Security Behavior Observatory, as in the previous two chapters. We chose to collect and analyze the data of SBO participants since we were able to instrument their browser extensions to collect non-public Facebook data, which is notoriously difficult to collect otherwise, and since we could additionally analyze their security behavior.

Perhaps surprisingly, we found that interactions (e.g., liking, commenting, sharing, retweeting) with security and privacy content were very scarce; only 131 of the 194,081 Facebook posts participants interacted with, and 44 of the 6,883 Twitter posts, were related to security and privacy (0.09% in total). In fact, of the 38 participants the majority (74%) did not interact with

any security and privacy content at all. We further found that the amount of interactions with security- and privacy-related posts was not correlated with any demographics or with technical savviness. To validate the above findings with an additional dataset, we constructed a dataset of 15,053 tweets made by 100 random Twitter users and identified security and privacy posts within this set. We found that only 0.08% of the posts were related to security and privacy—just barely fewer than in our main dataset.

We examined the few security and privacy posts further to understand what they were trying to convey, to understand, for example, whether the posts were educational in nature and could help encourage healthy security behavior. We examined these posts through a thematic analysis [63] and identified five themes in what the posts about security and privacy conveyed. One of the themes described posts that presented obviously constructive security and advice. Posts with this theme may be the most clearly helpful in encouraging desirable security practices, but only 10 participants interacted with such posts and not as often as with posts with the other themes. A more prominent theme described posts that talked about security and privacy but were either ambiguous about whether they were promoting desirable or undesirable practices or recounted anecdotes or jokes about security and privacy. Other themes included brief mentions of security and privacy topics without any substance (which 16 participants interacted with), encouragement or demonstrations of detrimental security and privacy practices (seven participants), and providing information about public policy topics related to security and privacy (six participants).

Our results showed that, for participants in our sample, security and privacy were not topics that people frequently interacted with or were exposed to on social media. Given that our sample was slightly skewed towards technically-savvy and educated people (see Section 5.2.2), we suspect that the general population may have even less exposure than we observed, as hinted at by the even smaller amount of security- and privacy-related posts we found in the random Twitter sample. Even when posts referenced security and privacy, rarely was this to deliver constructive security advice or recount secure behavior, with posts more often providing no useful commentary on security practices. As a result, participants who interacted with the security and privacy posts we analyzed may have had higher security awareness, but were not likely to have taken away actionable advice or adopted security-enhancing behavior. The latter is supported by an exploratory investigation we performed into the relationship between people's security behaviors and the posts they interacted with. While prior work showed through interventions, that demonstrating healthy security practices on social media can correlate with increased adoption of healthy security behavior, our findings suggest that without interventions, the amount of constructive security advice being shared could be too low to effectively encourage healthier security behavior. Achieving more widespread adoption of security best practices could potentially be helped by two steps: increasing the spread of security and privacy advice using wide-reaching channels such as social media; and presenting the advice in such a way that user burden is reduced and such that it incites changes in their security behavior.

We first describe the dataset we analyze in this chapter. We next discuss our methodology for quantifying the amount of security and privacy posts in our social media datasets and the corresponding results (Section 5.2). We then describe our thematic analysis of security and privacy posts (Section 5.3) and exploratory statistical analyses studying security behavior in relation to social media interactions (Section 5.4). We conclude with limitations of our study (Section 5.6) and a discussion of the implications of our work (Section 5.7).

5.2 Data collection and dataset

5.2.1 Data collection

As in Chapters 3 and 4, we use data collected by the Security Behavior Observatory (SBO) [97, 98]. The SBO started recruiting participants and collecting data in October 2014 and ceased data collection in July 2019; we used data collected upto April 2018.

We used data collected by the Security Behavior Observatory (SBO) project. The SBO is a longitudinal study of the usage patterns and security behaviors of Windows computer users [97, 98]. The SBO started recruiting participants and collecting data in October 2014 and ceased data collection in July 2019; we used data collected upto April 2018.

Collecting FB and Twitter data We collected additional social media data from a subset of consenting SBO participants through another study approved by Carnegie Mellon University's review bard. Specifically, we collected Facebook and Twitter data between October 2014 and April 2018. For Facebook, we collected the content on the "Activity Log" page (a log of all activity made by a Facebook user) and the list of public pages they follow. For Twitter, we collected participants' tweets (including retweets and replies to tweets), the tweets they've favorited, the Twitter accounts they follow, and the Twitter accounts that follow each participant. The logs we collected covered *all* historical social media activity irrespective of the device on which participants may have used the application. Participants who consented to this additional study received $15 compensation in the form of an Amazon gift card.

We created developer applications for both Facebook and Twitter [8, 10] and asked participants to log into our applications through a webpage presented after the consent page to our additional study. Participants had the option to provide data for one or both platforms. We collected a participant's Facebook data as follows: using the participant's login credentials, we fetched their name using the Facebook API [8]. We then exclusively stored the MD5 hash of the name for subsequent use, ensuring that we were never directly working with the participant's identifying information. Since the API does not provide functionality to retrieve the activity log of a user, the SBO browser extension was instrumented to trigger data collection when the participant logged into Facebook and visited the Facebook homepage from their SBO computer. Before starting to collect data, the extension ensured that the hash of the name on the visited Facebook homepage matched the hash of the name fetched by the API. When the above criteria were met, the extension loaded the participant's activity log or page likes webpage in an invisible browser tab, scrolled through the page, and downloaded the contents. If data collection was halted due to logging out, it was resumed the next time the participant logged in to Facebook. Some posts listed in the activity log had content that was not displayed in the log we downloaded. For each of these posts, we attempted to visit the link of the post during a second round of data collection to collect the full post text after the initial log was collected by the extension. The contents of these individual posts were collected in a similar way to the log, by visiting the post link in the background and downloading their contents[1]. For Twitter, we used the credential

[1]We were not able to collect the full content of posts for all users or for all of their activity because not all participants visited Facebook frequently enough after the first round of data collection for the second round to complete.

tokens provided when participants logged into our Twitter app and used the Twitter API [10] to fetch the above-mentioned Twitter data for each participant.

5.2.2 Dataset

Our study is based on the longitudinal data collected by the browser extensions and the additional collected social media data for 38 participants. Specifically, 34 participants provided us with Facebook data, 16 with Twitter data, and 12 with both Facebook and Twitter data. The participants who provided Facebook data interacted with 5708 Facebook posts on average over the four-year period. Participants who provided Twitter data interacted with 601 tweets on average. We use the following datasets in this chapter.

Facebook and Twitter data: This dataset contains information about every post or tweet participants interacted with and the public pages or Twitter users that the participants followed.

Browsing history: This dataset contains participants' browsing history in Google Chrome, Mozilla Firefox, and Internet Explorer. The dataset contains information about every URL visited in the web browser, along with page titles and timestamps.

Password data: This dataset includes information about every entry made into a password field in a web page, as determined by a browser extension, including: a salted one-way hash of the password and the URL of the form in which the password was submitted. We filtered this dataset to exclude passwords used during failed login attempts or entered by a user other than the main computer using the approach from Section 3.2.1.

Installed software update history: This dataset contains information about all events related to installed software and software updates on participants' computers along with Windows updates. Data about one event includes information about whether it corresponds to an update or new software installed, the version of the software, and the timestamp for when the software was installed or update was executed.

File system data: This dataset contains information about all the files present on participants' filesystems. Specifically, data about one file includes the file name and path, the hash of the file in both MD5 and SHA1, and the timestamp for when the file was created.

Operating system history: This dataset contains the history of the different Windows operating system versions (e.g., "Vista", "7", "8", "10") participants have used throughout the duration of the SBO study. Data about one version includes the operating system version and the timestamp at which that specific version was installed.

Software update settings history: This dataset consists of the Windows software update settings participants set for their computers at various points in time. The settings are a combination of four different preferences: (1) whether Microsoft updates are enabled; (2) whether recommended updates are enabled; (3) whether the update service is enabled; and (4) when notifications are desired. Data about one settings combination at a point in time contains the answers to the above four questions as well as the timestamp that particular settings combination was set.

WiFi connection history: This dataset contains information about all the WiFi networks participants connected to from their SBO-instrumented computer. Data about one instance of connecting to a WiFi network includes the name of the WiFi profile, whether the network required authentication, the type of encryption used by that network, and the type of shared key authentication used.

Browser content settings: This dataset contains information about what content participants allow various websites to use within their browser, i.e., microphone, location, or camera permissions. These permissions may be set or modified for a website at various points in time. Data about one event corresponding to setting or modifying permissions for a website includes which of the above permissions the participants granted to that website as well as the timestamp for when those particular permissions were set.

Participants' ages ranged from 21 to 81 years with a mean age of 33. Participants were female-skewed (74%). A little more than half (53%) knew at least one programming language and 18% had programming as their primary profession. 61% were students and 53% had a Bachelor's degree or higher.

5.3 How common were security and privacy discussions on social media?

Our first research question examines how often participants interacted with content that talked about digital security and privacy on Facebook and Twitter.

We compiled a set of all the Facebook and Twitter posts the participants interacted with during the timeframe discussed in Section 5.2; this included all the Facebook posts for the 34 Facebook users and Twitter posts for the 16 Twitter users. For a given interaction with a post on Facebook or Twitter, we then extracted and concatenated all pieces of text corresponding to the components of the post. For example, if a post was being re-shared, we considered both the original text and the shared text. Similarly, if a post was commented on, we considered the original post and the comment text. We identified posts related to digital security and privacy posts by examining this text for each post.

We iteratively categorized posts as related to digital security and privacy as follows. For security and privacy posts, two researchers who are domain experts in security and privacy initially created a list of regular expressions (regexes) by reviewing several posts and by brainstorming for regexes related to digital security and privacy (see Table 5.1 for this initial list). We then

systematically built a list of regexes, starting with the aforementioned initial list, by iterating through the following steps:

1. Match the regexes in the list to the set of all collected posts.
2. Look through each post matched by any of the regexes. Identify strings within the matched posts that are relevant to security and privacy, but which do not yet have corresponding regexes in our list.
3. Manually examine each post in a random sample of one hundred of the unmatched posts. Identify posts that should be classified as security- or privacy-related and strings relevant to security and privacy within each post.
4. Construct regexes for each of the new strings from the previous two steps and add them to the list of regexes.
5. Repeat steps 1–4 until steps 2 and 3 yield no new strings from which regexes could be created.

After constructing the final list of regexes, we flagged all the posts that matched any of the regexes. To ensure the absence of false positives, we manually examined all the flagged posts and unflagged them if they appeared not to be related to security and privacy. We verified we did not incur false negatives by inspecting random samples of 100 unmatched posts at each iteration, and once at the end before stopping which showed no false negatives. After we stopped the iterations, we further verified that we likely did not miss relevant posts by examining a random set of 1000 unmatched posts and found consistent results. The final list of regular expressions for identifying security and privacy posts is in Table 5.1.

We followed a similar process to identify posts related to technology in general, as well as posts specifically related to data breaches, with different initial lists for each category (see Table 5.1).

We used the above approach after exploring a few possibilities. Originally, one approach was to construct a topic model based on the text of individual, uncategorized posts. However, the resulting topic models did not produce coherent topics. Another approach we considered was to manually divide a set of posts into posts related and unrelated to security and privacy according to some criteria; then build a topic model over posts in these two categories; and then use the topic model to categorize additional posts. However, it was difficult to obtain a sound initial categorization without an approach like the iterative approach described above.

We found 131 of the 194,081 Facebook posts (0.07%) and 44 of the 6,883 Twitter posts (0.6%) in our dataset to be related to security and privacy. To confirm that this proportion was not an artifact of our dataset, we compared it to the proportion of security and privacy content in a dataset of 100 tweets made by each of 100 randomly selected Twitter users from a public dataset [64]; the resulting dataset contained 15,041 tweets, as not all Twitter users had 100 tweets to collect. Matching our final list of regexes against the set of random tweets, 0.08% of posts were flagged as related to security and privacy. To ensure the absence of false negatives as a result of using the list of regular expressions for our collected data on the separate random Tweets dataset, we sampled 100 random unflagged tweets of the 15,041 posts and found that they were correctly categorized as not related to security and privacy.

Based on the results for two different datasets, we find that security- and privacy- content was

scarcely interacted with on social media by the participants.

5.4 How did posts actually talk about security and privacy?

In Section 5.3 we reported on the low occurrence of interactions with security- and privacy-related content in our set of Facebook and Twitter posts. In this section, we dive deeper into the posts themselves by conducting a thematic analysis of the Facebook and Twitter posts. The purpose of this analysis is to characterize the types of security-related content people may come across as they're browsing social media. While prior work has studied how helpful security and privacy information is presented across the web [184, 192], in this section we study what information people may actually be exposed to which may go beyond only helpful content. In particular, we identify themes of how the posts discuss security and privacy and what they convey.

5.4.1 Methodology

As the first step of the thematic analysis, two researchers familiarized themselves with every security- and privacy-related Facebook and Twitter post in our dataset. If a link was present in the post, we considered the link's content to be part of the post. Using an inductive approach, one of the researchers conducted a round of open-coding of each post [51]. Axial coding was then used to derive higher-level codes or categories related to what the post appeared to convey [130] (e.g., "sarcasm about security advice," "demonstrating a constructive action," "story about bad experience"). They then coded each post according to these derived codes. Following this, the researcher identified a set of overarching themes across the identified higher-level categories. Through frequent discussions among the research group, the researchers iterated on the categories and the resulting themes to ensure that every post could be described by at least one of the themes. Because we conducted a thematic analysis, we did not compute the inter-rater reliability between two coders, a decision that is supported by prior work [35, 41, 157].

5.4.2 Results

We identified five major themes in how the Facebook and Twitter posts talked about security and privacy.

Quick mentions of security or privacy

Some posts mentioned a security or privacy topic (e.g., bitcoin, passwords, net neutrality) but did not elaborate on that topic. Such posts also included announcements about something related to security and privacy but without details of the actual phenomenon (e.g., a post that announces a cybersecurity article or paper but does not describe the topic). The following are examples of posts that exhibit this theme.

> *Overseen: Girl from The Ring supports data privacy*

> *bitcoin is cool but have you guys heard of kohl's cash*

Security and privacy	Technology		Data breaches
password	matlab	password	breach
social security number	airport.*security	software development	captcha
security camera	net.*neutrality	web design	data.*hack
key.*security	gprs	data.analy	data.*stolen
security.*account	[0-9]mb	data.economy	data.*compromise
security question	Java	coinbase	data.*breach
phishing	3g(\W—$)	JSFoo	
cybersecurity	computer	uptime	
cyber.security	Android [0-9]	plugin	
de-verify	internet.*service	motherboard	
security.*protocol	apple.*ios	autonomous	
privacy	ios.*apple	browser.*code	
bitcoin	ios.*android	frontend	
net.*neutral	android.*ios	webapp	
secure	github	jschannel	
jailbreak	wireless	\.js	
jail.*break	browser extension	3d model	
comp.*virus	api	apple.*invent	
hack	bitcoin	chat.*dm	
	Nokia	Dell.*laptop	
	smart device	screens.*digital	
	hyperloop	game.*app	
	browser.*bug	(\W)Siri(\W)	
	prototype	programming	
	smartphone	matlab	
	jquery	programmer	
	callback	sensor	
	app.*service	robot	
	js.*app	algorithm	
	lg.*screen	python	
	c\+\+	arduino	
	ux.*ui	autolab	
	ui.*ux	package	
	js_channel	linux	
	node.*js		
	debug		

Table 5.1: Lists of regexes used to flag Facebook or Twitter posts within three categories. Initial regexes are in black while the regexes in red were added via the iterative process.

Atoms, stars, the solar system, cyber security, creativity - just a few of the amazing things you can learn about in your spare time.

In all three posts, terms related to security and privacy are mentioned (i.e., "data privacy," "bitcoin," and "cyber security"). However, these terms do not describe the main topic of the post.

Of the 34 Facebook users and 16 Twitter users, 14 and five respectively interacted with posts corresponding to this theme; a total of 16 participants interacted with posts in this theme across both Facebook and Twitter. When reporting on the "total" for the subsequent themes, we mean the total over the union of the Facebook and Twitter users.

Ambiguity in the intended message

Some posts talked about security and privacy in more detail but did not convey a clear message of encouraging or discouraging certain security behaviors. These posts were sometimes sarcastic, included anecdotes about security experiences, or joked about security topics.

Right to Privacy. LOL.

This post talks about the right to privacy but follows that mention up with "LOL" (the acronym for "laugh out loud") which indicates that the poster thought the idea of right to privacy is funny. It is unclear whether they were trying to convey something constructive or otherwise with this post.

you ever just stare at the security questions options when setting up new accounts online and think "i really don't know myself"

Here, the poster talks about security questions used to access online accounts. However, it appears to be joking about this concept without conveying their implications for account security.

The mandatory computer security training I have to do for work just advised me to "only install mobile apps that are absolutely necessary". Also apparently hackers can kidnap your child, steal your laptop to sell on the black market (specifically the black market), and also take your job. Those tricksy hackers.

This post appears to express sarcasm about recommended security behavior. While the poster does speak about the advice to only install necessary apps, the second hyperbolic sentence suggests that they may be ridiculing the advice.

A total of nine participants interacted with posts in this theme, encompassing eight Facebook users and three Twitter users.

Constructive security or privacy advice

A less frequent type of post encouraged constructive security practices directly or indirectly; for example, by sharing an anecdote about how certain security practices helped, criticizing poor security practices, or explicitly delivering constructive advice.

Hi everyone. If you received an email from me. Don't open it, good ole gmail was hacked. # thanksgmail

In this post, the poster tells their friends to not trust and open emails sent by their hacked account. This advice is constructive to making sure their friends are not victims of the hacker given the specific situation.

> *If it said you posted it, you may have gotten "phished" and need to change your password.*

In the above post, the poster suggests how to tell if someone's account was compromised as a result of phishing and describes the recommended remediation to protect their account.

The following post again explicitly provides constructive advice for healthy security behavior.

> *MINER ALERT: Should anyone see a AD here on FB for free downloadable photo enhancer called InPixio, do not...I repeat.. DO NOT download as it is harmful to your computer, my Internet Security scanner raised red flag that it was automatically removed on account of being not just virus related but hacks your computer as well*

In total, 10 participants interacted with constructive posts in this theme, including nine Facebook users and two Twitter users.

Demonstration of detrimental security practices

Posts can often demonstrate undesirable security practices (e.g., by asking for advice to help execute these practices, sharing passwords inside posts, or speaking positively about unsafe practices).

> *The wifi password is probably puravida Thank you Costa Rica...*

> *I want WomanlyAlways1895 on my gravestone Lol does anyone know what the for sisters password on the chi o wix page is*

In the above two posts, the posters mention or ask for a password in a post. Whether the password information was intended to be public or whether it was posted to a private group, the post demonstrates a bad practice of posting or asking for passwords online which may encourage the same behavior in others.

A total of seven participants interacted with these negatively constructive posts, all of whom were Facebook users.

Information or advice about public policy topics

The final theme describes a different flavor of posts that are less about the technical details of security and privacy but provide information on security and privacy topics in the public policy sphere. For example, posts that provide information on net neutrality and its implications or links to events or talks about the intersection of public policy and security or privacy would be described by this theme. The following are two examples of this kind of post:

The basic principle of Net Neutrality is that access to all websites should be treated equally. What the FCC wants to do is empower broadband service provides to distort the online marketplace and set up a pay-for-play system. This would be a terrible mistake.

The FCC just announced its plan to slash net neutrality rules, allowing ISPs to block apps, slow websites, and charge fees to control what you see and do online. They vote December 14th. Call your representatives today to tell them to fight for net neutrality! Learn how to do that at http://battleforthenet.com

Posts in this theme occurred the least frequently among participants with only four Facebook users, four Twitter users, and six participants in total interacting with such posts.

Out of the posts related to security and privacy we analyzed, we could consider that the kinds of posts that may actually trigger changes in security behavior or habits are the posts in the third theme ("Constructive security or privacy advice"). However, only 10 participants interacted with those posts and such posts occurred infrequently (20% of the 175 security and privacy posts and 0.02% of all social media posts in our dataset). With the low number of security and privacy posts we found overall in Section 5.3 and the even scarcer amount of constructive advice in these posts, we hypothesized that though these participants' security behavior may be affected by social influence in a social network in experimental settings [77], the discussions we observed about security and privacy on Facebook and Twitter may not have been prevalent enough to trigger such a correlation in the wild.

We test this hypothesis through an exploratory statistical analysis in the next section.

5.5 Exploring the relationship between consuming security content and security behavior

We observed a low amount of constructive security and privacy posts in Section 5.4. Therefore, although prior work has shown that in an experimental setting, social influence within a social network plays a role in adoption of security-enhancing behavior [76, 77, 81], we hypothesized that the amount of security and privacy content people may interact with on Facebook and Twitter may not be enough to encourage improved security behavior. We test the hypothesis in this chapter; as a proxy for exposure to content, we measure participants' interactions with posts. For example, we study the relationship between people's interactions with security- and privacy-related posts (as identified and analyzed in Sections 5.3 and 5.4) on social media and their measured security behavior.

5.5.1 Methodology

To study the aforementioned relationship, we used behaviors describing interactions with social media content on Facebook or Twitter and participant demographics (age, gender, whether they're a student, and whether they knew a programming language) as input variables to statistical analyses. The outcomes we studied were behaviors indicating better or worse security practices.

The behaviors we were interested in modeling typically were not directly part of the raw data we collected: to compute them, we processed the SBO system-level and social media data such that each participant was associated with all events pertaining to them from each dataset described in Section 5.2. We distilled and augmented this data into features that captured participants' interactions with content on social media and measures of their security behavior.

Social media features

We categorized the type of interactions participants had with Facebook posts as: (1) liking a post; (2) commenting or replying to a comment on a post; (3) saving a post; and (4) sharing a post (either on the Facebook user's own timeline or someone else's timeline). We consider the first three types of interactions to be consumption of content and the last one to be sharing content. We categorized Twitter interactions as follows: (1) liking a tweet; (2) replying to a tweet; and (3) sharing a tweet (including retweets). Similarly, the first two interactions are considered to be consuming content. If a post contained a link, we considered the content of the link as part of the content of the post whenever the link content was available. We did this because we observed instances of people posting links to articles with substantial information but the post text itself was minimal. When a post was re-shared or re-tweeted, commented on, or replied to, we stored information about each individual poster when the data was available (e.g., whether the poster was the participant themselves, a friend, or in the case of Facebook, a page). For example, if the participant re-shared a post on Facebook, we stored information about the main poster (the user who re-shared) and the original poster. On Twitter, if a participant re-tweeted a tweet that retweeted another tweet with a quote, we fetched and stored information about the main poster and the two other posters involved, up to five posters overall. For Facebook posts originally posted or reshared by a public Facebook page and all posts on Twitter, we also considered the popularity of the Facebook page or Twitter user, i.e., its number of followers. Therefore, we collected additional information about each social media post as follows: when a link was present inside the post's text, we downloaded the content of the link using the newspaper library [172] and replaced the link in the post text with the downloaded text. For Facebook posts posted by a page, we scraped the page's URL and extracted and stored the follower count; Twitter follower counts were already included in our collected data.

Based on research showing the effectiveness of social influence by friends or known people [76, 77, 78, 81, 88, 190, 235], we also hypothesized that advice from a familiar poster may be correlated with uptake of constructive security behavior. We consider the familiarity of a poster specific to a post to be the number of previous posts in which the participant interacted with that poster. The familiarity of one poster will not be the same for all posts involving that poster and takes into account only the posts made by the poster prior to the post in question.

For each participant, we created separate features for Facebook and Twitter since the way content is shared on each platform differs. These features span interactions with the following three topics whose posts we identified in Section 5.3: security and privacy, technology, and data breach. We designated three features to describe the total number of interactions with posts in each category on Facebook and three more to describe this number for Twitter. Prior work has found that the source of security and advice plays a role in people's inclination to act upon it, often heeding the advice of experts or influential people, or friends and family [78, 88, 190,

235]. Therefore, we describe interactions with posts along dimensions related to the source of a social media post such as: the type of poster (friend or page), the familiarity of the poster, and the poster's popularity. Common to both platforms, for posts in each category, other features describe: the number of posts that were consumed versus shared by the participant, the number of posts where the participant was the sole author of the post they were interacting with, and the average familiarity of the posters for each post participants interacted with. Features specific to Facebook included: the average number of followers of the public page posters, the number of posts either originally posted or re-shared by a regular Facebook user (likely a friend), and the number of posts either originally posted or re-shared by a public page. Features specific to Twitter included: the average number of followers of all the posters of posts interacted with, the average number of favorites of all posts interacted with, and the average number of retweets of all posts interacted with. A list of all features and their descriptions can be found in Table 5.2. If a participant did not have either a Facebook or Twitter account (including those who had no social media account or visits), their corresponding feature values were set to 0.

Security behavior indicators

Each participant was also associated with outcomes representing their security behavior (as introduced in Section 5.2). Inspired by and building upon previous work [52, 67, 98, 174, 222], indicators of security behaviors span each dataset in Section 5.2. For example, in the realm of browsing behavior, visiting malicious web sites or web sites rarely visited by others may be signs of poor computer security hygiene. Therefore, indicators based on browsing behavior describe: the number of webpages visited related to security and privacy, the number of malicious websites visited, the number of times security or privacy errors were prompted by the browser and the number of times they were ignored, and the number of visits to less popular websites or websites rarely visited by others. Google Safe Browsing and VirusTotal have been extensively used in prior work as services that report whether a URL has been flagged as malicious [50, 145, 152, 176, 200]. Therefore, for every URL event in the browsing dataset, we augmented the data about each webpage visit with the reported results from VirusTotal [11] and Google Safe Browsing (GSB) [9] when queried with the URL. As in Chapter 3, we again use Alexa Web Information Services (AWIS) [6]—a service that provides information about a website and about its popularity or traffic (e.g., the category of the website, the number of URLs that link to the website, the website's Alexa rank)—to compute browsing behavior features. We incorporated the reported results from AWIS when queried with the URL into the features related to browsing behavior.

Features related to installed software and updates included: the number of times participants updated software, the number of times participants computers underwent security updates, and the number of antivirus software installed. Features related to operating system (OS) updates described the number of OS updates and the number of times the OS update settings were modified. Other features included the average strength across each domain's latest password, the number of open Wi-Fi networks participants connected to, the number of websites to which participants granted either microphone, camera, or location permissions, and the number of files in participants' filesystems flagged by VirusTotal as malicious. A list of all security behavior indicators along with their descriptions can be found in Table 5.3.

feature_set	*feature_name*	*description*
Facebook	has_fb	Whether the participant had a Facebook account
	fb_num_{category}	Number of posts a participant interacted with that fell into {category}
	fb_num_{category}_consumed	Number of posts in {category} wherein the interaction involved consuming content
	fb_num_{category}_by_friend	Number of posts in {category} wherein at least one of the posters involved in the post was a friend
	fb_num_{category}_by_page	Number of posts in {category} wherein at least one of the posters involved in the post was a public page
	fb_num_{category}_was_poster	Number of posts in {category} wherein the Facebook user themself was the sole author of the post
	fb_{category}_pages_avg_popularity	Average number of followers of the public pages involved in posts in {category}
	fb_{category}_posters_avg_familiarity	Average familiarity of each poster involved in all posts in {category}
Twitter	has_twitter	Whether the participant had a Twitter account
	twitter_num_{category}	Number of posts a participant interacted with that fell into {category}
	twitter_num_{category}_consumed	Number of posts in {category} wherein the interaction involved consuming content
	twitter_num_{category}_was_poster	Number of posts in {category} wherein the Twitter user themself was the sole author of the post
	twitter_{category}_posters_avg_popularity	Average number of followers of the posters involved in posts in {category}
	twitter_{category}_posts_avg_favorites	Average number of favorites on all posts in {category}
	twitter_{category}_posts_avg_retweets	Average number of favorites on all posts in {category}
	twitter_{category}_posts_avg_familiarity	Average familiarity of each poster involved in all posts in {category}

Table 5.2: Features related to social media consumption describe posts in each of the following categories: sec_priv, tech, and breach. Features with "{category}" are repeated for each of the three categories.

feature_set	*feature_name*	*description*
Browsing data	browsing_num_sp_related_pages	Number of visits to webpages related to privacy policies or security settings
	browsing_num_ sp_errors	Number of visits to webpages that signaled a privacy or security error
	browsing_num_sp_ignored_errors	Number of times the above security or privacy errors were ignored by detecting if the participant continued to the page after seeing the error
	browsing_num_vt_domain_ip	Number of visits to webpages on domains flagged by VirusTotal's domain report and IP address report APIs
	browsing_num_vt_domain_as_url	Number of visits to webpages on domains flagged by VirusTotal's URL report API
	browsing_num_vt_url	Number of visits to webpages flagged by VirusTotal's URL report API
	browsing_num_gsb	Number of visits to webpages flagged by Google Safe Browsing
	browsing_num_uncommon_tlds	Number of visits to webpages under uncommon Top Level Domains (TLD) [2]
	browsing_num_private	Number of visits to webpages in private browsing mode
	browsing_avg_links_in_count	For the webpages visited, the average number of links to each page from other webpages as reported by AWIS
	browsing_avg_website_ranks	For the webpages visited, the average Alexa rank as reported by AWIS
	browsing_num_no_links_in_count	The number of webpages visited that did not have any links to it from other webpages as reported by AWIS
	browsing_no_rank	The number of webpages visited that did not have an Alexa rank as reported by AWIS
Updates	software_num_updates	The number of times installed software was updated
	software_num_antivirus	The number of antivirus programs installed on a participant's operating system [3]
	software_num_sec_updates	The number of Windows security updates executed
OS	os_num_updated	Number of times the operating system version was updated
	os_num_updatesettings_changed	Number of times the operating system update settings were modified

[2]Common TLDs were determined from https://www.lifewire.com/most-common-tlds-internet-domain-extensions-817511.

[3]The list of antivirus software we checked for can be found at https://support.microsoft.com/en-us/help/18900/consumer-antivirus-software-providers-for-windows#avtabs=win7.

Filesystem	`fs_num_vt`	The number of files on the filesystem whose hash was flagged by VirusTotal's file report API
Passwords	`pwds_avg_strength`	The average strength of the latest passwords being used on each domain
WiFi profiles	`wifi_num_open`	The number of times a participant connected with an open WiFi network that did not require authentication
Browser content	`content_num_camera`	Number of URLs for which the participant granted camera access
	`content_num_location`	Number of URLs for which the participant granted location access
	`content_num_microphone`	Number of URLs for which the participant granted microphone access
	`content_num_all_allowed`	Number of URLs for which the participant granted camera, location, and microphone access

Table 5.3: Security behavior indicators across browsing and system-level datasets.

In total, participants were represented by 44 features related to social media including interactions with security and privacy, technical, and breach content on Facebook and Twitter and 25 indicators related to security behavior. To reduce the number of features to analyze, we grouped highly correlated features together through factor analysis [165] on the social media features and the security behavior features separately. As a result, 44 social media features were reduced to six factors and 25 security behavior indicators were reduced to two factors as described in Section 5.5.2.

We then computed statistical relationships between the factors describing social media consumption of content in the three categories and the factors describing the various security behaviors.

5.5.2 Results

Factor analysis

After collapsing highly correlated social media features into one feature and transforming the values for each feature into its Z-score, factor analysis yielded eight total factors based on a scree plot showing that eight factors had eigenvalues above one [199].

We considered six of those factors that had at least two features with an absolute factor loading greater than 0.7 [82, 182, 201]. The remaining two factors did not have a sufficient number of factor loadings greater than 0.7 and were therefore, excluded from consideration. Based on the features with loadings above 0.7 in each factor, the factors described the number of: (1) interactions overall with technical posts on Twitter and those made specifically by familiar posters; (2) interactions with security- and privacy-related posts on Facebook made by familiar posters and the number of technical posts the user made on Twitter; (3) interactions with security- and privacy-related posts and breach-related posts on Facebook; (4) interactions overall with technical posts on Facebook and specifically, the number of technical posts the user themselves made;

(5) security- and privacy-related posts and breach-related posts made by the user themselves; and (6) interactions with technical and security- and privacy-related posts made by familiar posters on Facebook.

Applying the same criteria as above, factor analysis of security behavior indicators (each transformed to their Z-scores) yielded eight factors of which we considered two.Again, based on the features in each factor with loadings above 0.7, the factors described: (1) the number of visits to websites flagged by VirusTotal or GSB and the number of links from other websites to the websites visited; and (2) the frequency of visits to URLs with uncommon TLDs and the number of antivirus software installed on participants' operating systems.

Analyzing relationships

Due to the size of our sample and the distribution of the data about participants, similarly to in Chapter 3, we constructed a non-parametric linear quantile regression model of the relationship between all six social media-related input factors in addition to the four demographic features, and each of the two outcome factors [132]. We computed each of the two quantile regression models for the 25th, 50th (median), 75th, and 90th percentiles [62], resulting in a total of eight regression models (Tables 1, 2, 3, 4, 5, 6, 7, and 8 in Appendix D). We did not find any of the social media factors to be consistently correlated with either of the security behavior factors. In particular, the models for the 25th and 50th percentiles revealed that: being a student was correlated with a higher tendency to visit malicious URLs (first security behavior factor); and more interactions with security and privacy content on Facebook made by familiar posters and technical content on Twitter were correlated with having more antivirus software installed (second security behavior factor). Similarly, the models for the 75th and 90th percentiles revealed that: more interactions with security and privacy- and breach-related content was correlated with a higher value of the first security behavior factor. Being older was positively correlated with the first security behavior factor for the 90th percentile. Finally, more interactions with technical posts on Twitter made by familiar posters and with security and privacy and technical posts made by familiar posters on Facebook were positively correlated with the second security behavior factor for the 50th percentile. These features indicate that certain demographics such as being a student and age as well as interactions with social media posts were associated with better security behavior for both the least, typical, and most desirable levels of security behavior (outcomes in the low, median, and higher percentiles). However, no social media factor was correlated with any of the security behavior factors across all or a majority of the percentiles, thus implying a lack of strong correlation between the demographics and social media factors and security behavior factors [62].

5.5.3 Summary of findings

In summary, while some amount of interactions with security and privacy, technical, or breach-related content on social media, whether as a result of sharing or consuming, along with demographics were correlated with better security behavior, no factor was consistently correlated with this behavior.

5.6 Limitations

We analyzed the data of a (relatively) small subset of participants in the SBO due to the difficulty of recruiting participants to agree to additional data collection. Popular social networks such as Facebook do not provide access to activity logs through their API, and obtaining access to such data is otherwise difficult [8]. Our dataset is a tradeoff between a large amount of participants (which it does not have) and extensive, hard-to-obtain, longitudinal Facebook data (which it does have).

The data we collected contains information only about posts that participants explicitly interacted with on Facebook or Twitter. Therefore, we could analyze only those posts, and not, for example, posts that participants may have seen but with which they did not interact. While such posts may further contribute to our understanding of how security and privacy is discussed on social media and how often, we were particularly interested in studying active engagement with content as opposed to passive engagement (i.e., seeing posts without interaction), including because evidence of active engagement implies that the participant has seen a post.

We validated our results regarding the low number of security and privacy posts on Facebook and Twitter by applying the same post-detection methodology on a set of random, public Twitter posts. While those numbers supported our findings, evaluating random posts does not account for the visibility or impact of these posts. For instance, a random post by a popular or famous user would have more of an impact and reach more people than a random post by an ordinary user.

The participants in our study were female-skewed with a relatively high percentage of people who knew at least one programming language. Even though our results are not based on a representative population, we believe the frequency of security and privacy posts will likely be even lower with a less technically savvy population. Furthermore, our thematic analysis reveals important patterns for how security and privacy may be discussed on social media that are likely independent of demographics (mentioned in Section 5.1).

As in Chapters 3 and 4, the SBO only contains data about Windows users. However, as Windows is the most commonly used OS for personal computers [61], we do not believe our findings are fundamentally affected by this.

Finally, again as in Chapters 3 and 4, due to the nature of the SBO data collection infrastructure, participants may be skewed towards people with lower concerns about security and privacy. Despite this, participants tended to self-report high on the SeBIS intentions scale [86], answers to which they were optionally asked to provide at the time of enrollment to the SBO. On average, 64% of the participants whose data we studied indicated a frequency of "sometimes" or higher for the extent that they: a) secure their devices, b) generate strong and varied passwords, c) demonstrate proactive awareness of security issues, and d) update their computer software.

5.7 Discussion

We discovered that security and privacy were not topics of frequent discussion on social media, based on the social media interactions of 38 participants spanning over 200k posts. Furthermore, only a few participants accounted for a majority of interactions with the security and

privacy posts we identified. Through our thematic analyses, we found that, more often than not, discussions about security and privacy revolved around jokes or sarcasm or merely mentioned a security- or privacy-related buzzword without discussing it further. Posts sometimes shared security experiences with clear lessons but very rarely did posts about security and privacy include constructive, actionable advice. Our findings provide a characterization of the security and privacy content that people may naturally encounter as part of social media browsing. While prior work has characterized the different ways constructive security advice is available in articles across the web [184, 192], our work characterizes the security- and privacy-related content people encounter on social media.

We next discuss the implications of our findings on increasing the dissemination of security and privacy information through social network, how disseminated information could encourage healthy security behaviors, and what may constitute effective security and privacy education for the consumers of this information.

5.7.1 Disseminating security and privacy advice in social networks

We were surprised to find only 175 posts related to security and privacy out of the 200,964 social media posts that we examined. It is possible that security and privacy information information is being de-prioritized by social media algorithms in comparison to other topics. In fact, prior work has also shown that security and privacy content does not reach all computer users equally, which could also be reflected in social networks [189, 190]. Researchers have found that social influence—particularly within social media—can in experimental settings encourage people to taking security-enhancing actions [76, 77, 78, 81, 88]. However, through measurements of historical social media interactions and behavior, our findings and exploratory statistical analyses suggest that the positive influence of real social-media posts and discussions on security behavior may not be as high or likely as experimental results suggested.

Similarly to in-person social circles, it appears that people who do not regularly interact with people on social media who are knowledgable about security and privacy are unlikely to come across information about it unless they seek it out. Future work could explore these social group dynamics within a social network further to understand to what extent information is localized to specific communities and who is receiving this information on social media. Advertising and public relations campaigns do not rely solely on existing social structures to diffuse information; they manipulate the spread of information through the social network [57, 109, 221, 233]. Furthermore, social media has been highlighted as having an important role in information diffusion [46]. Future work could explore how to launch large-scale "security and privacy" publicity campaigns on social media using lessons learned from older cybersecurity awareness campaigns [19, 198]. By combining the mechanics behind viral marketing and leveraging the properties of social networks shown to be effective at spreading awareness [164, 227], information about security and privacy best practices has the potential to reach social communities in a network previously out of the loop.

5.7.2 Inciting changes in security behavior

After discovering how little constructive advice was present in the social media interactions we analyzed, we hypothesized that it was extremely unlikely that people generally come across enough security-related advice on social networks to have their security hygiene affected. However, not only is increasing the spread of security information in a network important, but when the information reaches more people in a network the information still needs to incentivize people enough to act. Previous research has explored in depth what influences users to change their security behavior and found that often the perceived burden of completing an action wins over the potential benefit [43, 118]. Even if the cost or burden is low, people often believe that they are not susceptible to attacks and do not need to take action [33, 235]. In short, people may not be motivated to implement even low-cost actions that would provide a clear benefit. Researchers recently studied security and privacy advice available across the web and found that people perceived most advice to be somewhat actionable [192]. However, they found that the hurdle to implementing advice was the burden the advice articles placed on the users to prioritize different advice. Self-reporting that advice is actionable also does not accurately represent motivation to implement advice or the likelihood of doing so.

In addition to the methods of increasing the spread of constructive security-and-privacy advice discussed in Section 5.7.1, the presentation of security advice on social media could be improved in ways suggested by existing work: outlining clear incentives to take a security-enhancing action can play a role in influencing action [118]. Conveying risk clearly is important [101], and prior work has found that conveying security advice through storytelling and sharing of experiences is effective [185]. Furthermore, presenting advice in different formats may garner more results. For example, communicating security and privacy advice and risks through video was shown to be effective [101]. Many of these approaches could be effectively implemented on social media, such as through the popular Instagram or Facebook reels [20] for sharing video snippets.

5.7.3 Security education

Although our work points out the lack of engagement with security information on social media, the degree to which users *should* be educated about security and privacy remains an open question. Implementing security in systems has often been an afterthought in system design [110, 210], with its usability being even less of a priority [100]. As a result, systems may place the burden of deciding to implement security on its users, requiring them to make complex decisions in the process [72, 80, 125]. Much of the research in the security and privacy community advocates for factoring security into system design from the start and relieving people of the burden of security decision-making as much as possible [34, 118, 197].

Many current systems, tools, and services still require their users to be partly responsible for maintaining security (e.g., creating and remembering strong passwords). Social engineering attacks (e.g., shoulder-surfing, phishing) are multi-faceted with no fix-all technical solution and may also require some savviness from the user [65, 131]. Security awareness plays an important role in educating people about the decisions and steps to take to achieve the level of digital security they expect [43, 115, 135]. Amidst the complex security requirements of systems, several

tools have been created to assist users. For example, password managers have been shown to be an effective way to create different strong passwords across websites [90]. However, the use of password managers is uncommon [21]. Our study suggests that computer users are unlikely to learn to improve their security behaviors through their general interactions with online content, at least in the context of social media. This suggests that more intentional security-and-privacy education and designing systems that further remove the burden of making security-and-privacy decisions are both necessary.

5.8 Conclusion

We analyzed real Facebook and Twitter logs and security behavior data from 38 participants over almost four years to empirically study *how often* and *how* security and privacy is talked about on social media, with the goal of understanding whether social media posts could be helpful in encouraging healthy security behavior. We identified a surprisingly low number (0.09%) out of over 200k posts in the social media logs we analyzed to be about security and privacy. As the participant sample was slightly skewed toward technically savvy participants, it is likely that the fraction of security-and-privacy posts interacted with by the general population is even lower, which is supported by our analysis of the frequency of security-and-privacy posts in a random sample of Twitter posts. We gained insight into the underlying nature of the security-and-privacy-related posts on Facebook and Twitter by conducting a thematic analysis. We uncovered five major themes in how the social media posts we analyzed talked about security and privacy. Only one theme described posts that spoke constructively about security and privacy, and these constructive posts made up only 20% of the already small pool of security and privacy posts we identified. Posts often spoke about security and privacy in passing, through jokes or sarcasm, or through anecdotes without conveying information about healthy or unhealthy security behavior. Though prior work suggests social media can be a highly effective platform for influencing healthy security habits, our findings suggest that there may not be enough constructive discussions on social media in practice such that people are educated and incentivized enough to make changes in their security habits, which exploratory statistical findings substantiated. Based on our findings, we discussed directions for future work toward achieving widespread constructive security awareness and behavior—both in terms of how to increase the dissemination of advice using social networks and how to improve the efficacy of security-and-privacy posts in encouraging healthier security behavior.

Chapter 6

How can the spread of security and privacy posts in social networks be improved?

6.1 Introduction

In the previous chapters, we found that discussions or information about security and privacy were not encountered often through web browsing or in the social media datasets we studied. This trend was surprising. In particular, we were surprised to see these results for social media, given its prevalence among adults [204] and findings from prior work showing that social media are some of the most frequently used platforms for discussing security and privacy [85]. However, since social media has been shown to be a very effective educational tool (see Section 2.4), we believe that security awareness can be increased through social media. In particular, a higher volume of posts about security and privacy in social networks or increasing the spread of posts about security and privacy in social networks can be beneficial to helping people improve their security practices and habits. In this chapter, we focus on determining how posts about security and privacy can be popularized and spread further.

We determine recommendations for distributing security and privacy information effectively in social networks. While in previous chapters we studied security and privacy content consumed or created by a set of users in relation to their security behavior, in this chapter we study security and privacy posts on a large scale without considering the security behavior of individual users. We collect and analyze a large dataset of security and privacy-related posts on the social network Reddit to analyze what features of these posts could be associated with further spread. For 30,337 Reddit posts, we identified and extracted several features to represent post properties and two outcomes that represent the spread of a post. We analyzed the relationships between these features and outcomes across the set of posts through statistical models and identified several properties of posts that may help posts obtain wider spread on a social network.

For example, the presence of visual attributes in posts in the form of emojis or images displayed next to posters' names or text was positively correlated with higher post engagement. We also found that posts that conveyed strong emotions or sentiment (e.g., opinions that conveyed anger or joy) or in contrast, tentative emotions, were both associated with higher visibility and spread. The length of the posted content was also important, perhaps because longer posts are

likely to contain detailed, helpful information [144] and hence, garner more attention. However, engagement decreased when the content relied largely on links instead of text. Based on these findings, we conclude this chapter with recommendations and insight regarding how to better distribute security and privacy information in social networks in the form of posts.

6.2 Data collection and dataset

To study what properties of posts about security and privacy contribute to their spread, we analyzed a set of Reddit posts related to security and privacy[1]. Posts on Reddit are made inside "subreddits," which are communities (i.e., forums) in which people create posts about shared interests that are specific to the topic of that subreddit. All subreddit names are of the form "r/{topic}" and Reddit users are often called "Redditors". Redditors can engage with posts by commenting on them or voting on how positively or negatively they perceive them. In particular, they can either give the post an "upvote" to indicate a positive reaction or a "downvote" for a negative reaction.

We collected our set of security- and privacy-related posts to analyze by querying the Reddit search API against each in a set of search terms related to security and privacy.

We determined the set of search terms by first manually examining the "hot" 100 posts (i.e., the first 100 posts on the homepage of a subreddit) from each of "r/cybersecurity" and "r/privacy". These subreddits were selected because they were the most popular subreddits whose names contained the words "security" and "privacy" when searching for these strings in Reddit's search engine. We extracted as many terms (i.e., a word or phrase) as possible from each "hot" post that related to computer security and privacy (criteria described in Section 6.2). We then trimmed that set of terms to only "good-quality" terms by ensuring that searching for that term on Reddit resulted in few false positives (i.e., posts that are not related to computer security and privacy). Finally, we collected up to 250 posts that Reddit returned as search results for each of the final good-quality terms. We describe this process in detail in Section 6.2.

What posts do we consider relevant to security and privacy?

While the terms "security" and "privacy" are used in many domains (e.g., national security, security guards, privacy in a physical space), we considered posts to be related to security and privacy if the topics were discussed in the context of computers or online activity. Often posts discussed or asked about security and privacy explicitly. We also considered posts that provided or asked for help with setting up tools whose purpose was for computer or online security and privacy, even if the post did not clearly discuss security and privacy implications. Additionally, posts that talked about security of physical objects were not relevant unless their implications were tied to the theft of computer or online data and those implications were described in the post. When the purpose of the post for a given search term was not clearly related to security and privacy, we considered the post relevant if the default meaning or purpose of the search term was related to security and privacy.

[1]Although Reddit refers to each post as a "submission," we use the term "post" throughout this chapter for consistency.

When examining posts, if it was unclear from the post title and text whether the content was related to computer security and privacy, we additionally visited links in the post (if any) to determine if the above criteria were met.

Constructing search terms and collecting posts

The following steps describe in detail our algorithm for compiling and evaluating search terms related to security and privacy. As mentioned earlier, after executing this algorithm, we are left with a set of "good-quality" search terms. For each search term in this resulting set, we then fetch the maximum number of posts returned by the Reddit search API. We refer to this final set of posts to be collected for a search term as the "to-be-collected" set of posts in the algorithm below. After these "to-be-collected" posts are collected for a particular term, we further filter them in a manner as specified by the criteria in the algorithm. The posts collected for a search term that remain after this filtering process constitute what we call the "final set" of posts, which we later use in our analyses.

1. Fetch the "hot" 100 posts displayed in each of the two sub-reddits "r/cybersecurity" and "'r/privacy" (200 posts in total).
2. For each post, identify terms in the post—by manually examining its title and text—that are related to security and privacy. Here, a term can be either a combination of words or a single word. Not all posts may have clear security and privacy terms in their titles or text even if a post is about security and privacy (by virtue of being posted in the security and privacy subreddits). Ensure that the recall (percentage of false negatives or posts in which there are no extractable terms related to security and privacy) on this set of 200 posts is not below 75%.
3. For each term, fetch the first 20 search results from Reddit where the results are sorted by "relevance". Examine how many of these 20 posts are false positives based on their titles and text bodies according to the criteria in Section 6.2. Depending on the value of the precision (percentage of true positives) over the 20 posts as outlined below, decide how the to-be-collected posts for each search term will be handled (i.e., how the collected posts for a search term will be filtered in order to construct the final set of posts to analyze). The possible cases are defined as:
   (a) precision == 100%: do not filter any of the to-be-collected posts for this search term and add all collected posts to the final set
   (b) $80\% \leq$ precision $\leq 95\%$: filter out from the to-be-collected posts for this search term, only posts whose content matches the false positives found in the 20 posts; add the remaining unfiltered posts from the to-be-collected set of posts to the final set
   (c) $50\% \leq$ precision $< 80\%$: manually examine each of the to-be-collected posts for the search term and filter out false positives; add the remaining unfiltered posts from the to-be-collected set to the final set
   (d) precision $< 50\%$: discard search term and do not collect posts for the search term; ensure that the recall on the original set of 200 posts from the two subreddits does not fall below 75% by computing the percentage of the 200 posts that no longer have

terms matching them because of discarded terms[2]

After step 2, the recall on the original 200 posts was 82.5%. The final recall after discarding some search terms in step 3(d) was 77.5%. Using the final list of search terms, we fetched a test set of 50 unseen posts, 25 each from "r/cybersecurity" and "r/privacy". We achieved a 66% recall on this test set, which we computed by manually examining each of the 50 posts and determining whether any of the final terms (including stemmed variations) were present in each post.

For each search term in the resulting list of search terms, we fetched 250 Reddit posts or less where 250 is the maximum number of posts Reddit returns for a search term[3]. We then removed duplicates from this set by checking for posts with identical post IDs. The list of search terms we considered is shown in Table 6.1. We removed duplicate posts and then manually filtered the posts collected against search terms we marked for filtering in the manner specified above to remove false positives. We then removed non-English posts as detected by SpaCy language model tools [121]. As a result, we ended up with 30,337 total posts related to security and privacy which we used for our analyses.

6.3 Analysis

In this section, we describe how we studied what features of security and privacy posts are associated with higher spread. To study this problem, we analyze the 30,337 posts described in the previous section. For each post, we extracted a set of features and outcomes and study the relationship between features and each outcome via statistical inference models. We start this section by describing the features we extracted and studied for each post. We then describe the outcomes we defined for spread for each post. Finally, we discuss the statistical models we built to make inferences about relationships between the features and outcomes.

6.3.1 Features

We computed a set of features about Reddit posts inspired by and adapted from prior work that studied the popularity and spread of posts in social networks in different topic domains [120, 139, 155]. We computed the features in each feature set below over some combination of the title and text of each post.

Feature set 1: Features describing the content of the post

Previous work has studied various features of social media post text in relation to the post's popularity or the amount of engagement it gets [120, 139, 155]. Examples of these features

[2]When checking which of the original 200 became false negatives after discarding a term, if a particular post originally matched the discarded term but had given way to multiple non-discarded terms in step 2, the post does not become a false negative. Additionally, because Reddit stems search terms [29], if a particular post originally matched the discarded term but contains a stemmed variant of a non-discarded term, do not mark the post in the original 200 as a false negative.

[3]Reddit by default only stores and returns 250 posts. However, before returning, internal preprocessing may reduce the number of posts even further from 250.

keyword	***action***
2fa	keep
account compromise	keep
account privacy	keep
account tracking	remove
ad blocking	keep
algorithms surveillance	filter
anonymity privacy	keep
anonymous DNS	keep
antivirus	keep
app spying	keep
app tracking	keep
authentication	remove
authenticator	keep
bitcoin	remove
block domains	keep
block website	keep
blocked verification	keep
blocklist	selectively filter
blue/red activities	remove
brave privacy	selectively filter
bromite privacy	keep
browser fingerprint	keep
browser privacy	keep
browser security	keep
browser telemetry	selectively filter
browsing privacy	keep
browsing tracking	filter
CEH	filter
cipher spec	remove
CISM	filter
CISSP	selectively filter
cloud leak	filter
collect data	remove
collect user data	keep
company collecting personal data	keep
computer data owned	remove
cryfs	filter
cryptocurrency	remove
cryptocurrency scams	keep
cryptography	keep
CTF	remove
CVE	selectively filter

cyber	filter
cyber attack	keep
cyber hacker	keep
cyber incident	keep
cyber jobs	remove
cyber sec	keep
cyber security	keep
cyberattack	keep
cybersec	keep
cybersecurity	keep
cybersquatting	keep
dark patterns	keep
data breach	keep
data desensitization	remove
data extortion attack	filter
data masking	keep
data privacy	keep
data protection	keep
data security	keep
DDoS	selectively filter
decrypt	remove
deobfuscating javascript	remove
detect camera	remove
digital forensics	selectively filter
digital privacy	keep
digital security	remove
duckduckgo	keep
duckduckgo browser location	selectively filter
DVWA	keep
email privacy	keep
email spam	selectively filter
encrypted	keep
encryption	keep
ente.io	remove
ethical hacking	keep
exploit bugs	remove
exploit vulnerabilities	keep
exploited	remove
exploits	remove
exploits computer	selectively filter
expose data	keep
firefox privacy	keep
firewall	filter
forensics	remove

GCIH	keep
GDPR	keep
grapheneOS	keep
GWAPT	keep
hack	remove
hacked	selectively filter
hackers	remove
hacking	remove
hijacking	remove
honeypot	remove
ID fingerprint	remove
identity stolen	remove
infection security	filter
info security	keep
information privacy	keep
internet flaw	keep
internet safety	keep
ISS	remove
ISSAP	keep
ISSE	remove
IT GRC	remove
kali linux	remove
key pairs	keep
least privilege policy	selectively filter
location data	selectively filter
malicious software	keep
malware	keep
metasploit	keep
metasploitable	keep
MiTM attack	keep
monitor social media	selectively filter
netguard	keep
network security	keep
offensive apps	remove
onion routing	keep
online accounts privacy	keep
online defense	remove
online privacy	keep
online security	selectively filter
paranoid data	filter
password	keep
password manager	keep
pen testing	keep
pentest	keep

personal data	keep
personal data exposed	keep
personal data privacy	keep
personal info	keep
personal information	filter
personal key	remove
PGP	filter
phishing	keep
phone hijack	keep
phone privacy	keep
phone security	selectively filter
privacy	selectively filter
privacy campaign	remove
privacy data	keep
privacy device	keep
privacy focused	keep
privacy IDE	remove
privacy information tracking	keep
privacy laptop	keep
privacy location	selectively filter
privacy OS	keep
privacy policy	keep
privacy risk	keep
privacy service	keep
privacy software	keep
privacy threat	keep
privacy whatsapp	keep
private data	keep
private DNS	remove
private information	filter
private key	keep
private sync devices	keep
public key	keep
ransomware	keep
remote code execution	keep
RSA	filter
safe account	remove
security	remove
security breach	remove
security incident	selectively filter
security key	keep
security privacy	selectively filter
security risk	remove
security weakness	keep

sensitive information	filter
signal	remove
spammy	remove
spy tech	remove
spyware	keep
sql injection attack	keep
surveillance[4]	filter
surveillance collection of data	filter
surveillance network	remove
surveillance privacy	remove
system permission	remove
third party tracking	selectively filter
threat	remove
threat model	keep
TLS	remove
TOR	remove
tracker blocker	keep
tracker control	keep
tracking url	remove
trojan	remove
tryhackme	keep
typosquatting	keep
unencrypted	remove
vpn	selectively filter
vulnerabilities	remove
vulnerability scan	keep
vulnerable hacking devices	keep
web privacy	keep
whatsapp safe	keep
whatsapp trust	remove
windows privacy	remove
wireshark filters	remove

Table 6.1: List of all terms extracted from the 200 Reddit posts from "r/cybersecurity" and "r/privacy" (first column). The second column contains the "action" we assigned to the search terms when collecting Reddit posts against each of them. The "keep" action corresponds to criterion 3(a) of the algorithm in Section 6.2. The action "filter" corresponds to criterion 3(c) of the algorithm. The "selectively filter" action corresponds to criterion 3(b) of the algorithm. The "remove" action corresponds to criterion 3(d) of the algorithm.

include the emotionality and sentiment of the post as well as the weights of topics computed

[4]If, for a post collected against this term, it was unclear from the post text whether it was relevant a link was present, and the link did not clarify whether the post was about computer or online surveillance, we considered the post to not be relevant.

through topic modeling. We adapt these features into some of the features listed below except for topics from topic models because when computing topic models, we were not able to identify coherent topics. Furthermore, several features related to the emotionality and tone of the text in prior work were determined through manual analysis [139] whereas we compute all our features via automated means.

We also define features that consider aspects of post text that have been shown to be important in the readability of privacy policies [89, 136] or texts about security [191] such as the readability or length of texts.

sentiment_pos: Inspired by prior work finding post sentiment related to popularity [155], this feature describes how much of a positive sentiment the post title and text exhibit according to SentiStrength [215]. This is a numeric feature that can take values 1, 2, 3, 4, or 5 where 1 indicates no positive sentiment and 5 indicates an extremely positive sentiment.

sentiment_neg: Similarly to the above feature, this feature describes how much of a negative sentiment the post title and text exhibit, again, according to SentiStrength [215]. This is a numeric feature that can take values -1, -2, -3, -4, or -5 where -1 indicates no negative sentiment and 5 indicates an extremely negative sentiment.

tone_anger: This feature describes to what degree the tone of the post text is angry according to IBM Watson Tone Analyzer [18]. We included features computed by the tone analyzer to represent the emotionality of text as inspired by prior work that studied text emotionality manually [139]. This particular feature represents the amount of anger in a post as a continuous value in $[0, 1]$ where 1 indicates a strong presence of the anger tone. The tone analyzer only returns a score for anger if it was greater than 0.5. Therefore, if a score for anger was not returned, we consider this feature's value to be 0.

tone_fear: This feature describes to what degree the tone of the post text is fearful according to IBM Watson Tone Analyzer. We included it for similar reasons as for the `tone_anger` feature and its possible values are the same as for that feature.

tone_joy: This feature describes to what degree the tone of the post text is joyful according to IBM Watson Tone Analyzer. We included it for similar reasons as for the above two features and its possible values are the same as for those features.

tone_sadness: This feature describes to what degree the tone of the post text is sad according to IBM Watson Tone Analyzer. We included it for similar reasons as for the above three features and its possible values are the same as for those features.

tone_analytical: This feature describes to what degree the tone of the post text is analytical according to IBM Watson Tone Analyzer. We included it for similar reasons as for the above four features and its possible values are the same as for those features.

tone_confident: This feature describes to what degree the tone of the post text is confident according to IBM Watson Tone Analyzer. We included it for similar reasons as for the above five features and its possible values are the same as for those features.

tone_tentative: This feature describes to what degree the tone of the post text is tentative according to IBM Watson Tone Analyzer. We included it for similar reasons as for the above six features and its possible values are the same as for those features.

readability_flesch: This feature describes the readability of the post text on the Flesch Reading Ease scale [95], inspired by prior work studying the readability of articles about security [191] and privacy policies [89, 136]. This readability score is a value that typically falls in $[0, 100]$ where 0 indicates very unreadable or confusing texts and 100 indicates text that is very easy to read. The score can be outside of this range in practice; though the score's maximum is 121.22 [2], the readability was calculated to be 206.84 when the post text was empty or did not contain any alphanumeric characters. There is no minimum possible value for this score.

text_num_characters: This feature describes the length of the post in terms of the number of characters in the post text. We included this feature because of prior work that found that different length-related properties of privacy policies were correlated with readability [136], a feature which we also considered in relation to post popularity.

num_emojis: This feature describes the number of emojis present in the post text. We included this feature because the presence of emojis has been shown to be correlated with text sentiment [74], a feature which we also considered in relation to post popularity.

Feature set 2: Features describing the poster of the post

Prior work studying the popularity posts has examined properties of posters in relation to the popularity of their posts. For example, the popularity of posters has been defined as the number of times posts made by a poster were retweeted on Twitter [120] or the number of contacts or followers a poster has in an unspecified social network [155]. Inspired by this previous work that found poster popularity to be important for post engagement, we define several features of the posters of each Reddit post, each of which describes some aspect of that user's activities and popularity specific to Reddit.

poster_is_gold: This feature is a binary indicator that describes whether the poster has been awarded "gold status". Reddit users are given "gold" when a post or comment they made is appreciated by another Reddittor who then gives them said gold [3]. Reddit users who have been given gold have "gold status" which we consider a measure of user popularity. Gold users have a symbol displayed next to their username to indicate this.

poster_num_submissions: This feature describes the number of posts made by the poster on Reddit. Prior work measured the popularity of a poster in terms of the number of posts they made on Twitter [120] which we adapted to Reddit by counting the number of posts made by a poster.

poster_total_karma: This feature describes the total karma a Reddit user has accrued based on all their activity over the entire time they've used their Reddit account. We included this feature because the total karma indicates how much good a Reddit user has done overall for the Reddit community [28] which can be considered a measure of user popularity.

poster_comment_karma: This feature describes the total karma a Reddit user has accrued based on all comments they've made on posts over the entire time they've used their Reddit account. We included this feature for reasons similar to the `poster_total_karma` feature.

poster_link_karma: This feature describes the total karma a Reddit user has accrued based on all posts they've made over the entire time they've used their Reddit account. We included this feature for reasons similar to the above two features.

poster_awardee_karma: This feature describes the total karma a Reddit user has accrued based on all awards they received over the entire time they've used their Reddit account. We included this feature for reasons similar to the above three features.

poster_awarder_karma: This feature describes the total karma a Reddit user has accrued based on all awards they gave out over the entire time they've used their Reddit account. We included this feature for reasons similarly to the above four features.

poster_avg_score: This feature describes the average score (the number of downvotes subtracted from the number of upvotes on a post; see Section 6.2 for an explanation of these terms) of all posts made by the poster. This feature was adapted from prior work that defined a user's popularity as the number of times that user's posts on Twitter have gotten retweeted on average [120]. Instead of post retweets, we average over the score of the poster's posts and expose this as a feature.

poster_avg_upvote_ratios: This feature describes the average upvote ratio (the number of upvotes divided by the total number of upvotes and downvotes combined) of all posts made by the poster. Similarly to the `poster_avg_score` feature, we adapted the popularity metric of the average number of retweets on a poster's posts on Twitter to average the upvote ratios of the user's Reddit posts.

Feature set 3: Features describing other relevant properties of the post

Here, we describe features related to other aspects of posts that did not fall into the first two feature sets. Inspired by prior work that found that properties of posts related to non-text elements

of the post were important for post engagement [139, 155, 163], we define a number of features describing properties of the posts that are not based on the text of the post. For example, the presence of media or the type of media (e.g., video, image, text, hyperlink) in a post or whether it was shared before were factors considered in this feature set.

has_link: This feature is a binary indicator of whether a post contains an external link in its text. We included this feature because of prior work's focus on the presence and type of media in a post [139] and any external media would be accessible in the post through a link. We computed this feature by searching for text that matched the regular expression for a URL. This approach worked for hyperlinks (links disguised as regular text) too as we were analyzing the raw text which contained the underlying URL.

only_link_no_text: This feature is a binary indicator of whether a post contains only an external link with no other text surrounding it. We included this feature for similar reasons to above but to capture more detailed information about the presence of external links in posts.

has_image: This feature is a binary indicator of whether a post contains an image or link to an image in its text. We included this feature inspired by prior work that studied the presence of pictures and visuals in posts related to the engagement they get [155, 163]. We computed this feature by first looking for links in the way we did for computing `has_link` and then by checking if the file extension on any of those links corresponded to an image. We checked this by comparing the file extensions against a list of most common image file extensions on the web [22].

only_image_no_text: This feature is a binary indicator of whether a post contains only an image with no other text surround it. We included this feature for similar reasons to above but to capture more detailed information about the presence of images in posts.

has_author_flair: This feature is a binary indicator of whether the poster of the post has flair (a combination of short text and small images that users select for themselves in a subreddit [27]) displayed next to their name. We included this feature, again, based on prior work that studied the presence of visual attributes in posts in association with post engagement [139, 155].

has_link_flair: This feature is a binary indicator of whether the post has flair displayed above the post text. This flair is different from `has_author_flair` since one user can make multiple posts with different version of link flair. We included this feature for similar reasons as for the `has_author_flair` feature.

is_crosspost: This feature is a binary indicator of whether the post is crossposted from another subreddit. Prior work studied whether a post was shared before in relation to post popularity on Twitter [120]. Hence, we included that feature but adapted it to Reddit by checking whether a post was cross-posted from another sub-reddit.

6.3.2 Outcomes

We considered two metrics as outcomes to represent spread in our analysis. The two outcomes (described below) are the total number of comments and the total number of votes on a post. We selected these two metrics because posting comments and voting are the two main ways that Redditors can interact with posts. Therefore, the numbers of each of these interactions indicate the amount of visibility and attention a post received, which we consider as spread in our analyses.

`number_comments`: The first metric of spread we studied is the total number of comments on a post which includes higher-level and nested comments as used in prior work studying spread and engagement on Reddit [169, 206]. Although it is not clear how much of an influence more comments on a post has on its visibility soon after it is posted on Reddit [7], we use it as a metric of spread as the number of comments provides an indication of how visible a post became such that it generated a higher or lower number of comments.

`total_votes_estimate`: The second metric we defined is the total number of votes (upvotes and downvotes) for a post. Although Reddit appears to consider a higher number or fraction of upvotes relative to downvotes as a factor in deciding to promote its visibility [7], we study the total number of votes to give us an indication of how much visibility and engagement the post received even in the case of bad publicity (i.e., downvotes). However, Reddit does not provide the total vote count directly and much of prior work studying Reddit spread and engagement in terms of votes is only able to study the amount of upvotes with respect to the number of downvotes [160, 206]. Therefore, we estimated this total from the `score` and `upvote_ratio` exposed by Reddit about a post which are defined as follows:

$$score = upvotes - downvotes$$

$$upvote\_ratio = \frac{upvotes}{upvotes + downvotes}$$

Solving the above system of equations, we ended up with the following equations for upvotes and downvotes whose sum is represented as `total_votes_estimate`.

$$upvotes = \frac{score * upvote\_ratio}{(2 * upvote\_ratio) - 1}$$

$$downvotes = score - upvotes$$

When the number of downvotes is greater than the number of upvotes, Reddit does not return or display a negative number and instead returns 0. Therefore, we were not able to accurately compute the estimated total votes when `score` was 0 since the formula may incorrectly yield 0 for both upvotes and downvotes. Additionally, `upvotes` is undefined when `upvote_ratio` is 0.5.

6.3.3 Statistical modeling

We studied the relationships between features of posts and the two outcomes via linear statistical models. For both of the outcomes, neither of the models satisfied the necessary assumptions of linearity and homogeneity of variance for parametric multiple linear regression. There were also several outliers in the dataset which were interesting for our analysis (i.e., the extremely popular posts). However, these would heavily influence the linear regression model due to its estimation of the conditional mean of the outcome, which is sensitive to outliers and its use of the L2 least squares loss function that amplifies the impact of outliers. Therefore, we modeled the relationships between features and outcomes via quantile regression models [132] (as also used in Chapter 3 and 5). Quantile regression models are non-parametric linear models that are robust to outliers due to what they estimate and the loss function they use. Rather than predicting how the conditional mean (which is affected by extreme values) of the outcome varies with feature values as linear regression does, quantile regression predicts how the conditional quantile (e.g., 0.5 quantile or median which is not affected by extreme values) for a given quantile varies with feature values. For example, quantile regression can estimate how the median outcome (outcome in the 0.5 quantile 50th percentile) increases or decreases with changes with respect to each of the individual features.

Quantile regression provides the advantage that we can study relationships between features and outcomes for various extremes or levels of the outcomes. For example, we can study what features are related to each of the highly spread posts (posts in higher quantiles) and the posts with low spread (posts in lower quantiles). Therefore, we studied how the features of posts are associated with the spread of posts according to the two outcomes through four quantile regression models each. The four models compute the regression line for the 0.25, 0.5, 0.75, and 0.90 quantiles of the outcome as done in prior work to study the associations with the least, typical, more than typical, and most popular posts [62].

For each of the models, we removed one feature from every pair of collinear features because multicollinearity can result in a model with inaccurate results [45]. Additionally, we removed rows with empty or null values to build an accurate model. In particular, we removed rows with empty poster-related features due to empty results provided by the Reddit API. When studying the `total_votes_estimate` outcome, we also removed posts for which we could not compute their total votes (i.e., posts that had a score of 0 or upvote ratio of 0.5).

6.4 Results

We computed four quantile regression models for the outcome `number_comments` over 29,042 posts after removing posts with empty features. The results of the first model can be found in Table 6.2. The remaining three models for this outcome are described in Tables 9, 10, and 11 in Appendix E. Similarly, we computed four quantile regressions model for the outcome `total_votes_estimate` over 27,598 posts after removing posts with empty features and posts for which we could not estimate the vote count. The results from these models can be found in Tables 12, 13, 14, and 15 in Appendix E.

As in Chapter 3, we considered a feature to be significant if its significance level was less than

0.05. For each outcome, a few features were consistently significant across all four models and all contributed to the outcome in the same direction (i.e., positive or negative). The coefficients reported can be interpreted as follows: the conditional quantile of the outcome (i.e., the outcome in the quantile being studied for a given set of feature values) increases by the value of the coefficient for every unit increase in the feature's value for continuous features. For binary variables, the outcome increases by the value of the coefficient when the feature's value changes from false to true. When a feature's effect size is higher for one quantile than another, it implies that that feature is differently important for different levels of the outcome value. For example, if a feature is significant with the highest coefficient for the 0.9 quantile, it implies that the value in the 0.9 quantile is affected the most by that feature. In other words, that feature played a larger role for the spread of the most popular posts.

We now discuss results specific to post features in each feature set from Section 6.3.1.

Features describing the content of the post

Only one feature in this feature set, `num_characters`, was consistently correlated positively with one of the outcomes (`number_of_comments`) for all four quantiles. Posts that were longer in length were correlated with more comments, although only marginally for all four quantiles.

A few features also showed different behavior for different quantiles for each of the outcomes (e.g., a feature could be significant for one quantile but not for another or the sign of the coefficient could be different for different quantiles). We describe results about these features here and only report on correlations for a specific outcome if the absolute value of the coefficient (increase/decrease in outcome variable) for one of the quantile models is at least one.

- Posts with a stronger positive sentiment were associated with more comments (0.75 quantile: coef. = 1.81, p <0.01). The effect of positive sentiment was marginal for the 0.5 quantile and not significant for the other two quantiles.
- Posts with a less strong negative sentiment were associated with fewer comments (0.75 quantile: -2.08, p <0.01) and votes (0.75 quantile: coef. = -1.47, p <0.01). The effects of a stronger negative sentiment were marginal or not significant for the other quantiles for both outcomes.
- Posts with an angry tone were associated with more votes (0.5 quantile: coef. = 2.11, p <0.01). The effects of an angry tone in the text were not significant for the other quantiles.
- Posts with a joyful tone were associated with more comments (0.5 quantile: coef. = 1.04, p <0.01) and votes (0.5 quantile: coef. = 1.21, p <0.01). The effects of a joyful tone in the text were lower or not significant for the other quantiles for both outcomes.
- Posts with a tentative tone were associated with more comments (0.5 quantile: coef. = 2.66, p <0.01). The effects of a tentative tone in the text were lower or not significant for the other quantiles.
- Posts with more emojis were associated with more comments (0.9 quantile: coef. = 1.56, p <0.01). The effects of more emojis were marginal or not significant for the other quantiles.

Longer posts were associated with a higher spread in terms of the number of comments. This

feature_name	*baseline*	*coef.*	*std. err.*	*t*	*p*
Intercept		**0.680**	**0.157**	**4.336**	**<0.01**
has_link: true	**false**	**-0.799**	**0.049**	**-16.332**	**<0.01**
has_image: true	**false**	**0.160**	**0.053**	**3.004**	**<0.01**
only_link_no_text: true	**false**	**-0.676**	**0.058**	**-11.694**	**<0.01**
has_author_flair: true	**false**	**0.268**	**0.046**	**5.787**	**<0.01**
has_link_flair: true	**false**	**0.846**	**0.033**	**25.539**	**<0.01**
is_crosspost: true	false	0.042	0.106	0.392	0.695
num_emojis		**0.033**	**0.010**	**3.344**	**<0.01**
poster_is_gold: true	false	-0.065	0.064	-1.006	0.315
poster_num_submissions		**-0.010**	**0.002**	**-5.197**	**<0.01**
poster_total_karma		**-1.059×10^{-6}**	**1.84×10^{-8}**	**-57.384**	**<0.01**
poster_comment_karma		**9.549×10^{-7}**	**9.61×10^{-8}**	**9.934**	**<0.01**
poster_avg_score		**0.002**	**1.69×10^{-5}**	**125.405**	**<0.01**
poster_avg_upvote_ratios		**0.449**	**0.164**	**2.727**	**<0.01**
poster_awardee_karma		**10×10^{-5}**	**2.94×10^{-7}**	**396.676**	**<0.01**
poster_awarder_karma		**1.905×10^{-5}**	**7.2×10^{-6}**	**2.645**	**<0.01**
readability_flesch		**0.003**	**6.54×10^{-5}**	**38.641**	**<0.01**
text_num_characters		**2×10^{-4}**	**4.72×10^{-6}**	**31.917**	**<0.01**
sentiment_pos		0.012	0.028	0.412	0.680
sentiment_neg		-0.027	0.020	-1.311	0.190
tone_anger		0.036	0.114	0.315	0.753
tone_fear		-0.001	0.102	-0.005	0.996
tone_joy		0.052	0.066	0.783	0.433
tone_sadness		**0.365**	**0.067**	**5.483**	**<0.01**
tone_analytical		-0.073	0.039	-1.874	0.061
tone_confident		-0.094	0.069	-1.369	0.171
tone_tentative		**1.326**	**0.046**	**28.661**	**<0.01**

Table 6.2: Quantile regression model studying the popularity of security and privacy posts for the 0.25 quantile for the num_all_comments outcome.

was surprising because intuition and prior work may tell us that longer posts are not effective for capturing people's attention [148]. Other findings were that stronger emotions or sentiments were correlated with either the number of comments or votes, particularly when the value of the outcome was higher than the median outcome. For instance, a stronger positive sentiment or a stronger negative sentiment was associated with both outcomes (though the coefficient of `sentiment_neg` is negative, a decrease in the feature value indicates a stronger negative sentiment). The presence of specific emotions also saw higher engagement in the form of comments or votes when the number of comments or votes was the median (i.e., for the posts with "typical" popularity). The effects on the outcomes were small but the presence of these effects suggest that users who view a post may connect with them if they convey a strong emotion. Finally, visual attributes in the text were associated with more comments specifically for the posts with a very high number of comments.

Features describing the poster of the post

For features describing the posters of Reddit posts, several of them were consistently correlated with either of the two outcomes for all four quantiles (described below). However, some features describing poster popularity had a positive effect on the spread of security and privacy posts while others had negative effects.

- Posts made by posters with higher average scores were associated with both more comments and votes (0.9 quantile: coef. = 4.44, p <0.01). The effects on the number of comments and the rest of the quantiles for votes were marginal.
- Posts made by posters with higher total karma were associated with more comments and votes, although the effects of total karma were marginal for all quantiles and both outcomes.
- Posts made by posters with higher awarder karma were associated with more comments and votes, although the effect of awarder karma were marginal for all quantiles and both outcomes.
- Posts made by posters with higher awardee karma were associated with more comments, although the effects for all four quantiles were marginal.
- Posts made by posters who have made more post submissions on Reddit were associated with fewer votes, although the effects for all four quantiles for both outcomes were marginal.
- Posts made by posters with higher comment karma were associated with fewer votes, although the effects of comment karma were marginal for all quantiles.

While the above results were based on features that were consistent in their impact direction across all quantiles, one feature resulted in different impacts for each of the quantiles for one or both of the outcomes. Posts made by posters with higher average upvote ratios were associated with more comments (0.5 quantile: coef. = 1.02, p <0.01) but fewer votes (0.5 quantile: coef. = -30.55, p <0.01). The effects of a higher average upvote ratio were lower, marginal, or not significant for the other quantiles for both outcomes.

While prior work showed the positive impact of poster popularity on post popularity in social

networks [120, 155], we were surprised to discover that different metrics of poster popularity varied both positively and negatively with either one or both of the outcomes for the security and privacy posts we studied. For instance, posts made by posters with typically high engagement on their previous posts did not guarantee more engagement on the security and privacy posts we analyzed. While both the score and upvote ratio attributes of a post (see Section 6.3.2) signal the amount of positive reception to a post, a poster with a higher average score saw more engagement with posts (both comments and votes) particularly for posts with a high number of votes. Meanwhile, a poster with a higher average upvote ratio saw the same effect for their security and privacy in terms of the number of comments but saw a negative impact on the number of votes, with a particularly large negative effect on posts with the median number of votes. Similarly, different forms of posters' Reddit karma had varying effects on the outcomes. Higher total, awarder, and awardee karma of posters were correlated with their security and privacy posts achieving a higher spread. However, a higher comment karma was associated with a lower spread (i.e., fewer votes). The final and perhaps most surprising finding related to this feature set was that posters having made more overall posts were correlated with fewer votes but very marginally for all levels of vote count (i.e., the four quantiles).

Features describing other relevant properties of the post

Only one feature in this feature set, `has_link_flair`, was correlated with the number of comments for all quantiles. Though the feature was positively correlated for all quantiles, it had the greatest impact on the number of comments in the 0.5 quantile (coef. = 2.69, p <0.01). The effects of the feature for quantiles other than the median were marginal (i.e., the absolute difference in the outcome was less than one for a change in the feature value).

Similar to the features in the previous feature set, a few features in this feature set had different impacts on either the number of comments or votes for different quantiles. As before, we only report on significant features if the absolute value of the coefficient is greater than or equal to one.

- Posts that had links in them were usually associated with fewer comments (0.5 quantile: -1.09, p <0.01). The presence of links was marginal for the low popularity posts and not significant for the other two quantiles.
- Posts that only had an external link with no text were associated with fewer comments (0.5 quantile: coef. = -1.89, p <0.01) and more votes (0.5 quantile: coef. = 2.71, p <0.01). The effects of having a link in lieu of text were lower, marginal, or not significant for the other quantiles for both outcomes.
- Posts with author flair displayed were associated with more votes (0.75 quantile: coef. = 12.14, p <0.01). The effects of author flair were slightly lower for the 0.25 and 0.5 quantiles and not significant for the 0.9 quantile.

As with the significance of `num_emojis` in the first feature set, visual attributes such as flair on the post had a consistently positive effect on the number of comments when the number of comments was both low and high. Similarly, the presence of author flair next to a poster's username was associated with more votes but particularly for posts that had more comments than typical (i.e., when the number of comments was greater than the median). Interestingly, posts

including links (or when an external link was the only content in a post) saw fewer comments for the typical posts.

6.5 Comparison to popularity of baseline posts

In the previous section, we identified several properties of posts and their posters that were associated with posts receiving more engagement. To provide context as to how specific these findings are to security and privacy posts, we conducted a similar analysis over a set of baseline posts that are not only about security and privacy. We then compared the findings from both analyses. We first describe the baseline set of posts.

6.5.1 Collecting baseline posts

We collected posts that were about various topics (i.e., not restricted to security and privacy) by using Reddit's "random" API [26]. This random API returns a seemingly random sub-reddit or a seemingly random post within a specified sub-reddit (discussed further in Section 6.6). We collected one baseline post in the following way: (1) fetch a random sub-reddit using the API; (2) fetch a random post from the sub-reddit fetched in (1). We repeated this process multiple times such that we ended up with a set of baseline posts the same size as the set of security and privacy posts (i.e., 30.337 posts). If step (2) produced a duplicate of a post already fetched, we repeated step 2 until we encountered a non-duplicate post.

6.5.2 Analyzing the popularity of baseline posts

We extracted the features and outcomes described in Section 6.3 from these baseline posts and again analyzed the relationship between features and each outcome using quantile regression models for the 0.25, 0.5, 0.75, and 0.9 quantiles. Similarly to the previous analysis, we removed one feature from every pair of collinear features, posts with empty results about their poster, and when studying the `total_votes_estimate` outcome, posts for which we could not compute the total number of votes.

6.5.3 Results

Here, we describe results related to the significance of features in each of the three feature sets from Section 6.3.1 and how the results compare to those based on security and privacy posts. The quantile regression models for the `number_comments` outcome were computed over 30,067 posts after removing empty features. The model results for this outcome are described in Tables 16, 17, 18, and 19 in Appendix E. The models for the `total_votes_estimate` outcome were computed over 27,628 posts after removing posts with empty features and posts for which we could not estimate the vote count. The models results for this outcome are described in Tables 20, 21, 22, and 23 in Appendix E.

Features describing the content of the post

Only two features (described below) in this feature set were significant across all quantiles with consistent directions of the effect sizes (i.e., positive or negative) for either or both of the outcomes.

- Posts with a higher readability score were associated with more comments, although the effects of readability were marginal for all quantiles.
- Posts that were longer in length were associated with more comments and votes, although the effects for all four quantiles for both outcomes were marginal.

A number of features in this feature set showed different behavior for different quantiles for each of the outcomes. We describe these results below. Similarly to the analyses of security and privacy posts, we only report on correlations for a specific outcome if the absolute value of the coefficient (increase/decrease in outcome variable) for one of the quantile models is at least one.

- Posts with a stronger positive sentiment were associated with more comments (0.75 quantile: coef. = 1.8, p <0.01) and votes (0.9 quantile: coef. = 1.61, p = 0.04). The effects of positive sentiment were lower, marginal, or not significant for the other quantiles for both outcomes.
- Posts with a less strong negative sentiment were associated with fewer comments (0.75 quantile: -2.2, p <0.01) and votes (0.5 quantile: coef. = -2.04, p <0.01). However, a less strong negative sentiment was associated with more votes for the 0.9 quantile (coef. = 1.67, p = 0.03). The effects of negative sentiment were lower, marginal, or not significant for the other quantiles for both outcomes.
- Posts with an angry tone were associated with more comments (0.5 quantile: coef. = 1.85, p <0.01) and votes (0.5 quantile: coef. = 6.26, p <0.01). The effects of an angry tone in the text were lower or not significant for the other quantiles for both outcomes.
- Posts with a fearful tone were associated with fewer votes (0.5 quantile: coef. = -3.02, p <0.01). The effects of a fearful tone in the text were not significant for the other quantiles.
- Posts with a joyful tone were associated with more votes (0.5 quantile: coef. = 1.88, p <0.01). The effects of a joyful tone in the text were lower or not significant for the other quantiles.
- Posts with a sad tone were associated with fewer votes (0.5 quantile: coef. = -1.15, p = 0.03). The effects of a sad tone in the text were not significant for the other quantiles.
- Posts with an analytical tone were associated with fewer votes (0.5 quantile: coef. = -1.64, p <0.01). The effects of an analytical tone in the text were lower or not significant for the other quantiles.
- Posts with a confident tone were associated with more comments (0.5 quantile: coef. = 1.04, p <0.01). The effects of a confident tone in the text were not significant for the other quantiles.
- Posts with a tentative tone were associated with fewer votes (0.5 quantile: coef. = -2.1, p <0.01). The effects of a tentative tone in the text were lower or not significant for the other quantiles.

Similarly to the security and privacy posts, posts that were longer in length were associated with a higher spread for posts with a typical or high amount of engagement. While security and privacy post engagement was not correlated with readability, the baselines posts that were more readable were more likely to have a higher spread. Posts in the baseline set that had strong positive or negative sentiments or strong emotions also had impacts on the degree of spread for posts with a typical amount of engagement. However, stronger tones did not always have a positive impact on the amount of engagement a post received. In line with the findings related to security and privacy posts, angry and joyful tones were correlated with higher engagement while tentative tones, surprisingly, had the opposite impact than they did for security and privacy posts (i.e., negative impact). Of the features that were not significant for security and privacy posts, confident tones were correlated with more engagement for the baseline posts while fearful, sad, and analytical tones were associated with lower amounts of engagement.

Features describing the poster of the post

Three features in this feature set (described below) were significant across all four quantiles with consistent directions of the effect sizes for either or both outcomes.

- Posts made by posters with higher comment karma were associated with fewer comments, although the effects of comment karma were marginal for all quantiles.
- Posts made by posters with higher average scores were associated with both more comments and votes (0.9 quantile: coef. = 3.6, p <0.01). The effects on the number of comments and the lower quantiles for votes were marginal.
- Posts made by posters with higher awardee karma were associated with more comments and fewer votes, although the effects for all four quantiles for each outcome were marginal.

The following features were found to have differing impacts on the outcomes:

- Posts made by posters who were awarded gold status were associated with more comments (0.5 quantile: coef. = 2.35, p <0.01) and votes (0.5 quantile: coef. = 9.72, p <0.01). The effects of a poster having gold status were lower or not significant for the other quantiles for both outcomes.
- Posts made by posters who have made more overall post submissions on Reddit were associated with more comments (0.9 quantile: coef. = 1.06, p <0.01). The effects of a poster having made more Reddit posts in the past were marginal or not significant for the other quantiles.

Similarly to the popularity of security and privacy posts, metrics of poster popularity or activity on Reddit differently correlated with either or both of the outcomes. Consistently with those findings, posts made by posters with higher average scores and higher awardee karma were associated with more engagement with the average score having the most impact for the posts with a high amount of engagement. Also consistently with those findings, posts made by posters with higher comment karma were associated with lower engagement. In contrast with the security and privacy findings, however, posts made by posters who made more submissions were more likely to get more engagement (for posts with high popularity), although the outcome only increased by slightly more than one for every additional post made by the poster. Finally,

although it was not significant for the security and privacy posts, baseline posts made by gold posters were more likely to get engagement when the posts had typical popularity.

Features describing other relevant properties of the post

No features in this set were consistently significant with the same directions of impact across all quantiles for either outcome. However, multiple features (described below) were found to be correlated with either or both of the outcomes in different ways across the quantiles.

- Posts that only had an external link with no text were associated with fewer comments (0.25 quantile: coef. = -1.16, p <0.01) and more votes (0.5 quantile: coef. = 16.43, p <0.01). The effects of having a link in lieu of text were lower or not significant for the other quantiles for both outcomes.
- Posts that only had an image with no text were associated with more votes (0.5 quantile: coef. = 15.42, p <0.01). The effects of having an image in lieu of text were lower or not significant for the other quantiles.
- Posts with author flair displayed were associated with more comments (0.5 quantile: coef. = 1.81, p <0.01) and votes (0.5 quantile: coef. = 3.53, p <0.01). The effects of author flair were lower or not significant for the other quantiles for both outcomes.
- Posts with link flair displayed were associated with more comments (0.75 quantile: coef. = 1.18, p <0.01) and fewer votes (0.9 quantile: coef. = -3.07, p <0.01). However, the presence of link flair was associated with more votes for the 0.5 quantile with marginal effects. The effects of link flair were lower, marginal, or not significant for the other quantiles for both outcomes.
- Posts that were crossposted before were associated with fewer comments (0.5 quantile: coef. = -1.39, p <0.01). The effects of a post being crossposted before were not significant for the other quantiles.

A few of these findings are consistent with the results related to security and privacy posts. For instance, posts that only contained a link and no accompanying text garnered fewer comments and more votes, although this finding applies to the baseline posts with a low number of comments. Before conducting the baseline analyses, we found that `has_link` was correlated with `only_link_no_text` and hence we removed the former from the analysis. This implies that similarly to the security and privacy analyses, the presence of any links in the baseline posts were also correlated with fewer comments. Other mostly consistent findings include the presence of author and link flair being positively correlated with more engagement although the set of posts to which the findings apply (in terms of the quantile for which features were significant) were different between the analyses. However, the presence of link flair was anomalously negatively correlated with the number of votes for the posts with a high number of votes. Finally, two features that were not significant in the security and privacy analyses were significant with respect to the baseline posts. Posts that only had an image and no accompanying text were associated with more engagement. This was surprising because of the aforementioned finding that posts without text were less likely to gain popularity. Finally, posts that were crossposted before were associated with lower engagement. Both features were significant for the typically popular

posts.

6.6 Limitations

The work described in this chapter is subject to a few limitations.

We computed the total number of votes for a post by solving the two equations of score and upvote ratio for the individual upvote and downvote counts of each post. On each refresh or fetch of a Reddit post, the score and upvote ratios are computed and displayed after "fuzzing" the raw upvote and downvote counts in the system to prevent spam bots from verifying whether their votes count [25]. While these raw numbers may deviate minimally or drastically from their true values, the resulting scores and upvote ratios are said to be accurate [25]. Since we calculated the number of upvotes and downvotes from these two metrics and not from fuzzed individual votes, our findings are unlikely to be affected.

We were not able to compute the total number of votes for posts with a 50% upvote ratio or with a score of 0. Therefore, we removed such posts when analyzing the outcome of the total number of votes. Therefore, our dataset may not have sufficient representation of heavily downvoted posts or posts that garnered responses with perfectly neutral attitudes. However, we only removed 6% of posts on average between the security and privacy and baseline analyses because we could not compute their total votes. Additionally, neutral posts can still be represented by upvote ratios close to 50% without the ratio being exactly 50%. Therefore, we do not believe our results would have been significantly affected by this limitation. Similarly, we removed posts for which we could not retrieve poster information (likely because the users deleted their accounts). The percentage of posts removed because of this was 3% on average between the security and privacy and baseline analyses and therefore, is unlikely to have affected our analyses.

For every term we extracted related to security and privacy when collecting the security and privacy posts dataset, we were only able to collect upto 250 posts which is the maximum amount Reddit returned for a search term regardless of whether there were more relevant posts in Reddit's post history. As a result, our dataset may have missed some posts that were relevant to security and privacy even if Reddit determined them the least relevant to a given search term.

We used special tools to compute sentiment and tone in this study, specifically, the SentiStrength sentiment analyzer and the IBM Watson Tone Analyzer. While these two tools are considered to be highly accurate or state-of-the-art [24, 215], there is the potential for these tools to return inaccurate results or results inconsistent with human understanding of sentiment and tone. However, our findings about sentiment and tone were roughly consistent with prior work in other domains studying the impact of tone and sentiment on content popularity as well as findings from Chapter 3 of this thesis.

Finally, to conduct our baseline analyses, we used Reddit's "random" API. While Reddit does not provide official documentation on how it selects a random sub-reddit, Reddit discussions about this API suggest that the API pseudo-randomly selects a sub-reddit from a set of the top and active non-NSFW (Not Safe for Work) subreddits. We found that the baseline dataset contained 3,812 unique sub-reddits, which substantiates the likelihood of the sub-reddit being selected pseudo-randomly. Similarly, while Reddit does not provide official documentation for how the API selects random posts within a sub-reddit, 21% of the baseline posts had a score of 10 or less

which indicate low popularity (since Reddit uses the relative amount of upvotes in promoting posts' popularity as opposed to the total number of votes [7]) which suggest that the posts are not largely skewed towards higher popularity. The "random" API also favors more recent posts which may bias the timeline of the dataset; the earliest date in the baseline dataset is December 14th, 2014 while the earliest date in the security and privacy posts dataset is June 9th, 2005. However, we believe this baseline dataset is a reasonable approximation for a set of random Reddit posts suitable to be used as a baseline.

6.7 Discussion

In this chapter, we studied properties of posts about security and privacy on Reddit across multiple dimensions and how they are related to the posts' spread. We believe that our dataset covers a reasonably representative set of posts about security and privacy on Reddit and therefore, that we are well-positioned to identify actionable recommendations for distributing security and privacy content on Reddit and perhaps, social networks more generally. Moreover, we found that several of our findings were consistent with results related to general non-security and privacy posts while also finding that some features play a specific role in promoting the engagement of security and privacy posts.

6.7.1 Recommendations for creating and distributing security and privacy content

Based on insights from our findings related to the popularity of security and privacy content, we determined several recommendations for creating and sharing such content in a social network. We first discuss recommendations for creating individual posts that are intended to require minimal effort. We next discuss recommendations for creating posts that require a bit more effort but can still be implemented on a post-by-post basis. We then discuss higher-effort recommendations that could be implemented by posters of Reddit content over a longer term. Finally, we conclude with a discussion of insights about what the most popular and widely spread posts looked like.

Low-effort recommendations

Add visual attributes: Results from the models suggest that visual attributes on a post are correlated with higher spread. For example, visual attributes could be in the form of flair for which we measured the presence of two types, author flair and link flair. However, all subreddits may not have flair or allow users to set flair for themselves or for posts. Therefore, posters of security and privacy content may find it useful to apply flair whenever applicable to popularize their posts, specifically on Reddit.

Another type of visual attribute that correlated with spread specifically for the very popular posts were emojis. This is in line with prior work that found that images and visuals captured people's attention in social networks [139, 155, 163]. In particular, the more emojis there were in the post, the higher the engagement was. The number of emojis was important for the security

and privacy posts specifically as it was not significant for the baseline posts. Given this, posters of security and privacy content may try to use emojis or other visuals when applicable.

For general posts in the baseline set, adding images to a post increased its engagement. However, we did not find the presence of images in a post to be important for security and privacy posts' engagement.

Higher-effort recommendations

Write longer posts: Perhaps surprisingly, longer posts with a higher character count were correlated with higher spread. This could have been because longer posts were written out thoughtfully with helpful details [144] that made people inclined to engage with them. Therefore, providing as much as relevant information as possible when sharing content about security and privacy may be important for increasing its reach.

Convey strong sentiments and tones: In support of the above recommendation, the presence of clear emotions in posts was found to be associated with higher spread, particularly for the posts of typical popularity (measured by the model estimating the median or the 0.5 quantile) and the more popular posts. In particular, since a stronger negative or positive sentiment and stronger anger and joyful tones correlated with higher spread, it is likely that posts that convey a strong emotion evoke more responses and engagement. The effect of positive sentiment is also consistent with findings from Chapter 3. While angry and joyful tones were associated with higher engagement for security and privacy posts, most other tones were not correlated with higher spread. This is different from findings related to the baseline non-security and privacy posts. For the baseline posts, confident tones correlated with more engagement while fearful, sad, and analytical tones were correlated with lower engagement. Therefore, security and privacy posts may specifically benefit in terms of engagement from strong angry or joyful tones in their content.

Posts with a tentative tone that were perhaps explicitly or implicitly seeking guidance also correlated with higher spread. This was in contrast to the findings related to general posts (baseline posts); tentative tones were associated with lower spread for these posts. Therefore, seeking guidance appears to be particularly likely to generate responses and engagement for posts about security and privacy. Posters of security and privacy content could try to run their posts through online sentiment or tone analyzers (e.g., IBM Watson tone analyzer demo [18]) or attempt to ask a clear question while providing as much detailed information as is necessary to support the question.

Below are examples of posts with the aforementioned sentiments and emotions important for security and privacy content.

Negative sentiments: The following posts contained a strongly negative sentiment.

Cyber security experts hate Mr Rubio for this!

I recently found a file on my Mac called Adobe Flash Player.app.zip and have heard of the fake adobe pop ups. Was just thinking this morning about any security issues

and thought of that, so I searched my files and found it. I somehow must have fallen for this years ago and have no idea what all of this means if I installed this. I just know it's bad. ...

A terrifying stalking situation and the importance of internet safety!

Each of the above posts express either a dislike or a negative perception of a situation. These posts explicitly use adjectives and verbs that convey dislike such as "hate" or negative perceptions such as "bad" and "terrifying".

Positive sentiments: The following posts contained a strongly positive sentiment.

TryHackMe is awesome!

I don't live nowhere near this. I love privacy! iOS 15 B1

Brave Targets Google by Preparing to Launch Privacy-First Search Engine! Great news for BAT and user privacy

In these posts, the posters use adjectives or words that convey they like something or perceive something positively such as "awesome," "love," and "great".

Angry tones: The following posts contained strong anger tones.

Thanks, I hate Santa antivirus

What the fuck Facebook? Account-Privacy Settings-Applications and Websites - WHAT YOUR FRIENDS CAN SHARE ABOUT YOU.

I hate school software like this, it's a violation of privacy. Please don't support this shit.

Similarly to posts with negative sentiments, all the post examples here express strong dislike at some security-related phenomenon by using curse words or words such as "hate".

joyful tones: The following posts contained strong joyful tones.

Got me a Trezor. Super easy, transferred everything from blockchain.info... Security victory. Happy.

Very happy to finally see 2fa rolling out to Fitbit!

Cybersquatting at its best.

These posts all contain words that represent a strong like of something, similarly to the posts with a positive sentiment. For example, "happy" and "best" are words that convey positive or joyful tones even if the last post does not necessarily discuss a desirable phenomenon (i.e., cybersquatting).

Tentative tones: The following posts contained strong tentative tones.

Does anyone know anything about this? Spyware?

Seems like a phishing scam. Kind of curious about this one.

Why was "Incredimail" considered a privacy threat or PUP by AV software? ... Did it steal emails or browsing history or something?

As we suggested above, these examples each convey tentativeness by seeking guidance. This indicates that a post may get more engagement and be spread further when it requests responses through a question.

Avoid posts only containing external links: The presence of links especially when there was only a link and no text was generally correlated with a lower number of comments. In light of this finding and the above recommendations about longer text and emotion, posters of security and privacy content may create more popular posts if they provide more information or commentary and convey emotion in the text rather than just linking to external content.

Higher-effort recommendations over the longer term

Boost poster popularity over time: Though our findings suggest various relationships between posters' attributes and spread, it is not clear whether the popular posters tend to make "better" posts that get more engagement or if other users actually view the credentials of posters and then decide whether to vote or comment on a post. Whether the interpretation of the results is the former or the latter, posters can do their best to build up their popularity using the above recommendations when making posts. Though we were also surprised by the finding that posters with more submissions were negatively associated with lower spread on their security and privacy posts, this may suggest that quality over quantity is important. By focusing on creating quality posts in the above way, posters can build up their karma, awards received, gold status, and the popularity of their posts since each of these were found to have varying relationships with post spread.

6.7.2 Examining the most popular posts

To gain more insight into what makes posts popular and in turn, spread further, we manually examined the 500 posts in our dataset with the highest number of votes to understand what popular content about security and privacy looks like.

Many of the posts were announcements about a security phenomenon or just presenting information as opposed to asking questions. More than half of the 50 posts were making an announcement about either a security breach or incident or announcing an event related to security or privacy. Some of the posts were announcing data breaches or security incidents had titles such as the following:

Anonymous Hackers Target TikTok: 'Delete This Chinese Spyware Now'

Hackers breach Electronic Arts, stealing game source code and tools

However, not all announcement posts were about a specific security incident but sometimes were about studies, for example, the following title:

> *Studies reveal that location tracking apps do more than just monitor your whereabouts; they collect a lot of sensitive information about the user's residence, habits, interests, demographics, and personality traits*

Some posts discussed how companies are handling private data, for example:

> *Google to stop selling ads based on your browsing history and drop cookies support for Chrome citing privacy concerns*

A couple of these announcement posts were also constructive, for example:

> *We had a security incident. Here's what you need to know.*

Other times it seemed like the purpose of the post was to start a discussion, as in the post title below:

> *If Apple is the only organisation capable of defending our privacy, it really is time to worry.*

These examples along with the fact that the texts for many of these posts were longer than a few sentences seem to be in line with our findings specific to content length, sentiment, and emotion. Therefore, it seems that posts about security and privacy that are constructive and contain detailed helpful information have the potential to be the highest spread.

Chapter 7

Conclusions and future work

From the results derived from this thesis, it is apparent that people are not as engaged nor aware of security and privacy best practices. There is much room for improvement if users are to be properly equipped to keep themselves secure. Since security incidents are particularly common, it is important for people to be aware of these incidents as it will alert them to the potential dangers to their computers and data. However, in this thesis, we found that people often were either unaware of or unengaged with information regarding security incidents, even when the incident very likely affected them. Furthermore, they often did not take mitigating actions even when an incident affected them. These findings highlight a few important observations. First, people appear unlikely to become aware of incidents on their own. Therefore, particularly if a user is or may be affected by a security incident, the affected company or organization should be the party to inform their users not only of the occurrence of the incident but of additional steps to protect themselves beyond the affected organization. Second, relying on user awareness of security incidents to become cognizant of adopting security-enhancing behaviors may not be a tractable option given the low awareness we observed. This suggests that at a fundamental level, people require more direct assistance from the websites or companies that hold their user data. These parties have an obligation to consider their consumers' security from the start and implement measures to protect them without requiring them to make security-related decisions as a barrier to protecting their data. However, the state of today's digital world requires some degree of awareness on the part of computer users, particularly for protection against security issues without clear technical solutions (e.g., detecting phishing attacks). For most other aspects of security, the designers and maintainers of systems need to take measures to protect their users during the early stages of system design.

To promote the security awareness that is still needed in today's world using social media's ability to spread and promote information in other domains, we assessed the potential of social networks in spreading security awareness. Implementing interventions in social media involving showing users in a network what security-enhancing features their friends use was found to be an effective approach for encouraging security-enhancing behaviors. However, for social media to be an effective medium for spreading security awareness, there should be sufficient constructive security advice that people encounter on these platforms. We found that there were few discussions of security and privacy on social networks in practice. Of those few discussions, only 20% of the discussions contained constructive advice or anecdotes about improving one's security and

privacy. Therefore, it is unlikely that security awareness is effectively being increased through actual social media discussions outside of circles with security-aware users.

The few constructive discussions about security and privacy that we observed, again, support the aforementioned view that relying on user security awareness is not a viable approach to user security. However, social media has been shown to be a promising information dissemination channel across various domains. We hypothesized that its effectiveness in spreading security awareness could be increased by adopting approaches from other domains studying information spread (e.g., marketing) to increase the popularity and spread of security content. We found that several properties of posts helped promote content about other topics and were correlated with higher levels of engagement with security posts. For example, adding visual attributes to a post and conveying strong sentiments or emotions in the post may encourage more engagement with security and privacy content and in turn, increase the spread of this relevant information.

In summary, we, as a community that develops technologies for people, should do everything possible to protect users' security from the start. This should be carried out in a manner that does not burden users with the decision to act in order to be able to protect themselves. However, since some security awareness is still necessary, curators and distributors of security advice can take lessons from prior work in other domains. This thesis represents a key examination of how we can create more popular security- and privacy-related content to spread security awareness more effectively.

7.1 Future work

This section briefly describes future directions related to work presented in this thesis.

7.1.1 Follow-ups to studying and improving security information spread on Reddit and other networks

Findings from Chapter 6 give way to next steps that can help us better understand and improve the spread of security and privacy information in Reddit or other social networks. For example, future work could test through live experimentation how the spread of security information in a social network changes by altering properties of the content according to the findings from this chapter. This would involve actually posting content in a social network and measuring the spread over time. To measure spread, spread metrics such as those we studied in Chapter 6 can be used to describe the amount of engagement with content. However, metrics could also be computed by accounting for the number of distinct network communities the information reaches.

In Chapter 6, we also discussed the need for a higher volume of available security content so that security awareness is increased and in turn, security health is improved. However, people who already have some security knowledge are not likely to need as much education as people completely unexposed to security advice. Similarly, people who have more technical experience and backgrounds may benefit from advice that takes their background into account in contrast to advice that treats everyone as equal novices. Younger people may need less of an introduction to security dangers because they were likely introduced to the digital world at a young age whereas

elderly people might benefit from education that relates digital security to concepts they are familiar with. In general, a one-size-fits-all approach to designing security advice messages may not be effective. Therefore, a natural follow-up to the work described in this thesis is to understand how security education content can be tailored to effectively teach different demographic groups about security and privacy.

7.1.2 Localization of security information in social networks

To enable more users in a social network to receive security and privacy advice or information, it would be useful for future work to understand more about who the information is actually reaching within such a network. In particular, by characterizing online users within a full social network according to their interactions and the communities to which they belong in the network, we can understand whether there is any pattern to who is receiving security and privacy advice and information. Exploring this problem could involve studying whether interactors with security content in a network have varied interests according to the other content they interact with and studying whether security information is localized to specific communities within a social network. Such communities could be defined by their demographics or other common topics of interests. These avenues would provide insight into the role of social network structures in the spread of security and privacy information. For instance, if we observe that information is not extending beyond specific network communities, future work can specifically study how to propagate security and privacy through distinct communities within a network.

7.1.3 Security misinformation in social networks

As part of studying how security information is distributed in a network, it is important to also address the existence of security and privacy information that is misleading or incorrect (i.e., security and privacy misinformation). While throughout this thesis, we have been focusing largely on the spread of constructive security information, the spread of misinformation alongside this constructive advice may have polarizing effects on people's security and privacy knowledge. Thus, by studying the problems described in Section 7.1.2 but specifically how they relate to security misinformation, we can shed light on whether more efforts are needed to remove misinformation and promote constructive advice.

7.1.4 Using network influencers to increase information spread

Research in the social network analysis space has introduced the concept of influencers in a social network [117, 213]. Influencers in a network can be defined as the nodes with the maximum expected information spread [213]. One promising avenue of research for improving security awareness could be to explore how much the spread of security information would increase if it is shared by influential nodes in a social network (e.g., Instagram, YouTube). These influencers could be influencers in the technical sphere or outside to potentially achieve the maximum possible spread.

7.1.5 Incorporating lessons from health-related campaigns into cybersecurity campaigns

There have been a few cybersecurity campaigns attempted in the past [19, 198]. However, they have not been widespread or effective on a large scale as public security awareness is still low (as substantiated by the work in this thesis) [43]. On the other hand, as discussed in Section 2.7, public awareness campaigns in other domains such as public health and anti-smoking campaigns have been comparatively far-reaching, long-running, and successful [12, 47, 66, 94, 94, 105]. When designing security and privacy posts to contribute to a cybersecurity campaign on social media, we can take lessons learned from public health and anti-smoking campaigns and incorporate them into how we promote security advice. For example, fear appeals (persuasive messages that arouse fear) have also been found to be effective for both types of campaigns as long as individuals believe they are able to protect themselves [151, 226]. Future cybersecurity campaigns could similarly focus on the harm that may befall users related to their computers and online data as a result of poor security habits. In addition, to enable the self-efficacy required for fear appeals to be most effective, these campaigns must also convey that every user has the ability to stay secure by implementing safe practices. Security information campaigns could also improve their efficacy by tailoring messages to target populations that have different experiences with security and privacy (as also discussed in Section 7.1.1). For example, based on prior work [71], people who have had little exposure to security advice are likely to react to security advice in different ways than people who have had some exposure or prior knowledge. Future security and privacy campaigns could also conduct research on what types of media are effective for targeting specific demographics and should make use of traditional as well as modern forms of media. For example, public health campaigns have become successful on interactive media such as social media by posting messages in short videos and live streams [66]. Future security campaigns could test their effectiveness by conducting promotions through these or similar modern forms of media as well.

Bibliography

[1] What is multichannel retailing?: HCL technologies. *HCL Technologies*. URL https://www.hcltech.com/technology-qa/what-is-multichannel-retailing.

[2] textstat · PyPI. URL https://pypi.org/project/textstat/.

[3] Reddit coins - award and recognize community contributions. URL https://www.reddit.com/coins.

[4] Huge Equifax hack is even bigger than first thought. *Fortune*, 2017. URL http://fortune.com/2017/10/02/equifax-credit-breach-total/.

[5] Art. 33 GDPR. Notification of a personal data breach to the supervisory authority. *General Data Protection Regulation (GDPR)*, 2018. URL https://gdpr-info.eu/art-33-gdpr/.

[6] AWS — Alexa Web Information Service - Traffic metrics for any website, 2018. URL https://aws.amazon.com/awis/.

[7] How to make it to the front page of reddit, 2018. URL https://everyonesocial.com/blog/how-i-made-it-onto-the-front-page-of-reddit/.

[8] Facebook for Developers. 2018. URL https://developers.facebook.com/.

[9] Safe Browsing - Google Safe Browsing, 2018. URL https://safebrowsing.google.com/.

[10] Use cases, tutorials, & documentation — Twitter Developer. 2018. URL https://developer.twitter.com/en.

[11] VirusTotal - Home. 2018. URL https://www.virustotal.com/.

[12] CDC's anti-smoking ad campaign spurred over 100,000 smokers to quit; media campaigns must be expanded nationally and in the states, 2019. URL https://www.tobaccofreekids.org/press-releases/2013_09_09_cdc.

[13] Detect Compromised Credentials. Prevent fraud and ATO, 2019. URL https://www.enzoic.com/.

[14] Firefox Browser: Fast, Easy Password Manager, 2019. URL https://www.mozilla.org/en-US/firefox/features/password-manager/.

[15] Business Password Management — LastPass Business, 2019. URL https://

www.lastpass.com/products/business.

[16] Identity Theft Protection, 2019. URL https://www.lifelock.com/.

[17] Have I Been Pwned: Pwned websites, 2019. URL https://haveibeenpwned.com/PwnedWebsites.

[18] IBM Watson Tone Analyzer, 2019. URL https://www.ibm.com/cloud/watson-tone-analyzer.

[19] Get Safe Online, 2020. URL https://www.getsafeonline.org/.

[20] Introducing Instagram Reels. *Instagram*, 2020. URL https://about.instagram.com/blog/announcements/introducing-instagram-reels-announcement.

[21] 65% of people don't trust password managers despite 60% experiencing a data breach, 2020. URL https://www.passwordmanager.com/password-manager-trust-survey/.

[22] Image file type and format guide - Web media technologies: MDN, 2021. URL https://developer.mozilla.org/en-US/docs/Web/Media/Formats/Image_types.

[23] Security breach notification laws. *National Conference of State Legislatures*, 2021. URL https://www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notification-laws.aspx.

[24] IBM Watson Tone Analyzer review - Slant, 2021. URL https://www.slant.co/options/26844/~ibm-watson-tone-analyzer-review.

[25] faq - reddit.com, 2021. URL https://www.reddit.com/wiki/faq#wiki_how_is_a_comment.27s_score_determined.3F.

[26] reddit.com: api documentation, 2021. URL https://www.reddit.com/dev/api/#GET_random.

[27] Flair - Reddit Mods, 2021. URL https://mods.reddithelp.com/hc/en-us/articles/360002598912-Flair.

[28] karma - reddit.com, 2021. URL https://www.reddit.com/wiki/karma.

[29] search - reddit.com, 2021. URL https://www.reddit.com/r/reddit.com/wiki/search#wiki_limitations_and_caveats.

[30] Top 15 most popular news websites: March 2021. *eBizMBA — The eBusiness Guide*, 2021. URL http://www.ebizmba.com/articles/news-websites.

[31] Jacob Abbott, Daniel Calarco, and L. Jean Camp. Factors influencing password reuse: A case study. In *Research Conference on Communications, Information and Internet Policy*, 2018.

[32] Lillian Ablon, Paul Heaton, Diana Catherine Lavery, and Sasha Romanosky. *Consumer attitudes toward data breach notifications and loss of personal information*. Rand Corporation, 2016.

[33] 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. Nudges for privacy and security: Understanding and assisting users' choices online. *ACM Computing Surveys (CSUR)*, 2017.

[34] Anne Adams and Angela Sasse. Users are not the enemy. *Communications of the ACM*, 1999.

[35] Imtiaz Ahmad, Rosta Farzan, Apu Kapadia, and Adam J. Lee. Tangible privacy: Towards user-centric sensor designs for bystander privacy. *Computer Supported Cooperative Work and Social Computing (CSCW)*, 2020.

[36] Wasim Ahmed, Peter A Bath, Laura Sbaffi, and Gianluca Demartini. Measuring the effect of public health campaigns on twitter: The case of world autism awareness day. In *International Conference on Information*, 2018.

[37] Elham Al Qahtani, Mohamed Shehab, and Abrar Aljohani. The effectiveness of fear appeals in increasing smartphone locking behavior among Saudi Arabians. In *Symposium on Usable Privacy and Security (SOUPS)*, 2018.

[38] Nora Alkaldi and Karen Renaud. Why do people adopt, or reject, smartphone password managers? In *European Workshop on Usable Security (EuroUSEC)*, 2016.

[39] Manal Alohali, Nathan Clarke, Steven Furnell, and Saad Albakri. Information security behavior: Recognizing the influencers. In *Computing Conference*, 2017.

[40] Taylor Armerding. The 17 biggest data breaches of the 21st century. *CSO Online*, 2018. URL https://www.csoonline.com/article/2130877/data-breach/the-biggest-data-breaches-of-the-21st-century.html.

[41] David Armstrong, Ann Gosling, John Weinman, and Theresa Marteau. The place of inter-rater reliability in qualitative research: An empirical study. *Sociology*, 1997.

[42] Farzaneh Asgharpour, Debin Liu, and L. Jean Camp. Mental models of security risks. In *International Conference on Financial Cryptography and Data Security*, 2007.

[43] Maria Bada, Angela Sasse, and Jason Nurse. Cyber security awareness campaigns: Why do they fail to change behaviour? In *International Conference on Cyber Security for Sustainable Society*, 2015.

[44] Younggue Bae and Hongchul Lee. A sentiment analysis of audiences on twitter: Who is the positive or negative audience of popular twitterers? In *International Conference on Hybrid Information Technology*. Springer, 2011.

[45] A.S. Bager. Ridge parameter in quantile regression models: An application in biostatistics. *International Journal of Statistics and Applications*, 2018.

[46] Eytan Bakshy, Itamar Rosenn, Cameron Marlow, and Lada Adamic. The role of social networks in information diffusion. In *International conference on World Wide Web (WWW)*, 2012.

[47] Ambar Basu and Jian Wang. The role of branding in public health campaigns. *Journal of Communication Management*, 2009.

[48] BBC News. Panama Papers Q & A: What is the scandal about? *BBC News*, 2016. URL https://www.bbc.com/news/world-35954224.

[49] BBC News. Cyber-attack: Europol says it was unprecedented in scale. *BBC News*, 2017. URL https://www.bbc.com/news/world-europe-39907965.

[50] Simon Bell and Peter Komisarczuk. An analysis of phishing blacklists: Google safe browsing, openphish, and phishtank. In *Australasian Computer Science Week Multiconference*, 2020.

[51] Lucia Benaquisto and Lisa M. Given. The sage encyclopedia of qualitative research methods. *New York: Sage*, 2008.

[52] Sruti Bhagavatula, Lujo Bauer, and Apu Kapadia. (How) Do people change their passwords after a breach? *arXiv preprint arXiv:2010.09853*, 2020.

[53] Sruti Bhagavatula, Lujo Bauer, and Apu Kapadia. (How) Do people change their passwords after a breach? In *Workshop on Technology and Consumer Protection*, 2020.

[54] Sruti Bhagavatula, Lujo Bauer, and Apu Kapadia. What breach? Measuring online awareness of security incidents by studying real-world browsing behavior. In *Workshop on Technology and Consumer Protection*, 2021.

[55] Sruti Bhagavatula, Lujo Bauer, and Apu Kapadia. What breach? Measuring online awareness of security incidents by studying real-world browsing behavior. In *European Symposium on Usable Security (EuroUSEC)*, 2021. To appear.

[56] Sruti Bhagavatula, Lujo Bauer, and Apu Kapadia. "Adulthood is trying each of the same six passwords that you use for everything": The scarcity and ambiguity of security advice on social media. In *Computer Supported Cooperative Work (CSCW)*, 2021. Under submission.

[57] Saumik Bhattacharya, Kumar Gaurav, and Sayantari Ghosh. Viral marketing on social networks: An epidemiological perspective. *Physica A: Statistical Mechanics and its Applications*, 2019.

[58] J. Martin Bland and Douglas G. Altman. Multiple significance tests: The Bonferroni method. *The BMJ: Leading general medical journal*, 1995.

[59] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 2003.

[60] Donna Borak and Kathryn Vasel. The Equifax hack could be worse than we thought. *CNNMoney*, 2018. URL https://money.cnn.com/2018/02/09/pf/equifax-hack-senate-disclosure/index.html.

[61] Ed Bott. Latest OS share data shows Windows still dominating in PCs. *ZDNet*, 2013. URL https://www.zdnet.com/article/latest-os-share-data-shows-windows-still-dominating-in-pcs/.

[62] Kevin J. Boudreau, Nicola Lacetera, and Karim R. Lakhani. Incentives and problem uncertainty in innovation contests: An empirical analysis. *Management Science*, 2011.

[63] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative*

research in psychology, 2006.

[64] Jon Bruner. Tweets loud and quiet. *O'Reilly Media, Inc.*, 2013. URL https://www.oreilly.com/content/tweets-loud-and-quiet/.

[65] Jan-Willem H Bullée, Lorena Montoya, Wolter Pieters, Marianne Junger, and Pieter H. Hartel. The persuasion and security awareness experiment: Reducing the success of social engineering attacks. *Journal of experimental criminology*, 2015.

[66] Crispin Butteriss. The world's best public health social media campaigns. *Bang the Table*, 2021. URL https://www.bangthetable.com/blog/public-health-social-media-campaigns/.

[67] Casey Canfield, Alex Davis, Baruch Fischhoff, Alain Forget, Sarah Pearman, and Jeremy Thomas. Replication: Challenges in using data logs to validate phishing detection ability metrics. In *Symposium on Usable Privacy and Security (SOUPS)*, volume 12, 2017.

[68] Heather Catalano. Best free credit report site of 2018. *The Simple Dollar*, 2018. URL https://www.thesimpledollar.com/best-free-credit-report-site/.

[69] Sean Cavanagh. Education company Chegg acknowledges data breach, puts 40 million users on notice. *Market Brief*, 2018. URL https://marketbrief.edweek.org/marketplace-k-12/tutoring-company-chegg-acknowledges-data-breach-puts-40-million-users-notice/.

[70] Stephane Champely. *pwr: Basic Functions for Power Analysis*, 2018. URL https://CRAN.R-project.org/package=pwr. R package version 1.2-2.

[71] Hyunyi Cho and Charles T. Salmon. Fear appeals for individuals in different stages of change: Intended and unintended effects and implications on public health campaigns. *Health Communication*, 2006.

[72] Lorrie Faith Cranor and Simson Garfinkel. *Security and usability: Designing secure systems that people can use*. O'Reilly Media, Inc., 2005.

[73] Heidi Daitch. 2017 data breaches - the worst breaches, so far. *IdentityForce*, 2017. URL https://www.identityforce.com/blog/2017-data-breaches.

[74] Thomas A. Daniel and Alecka L. Camp. Emojis affect processing fluency on social media. *Psychology of Popular Media*, 2020.

[75] Anupam Das, Joseph Bonneau, Matthew Caesar, Nikita Borisov, and XiaoFeng Wang. The tangled web of password reuse. In *Network and Distributed System Security Symposium (NDSS)*, 2014.

[76] Sauvik Das, Tiffany Hyun-Jin Kim, Laura A. Dabbish, and Jason I. Hong. The effect of social influence on security sensitivity. In *Symposium on Usable Privacy and Security (SOUPS)*, 2014.

[77] Sauvik Das, Adam D.I. Kramer, Laura A. Dabbish, and Jason I. Hong. Increasing security sensitivity with social proof: A large-scale experimental confirmation. In *ACM Conference on Computer and Communications Security (CCS)*, 2014.

[78] Sauvik Das, Adam D.I. Kramer, Laura A. Dabbish, and Jason I. Hong. The role of social influence in security feature adoption. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, 2015.

[79] Sauvik Das, Joanne Lo, Laura A. Dabbish, and Jason I. Hong. Breaking! A typology of security and privacy news and how it's shared. In *Conference on Human Factors in Computing Systems (CHI)*, 2018.

[80] Rachna Dhamija and Lisa Dusseault. The seven flaws of identity management: Usability and security challenges. *IEEE Security & Privacy*, 2008.

[81] Paul DiGioia and Paul Dourish. Social navigation as a model for usable security. In *Symposium on Usable Privacy and Security (SOUPS)*, 2005.

[82] William J. Doll, T.S. Raghunathan, Jeen-Su Lim, and Yash P. Gupta. A confirmatory factor analysis of the user information satisfaction instrument. *Information Systems Research*, 1995.

[83] Geoffrey B. Duggan, Hilary Johnson, and Beate Grawemeyer. Rational security: Modelling everyday password use. *International journal of human-computer studies*, 2012.

[84] Olive J. Dunn. Multiple comparisons among means. *Journal of the American statistical association*, 1961.

[85] Paul Dunphy, Vasilis Vlachokyriakos, Anja Thieme, James Nicholson, John C. McCarthy, and Patrick Olivier. Social media as a resource for understanding security experiences: A qualitative analysis of #password tweets. In *Symposium on Usable Privacy and Security (SOUPS)*, 2015.

[86] Serge Egelman and Eyal Peer. Scaling the security wall: Developing a security behavior intentions scale (sebis). In *Conference on Human Factors in Computing Systems (CHI)*, 2015.

[87] Paul D. Ellis. *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge University Press, 2010.

[88] Pardis Emami Naeini, Martin Degeling, Lujo Bauer, Richard Chow, Lorrie Faith Cranor, Mohammad Reza Haghighat, and Heather Patterson. The influence of friends and experts on privacy decision making in IoT scenarios. *Computer Supported Cooperative Work and Social Computing (CSCW)*, 2018.

[89] Benjamin Fabian, Tatiana Ermakova, and Tino Lentz. Large-scale readability analysis of privacy policies. In *Proceedings of the international conference on web intelligence*, 2017.

[90] Michael Fagan, Yusuf Albayram, Mohammad Maifi Hasan Khan, and Ross Buck. An investigation into users' considerations towards using password managers. *Human-centric Computing and Information Sciences*, 2017.

[91] Amy L. Fairchild, Ronald Bayer, and James Colgrove. Risky business: New york city?s experience with fear-based public health campaigns. *Health Affairs*, 2015.

[92] Xitao Fan, Brent C. Miller, Kyung-Eun Park, Bryan W. Winward, Mathew Christensen, Harold D. Grotevant, and Robert H. Tai. An exploratory study about inaccuracy and invalidity in adolescent self-report surveys. *Field Methods*, 2006.

[93] Paul Fipps. Important message regarding MyFitnessPal account security. *MyFitnessPal*, 2018. URL https://content.myfitnesspal.com/security-information/notice.html.

[94] Brian R. Flay and Dee Burton. Effective mass communication strategies for health campaigns. 1990.

[95] Rudolph Flesch. A new readability yardstick. *Journal of applied psychology*, 1948.

[96] Dinei Florencio and Cormac Herley. A large-scale study of web password habits. In *International conference on World Wide Web (WWW)*, 2007.

[97] Alain Forget, Saranga Komanduri, Alessandro Acquisti, Nicolas Christin, Lorrie Faith Cranor, and Rahul Telang. Building the security behavior observatory: An infrastructure for long-term monitoring of client machines. In *Symposium and Bootcamp on the Science of Security (HotSoS)*, 2014.

[98] Alain Forget, Sarah Pearman, Jeremy Thomas, Alessandro Acquisti, Nicolas Christin, Lorrie Faith Cranor, Serge Egelman, Marian Harbach, and Rahul Telang. Do or do not, there is no try: User engagement may not improve security outcomes. In *Symposium on Usable Privacy and Security (SOUPS)*, 2016.

[99] Josh Fruhlinger. The 6 biggest ransomware attacks of the last 5 years. *CSO*, 2019. URL https://www.csoonline.com/article/3212260/the-5-biggest-ransomware-attacks-of-the-last-5-years.html.

[100] Simson Garfinkel and Heather Richter Lipford. Usable security: History, themes, and challenges. *Synthesis Lectures on Information Security, Privacy, and Trust*, 2014.

[101] Vaibhav Garg, L. Jean Camp, Katherine Connelly, and Lesa Lorenzen-Huber. Risk communication design: Video vs. text. In Simone Fischer-Hübner and Matthew Wright, editors, *Privacy Enhancing Technologies*, 2012.

[102] Shirley Gaw, Edward W. Felten, and Patricia Fernandez-Kelly. Secrecy, flagging, and paranoia: Adoption criteria in encrypted email. In *Conference on Human Factors in Computing Systems (CHI)*, 2006.

[103] C.J. Hutto Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *International Conference on Weblogs and Social Media (ICWSM)*, 2014.

[104] Vindu Goel and Nicole Perlroth. Yahoo says 1 billion user accounts were hacked. *The New York Times*, 2016. URL https://www.nytimes.com/2016/12/14/technology/yahoo-hack.html.

[105] Lisa K. Goldman and Stanton A. Glantz. Evaluation of antismoking advertising campaigns. *JAMA*, 1998.

[106] Maximilian Golla, Miranda Wei, Juliette Hainline, Lydia Filipe, Markus Dürmuth, Elissa M. Redmiles, and Blase Ur. What was that site doing with my facebook password?: Designing password-reuse notifications. In *ACM Conference on Computer and Communications Security (CCS)*, 2018.

[107] Nathaniel Good, Rachna Dhamija, Jens Grossklags, David Thaw, Steven Aronowitz,

Deirdre Mulligan, and Joseph Konstan. Stopping spyware at the gate: A user study of privacy, notice and spyware. In *Symposium on Usable Privacy and Security (SOUPS)*, 2005.

[108] Edward C. Green and Kim Witte. Can fear arousal in public health campaigns contribute to the decline of hiv prevalence? *Journal of Health Communication*, 2006.

[109] Adrien Guille, Hakim Hacid, Cecile Favre, and Djamel A. Zighed. Information diffusion in online social networks: A survey. *ACM SIGMOD Record*, 2013.

[110] Peter Gutmann and Ian Grigg. Security usability. *IEEE security & privacy*, 2005.

[111] Hana Habib, Jessica Colnago, Vidya Gopalakrishnan, Sarah Pearman, Jeremy Thomas, Alessandro Acquisti, Nicolas Christin, and Lorrie Faith Cranor. Away from prying eyes: Analyzing usage and understanding of private browsing. In *Symposium on Usable Privacy and Security (SOUPS)*, 2018.

[112] Hana Habib, Pardis Emami Naeini, Summer Devlin, Maggie Oates, Chelse Swoopes, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. User behaviors and attitudes under password expiration policies. In *Symposium on Usable Privacy and Security (SOUPS)*, 2018.

[113] Suraya Hamid, Mohamad Taha Ijab, Hidayah Sulaiman, Rina Md Anwar, and Azah Anir Norman. Social media for environmental sustainability awareness in higher education. *International Journal of Sustainability in Higher Education*, 2017.

[114] Ameya Hanamsagar, Simon S. Woo, Chris Kanich, and Jelena Mirkovic. Leveraging semantic transformation to investigate password habits and their causes. In *Conference on Human Factors in Computing Systems (CHI)*, 2018.

[115] Bartlomiej Hanus and Yu "Andy" Wu. Impact of users' security awareness on desktop security behavior: A protection motivation theory perspective. *Information Systems Management*, 2016.

[116] Luke Harding. What are the Panama Papers? A guide to history's biggest data leak. *The Guardian*, 2016. URL https://www.theguardian.com/news/2016/apr/03/what-you-need-to-know-about-the-panama-papers.

[117] Paul Harrigan, Timothy M. Daly, Kristof Coussement, Julie A. Lee, Geoffrey N. Soutar, and Uwana Evers. Identifying influencers on social media. *International Journal of Information Management*, 2021.

[118] Cormac Herley. So long, and no thanks for the externalities: The rational rejection of security advice by users. In *Workshop on New Security Paradigms Workshop*, 2009.

[119] Benjamin Herold. Florida virtual school reveals huge data breaches. *Education Week - Digital Education*, 2018. URL http://blogs.edweek.org/edweek/DigitalEducation/2018/03/florida_virtual_school_data_breaches.html.

[120] Liangjie Hong, Ovidiu Dan, and Brian D. Davison. Predicting popular messages in twitter. In *International conference on World Wide Web (WWW)*, 2011.

[121] Matthew Honnibal and Mark Johnson. An improved non-monotonic transition system for

dependency parsing. In *Conference on Empirical Methods in Natural Language Processing*, 2015.

[122] Roy Horev. The top 7 vulnerabilities of the decade. *Vulcan*, 2018. URL https://blog.vulcan.io/top-7-vulnerabilities.

[123] Archie Hughes-Hallett, Daisy Browne, Elsie Mensah, Justin Vale, and Erik Mayer. Assessing the impact of mass media public health campaigns. be clear on cancer 'blood in pee': A case in point. 2015.

[124] Bin Ju, Mincao Ye, Yuntao Qian, Rong Ni, and Chenxi Zhu. Modeling behaviors of browsing and buying for alidata discovery using joint non-negative matrix factorization. In *International Conference on Computational Intelligence and Security*, 2014.

[125] Johannes Kaiser and Martin Reichenbach. Evaluating security tools towards usable security: A usability taxonomy for the evaluation of security tools based on a categorization of user errors. 2002.

[126] Sowmya Karunakaran, Kurt Thomas, Elie Bursztein, and Oxana Comanescu. Data breaches: User comprehension, expectations, and concerns with handling exposed data. In *Symposium on Usable Privacy and Security (SOUPS)*, 2018.

[127] Kassambara. Linear regression assumptions and diagnostics in R: Essentials. *STHDA*, Mar 2018. URL http://www.sthda.com/english/articles/39-regression-model-diagnostics/161-linear-regression-assumptions-and-diagnostics-in-r-essentials/.

[128] Kassambara. Logistic regression assumptions and diagnostics in R. *STHDA*, 2018. URL http://www.sthda.com/english/articles/36-classification-methods-essentials/148-logistic-regression-assumptions-and-diagnostics-in-r/.

[129] Amandeep Kaur and H.S. Chahal. Role of social media in increasing environmental issue awareness. *Researchers World*, 2018.

[130] Judy Kendall. Axial coding and the grounded theory controversy. *Western journal of nursing research*, 1999.

[131] Iacovos Kirlappos and Angela Sasse. Security education against phishing: A modest proposal for a major rethink. *IEEE Security & Privacy*, 2011.

[132] Roger Koenker and Gilbert Bassett Jr. Regression quantiles. *Econometrica: Journal of the Econometric Society*, 1978.

[133] Andrey Kolmogorov. Sulla determinazione empirica di una lgge di distribuzione. *Inst. Ital. Attuari, Giorn.*, 1933.

[134] Brian Krebs. Krebs on security: Deloitte breach affected all company email, admin accounts. *Krebs on Security*, 2017. URL https://krebsonsecurity.com/2017/09/source-deloitte-breach-affected-all-company-email-admin-accounts/.

[135] Elmarie Kritzinger and S.H. Solms. Cyber security for home users: A new way of protection through awareness enforcement. *Computers & Security*, 2010.

[136] Barbara Krumay and Jennifer Klar. Readability of privacy policies. In *IFIP Annual Conference on Data and Applications Security and Privacy*, 2020.

[137] William H. Kruskal and W. Allen Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 1952.

[138] Mohit Kumar. Yahoo hacked once again! Quietly warns affected users about new attack. *The Hacker News*, 2017. URL https://thehackernews.com/2017/02/yahoo-hack.html.

[139] Tal Laor. "Hello, is there anybody who reads me?" Radio programs and popular facebook posts. *International Journal of Interactive Multimedia & Artificial Intelligence*, 5 (7), 2019.

[140] Selena Larson. 10 biggest hacks of 2017. *CNNMoney*, 2017. URL https://money.cnn.com/2017/12/18/technology/biggest-cyberattacks-of-the-year/index.html.

[141] Selena Larson. Uber's massive hack: What we know. *CNNMoney*, 2017. URL https://money.cnn.com/2017/11/22/technology/uber-hack-consequences-cover-up/index.html.

[142] Selena Larson. Every single Yahoo account was hacked. *CNNMoney*, 2017. URL https://money.cnn.com/2017/10/03/technology/business/yahoo-breach-3-billion-accounts/index.html.

[143] Daniel D. Lee and H. Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 1999.

[144] Frank Li, Zakir Durumeric, Jakub Czyz, Mohammad Karami, Michael Bailey, Damon McCoy, Stefan Savage, and Vern Paxson. You've got vulnerability: Exploring effective vulnerability notifications. In *USENIX Security Symposium)*, 2016.

[145] Peter Likarish and Eunjin Jung. Leveraging google SafeBrowsing to characterize web-based attacks. *Association for Computing Machinery*, 2009.

[146] Natasha Lomas. Imgur says 1.7m emails and passwords were breached in 2014 hack. *TechCrunch*, 2017. URL https://techcrunch.com/2017/11/27/imgur-says-1-7m-emails-and-passwords-were-breached-in-2014-hack/.

[147] Nate Lord. A timeline of the Ashley Madison hack. *Digital Guardian*, 2017. URL https://digitalguardian.com/blog/timeline-ashley-madison-hack.

[148] Arvind Malhotra, Claudia K. Malhotra, and Alan See. How to create brand engagement on facebook. *MIT Sloan Management Review*, 2013.

[149] Leslie Malone. Computer and internet use in the United States: 2018. *Computer and Internet Use in the United States: 2018*. URL https://www.census.gov/newsroom/press-releases/2021/computer-internet-use.html.

[150] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. 2008.

[151] Simon Manyiwa and Ross Brennan. Fear appeals in anti-smoking advertising: How important is self-efficacy? *Journal of Marketing Management*, 2012.

[152] Rima Masri and Monther Aldwairi. Automated malicious advertisement detection using virustotal, urlvoid, and trendmicro. In *2017 8th International Conference on Information and Communication Systems (ICICS)*, 2017.

[153] Arunesh Mathur and Marshini Chetty. Impact of user characteristics on attitudes towards automatic mobile application updates. In *Symposium on Usable Privacy and Security (SOUPS)*, 2017.

[154] Peter Mayer, Yixin Zou, Florian Schaub, and Adam J. Aviv. "Now I'm a bit angry:" Individuals' awareness, perception, and responses to data breaches that affected them. In *USENIX Security Symposium*, 2021.

[155] Masoud Mazloom, Robert Rietveld, Stevan Rudinac, Marcel Worring, and Willemijn Van Dolen. Multimodal popularity prediction of brand-related social media posts. In *Proceedings of the 24th ACM international conference on Multimedia*, 2016.

[156] Michelle L. Mazurek, Saranga Komanduri, Timothy Vidas, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, Patrick Gage Kelley, Richard Shay, and Blase Ur. Measuring password guessability for an entire university. In *ACM Conference on Computer and Communications Security (CCS)*, 2013.

[157] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for cscw and hci practice. *Computer Supported Cooperative Work and Social Computing (CSCW)*, 2019.

[158] Mary Meeker. Internet trends report 2018. *Kleiner Perkins*, 2018. URL https://www.kleinerperkins.com/perspectives/internet-trends-report-2018/.

[159] William Melicher, Blase Ur, Sean M. Segreti, Saranga Komanduri, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. Fast, lean, and accurate: Modeling password guessability using neural networks. In *USENIX Security Symposium*, 2016.

[160] Julia Mendelsohn and Lucy Li. Giving gold: Understanding appreciation in reddit communities. Technical report, Working paper, Stanford University, Stanford, CA, 2017.

[161] Renae Merle. Yahoo fined $35 million for failing to disclose cyber breach. *The Washington Post*, 2019. URL https://www.washingtonpost.com/news/business/wp/2018/04/24/yahoo-fined-35-million-for-failing-to-disclose-cyber-breach/.

[162] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[163] Gillian Moran, Laurent Muzellec, and Devon Johnson. Message content features and social media engagement: evidence from the media industry. *Journal of Product & Brand Management*, 2019.

[164] Julian M. Mueller-Herbst, Michael A. Xenos, Dietram A. Scheufele, and Dominique Brossard. Saw it on facebook: The role of social media in facilitating science issue aware-

ness. *Social Media+ Society*, 2020.

[165] Stanley A. Mulaik. *Foundations of factor analysis*. 2009.

[166] Francis Navarro. Ancestry.com suffers big data leak - 300,000 user credentials exposed. *The Kim Komando Show*, 2017. URL https://www.komando.com/happening-now/435921/ancestry-com-suffers-big-data-leak-300000-user-credentials-exposed.

[167] James Nicholson, Lynne Coventry, and Pam Briggs. Can we fight social engineering attacks by social means? Assessing social salience as a means to improve phish detection. In *Symposium on Usable Privacy and Security (SOUPS)*, 2017.

[168] James Nicholson, Lynne Coventry, and Pamela Briggs. If it's important it will be a headline: Cybersecurity information seeking in older adults. In *Conference on Human Factors in Computing Systems (CHI)*, 2019.

[169] Rishab Nithyanand, Brian Schaffner, and Phillipa Gill. Online political discourse in the trump era. *arXiv preprint arXiv:1711.05303*, 2017.

[170] Philippe Oechslin. Making a faster cryptanalytic time-memory trade-off. In *Annual International Cryptology Conference*, 2003.

[171] Kaan Onarlioglu, Utku Ozan Yilmaz, Engin Kirda, and Davide Balzarotti. Insights into user behavior in dealing with internet attacks. In *Network and Distributed System Security Symposium (NDSS)*, 2012.

[172] Lucas Ou-Yang. Newspaper3k: Article scraping & curation: http://newspaper.readthedocs.io/en/latest/. *Python Software Foundation*, 2013. URL http://newspaper.readthedocs.io/en/latest/.

[173] Pierluigi Paganini. Cashcrate cash-for-surveys site breached, 6 million accounts stolen. *Security Affairs*, 2017. URL https://securityaffairs.co/wordpress/60083/data-breach/cashcrate-data-breach.html.

[174] Sarah Pearman, Jeremy Thomas, Pardis Emami Naeini, Hana Habib, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, Serge Egelman, and Alain Forget. Let's go in for a closer look: Observing passwords in their natural habitat. In *ACM Conference on Computer and Communications Security (CCS)*, 2017.

[175] Sarah Pearman, Aerin Shikun Zhang, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. Why people (don't) use password managers effectively. In *Symposium on Usable Privacy and Security (SOUPS)*, 2019.

[176] Peng Peng, Limin Yang, Linhai Song, and Gang Wang. Opening the blackbox of virustotal: Analyzing online phishing scan engines. In *Proceedings of the Internet Measurement Conference*, 2019.

[177] Yilang Peng. What makes politicians? instagram posts popular? analyzing social media strategies of candidates and office holders with computer vision. *The International Journal of Press/Politics*, 2021.

[178] Marco Pennacchiotti and Siva Gurumurthy. Investigating topic models for social media user recommendation. In *International conference on World Wide Web (WWW)*, 2011.

[179] Nicole Perlroth. Yahoo says hackers stole data on 500 million users in 2014. *The New York Times*, 2016. URL https://www.nytimes.com/2016/09/23/technology/yahoo-hackers.html.

[180] Andrew Perrin and Monica Anderson. Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018. *Pew Research Center*, 2019. URL https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/.

[181] Jakob Petersen, Maurizio Gibin, Paul Longley, Pablo Mateos, Philip Atkinson, and David Ashby. Geodemographics as a tool for targeting neighbourhoods in public health campaigns. *Journal of Geographical Systems*, 2011.

[182] Kristopher J. Preacher and Robert C. MacCallum. Repairing tom swift's electric factor analysis machine. *Understanding statistics: Statistical issues in psychology, education, and the social sciences*, 2003.

[183] James O. Prochaska and Carlo C. DiClemente. Stages of change in the modification of problem behaviors. *Progress in Behavior Modification*, 1992.

[184] Emilee Rader and Rick Wash. Identifying patterns in informal sources of security information. *Journal of Cybersecurity*, 2015.

[185] Emilee J. Rader, Rick Wash, and Brandon Brooks. Stories as informal lessons about security. In *Symposium on Usable Privacy and Security (SOUPS)*, 2012.

[186] Whitney Randolph and Kasisomayajula Viswanath. Lessons learned from public health mass media campaigns: Marketing health in a crowded media world. *Annu. Rev. Public Health*, 2004.

[187] Steve Ranger. Ransomware attack: The clean-up continues after WannaCry chaos. *ZDNet*. URL https://www.zdnet.com/article/ransomware-attack-the-clean-up-continues-after-wannacry-chaos/.

[188] Elissa M. Redmiles. "Should I worry?" A cross-cultural examination of account security incident response. In *2019 IEEE Symposium on Security and Privacy (SP)*, 2019.

[189] Elissa M. Redmiles, Sean Kross, and Michelle L. Mazurek. How I learned to be secure: A census-representative survey of security advice sources and behavior. In *ACM Conference on Computer and Communications Security (CCS)*, 2016.

[190] Elissa M. Redmiles, Amelia R. Malone, and Michelle L. Mazurek. I think they're trying to tell me something: Advice sources and selection for digital security. In *IEEE Symposium on Security and Privacy (SP)*, 2016.

[191] Elissa M. Redmiles, Lisa Maszkiewicz, Emily Hwang, Dhruv Kuchhal, Everest Liu, Miraida Morales, Denis Peskov, Sudha Rao, Rock Stevens, Kristina Gligorić, Sean Kross, Michelle L. Mazurek, and Hal Daumé III. Comparing and developing tools to measure the readability of domain-specific texts. In *Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019.

[192] Elissa M. Redmiles, Noel Warford, Amritha Jayanti, Aravind Koneru, Sean Kross, Miraida Morales, Rock Stevens, and Michelle L. Mazurek. A comprehensive quality evaluation of security and privacy advice on the web. In *USENIX Security Symposium*, 2020.

[193] Diana Rofail, Antje Colligs, Linda Abetz, Marjaana Lindemann, and Laura Maguire. Factors contributing to the success of folic acid public health campaigns. *Journal of Public Health*, 2012.

[194] Robert Rosenman, Vidhura Tennekoon, and Laura G. Hill. Measuring bias in self-reported data. *International journal of behavioural & healthcare research*, 2011.

[195] Koustuv Saha, John Torous, Sindhu Kiranmai Ernala, Conor Rizuto, Amanda Stafford, and Munmun De Choudhury. A computational study of mental health awareness campaigns on social media. *Translational behavioral medicine*, 2019.

[196] Saima Salim. Revealed: The 21 biggest data breaches of 2018. *Digital Information World*, 2018. URL https://www.digitalinformationworld.com/2018/12/biggest-data-breaches-of-2018.html.

[197] Angela Sasse and Ivan Flechais. Usable security: Why do we need it? how do we get it? *O'Reilly Media, Inc.*, 2005.

[198] Matthew W. Savage, Sarah E. Jones, Jenna E. Reno, and Shari Veil. A case study: Targeting the stop. think. connect. cybersecurity campaign to university campuses. In *Oxford Research Encyclopedia of Communication*. 2017.

[199] Johanna Schönrock-Adema, Marjolein Heijne-Penninga, Elisabeth A. van Hell, and Janke Cohen-Schotanus. Necessary steps in factor analysis: enhancing validation studies of educational instruments. the pheem applied to clerks as an example. *Medical teacher*, 2009.

[200] Mahmood Sharif, Jumpei Urakawa, Nicolas Christin, Ayumu Kubota, and Akira Yamada. Predicting impending exposure to malicious content from user behavior. In *ACM Conference on Computer and Communications Security (CCS)*, 2018.

[201] Mark Shevlin and Jeremy N.V. Miles. Effects of sample size, model specification and factor loadings on the gfi in confirmatory factor analysis. *Personality and Individual differences*, 1998.

[202] Ruth Shillair, Shelia R. Cotten, Hsin-Yi Sandy Tsai, Saleem Alhabash, Robert LaRose, and Nora J. Rifon. Online safety begins with you and me: Convincing internet users to protect themselves. *Computers in Human Behavior*, 2015.

[203] Atsushi Shimada, Fumiya Okubo, and Hiroaki Ogata. Browsing-pattern mining from e-book logs with non-negative matrix factorization. In *EDM*, 2016.

[204] Aaron Smith and Monica Anderson. Social media use 2018: Demographics and statistics — Pew Research Center. URL https://www.pewresearch.org/internet/2018/03/01/social-media-use-in-2018/.

[205] Leslie B. Snyder and Jessica M. LaCroix. How effective are mediated health campaigns. *Public Communication Campaigns*, 2001.

[206] Ahmed Soliman, Jan Hafer, and Florian Lemmerich. A characterization of political com-

munities on reddit. In *Proceedings of the 30th ACM conference on hypertext and Social Media*, 2019.

[207] Janine L. Spears and Henri Barki. User participation in information systems security risk management. *MIS Quarterly*. URL https://pdfs.semanticscholar.org/387f/288a218a3e9c075c193910cd09f9b6874d88.pdf.

[208] Brian Stanton, Mary F. Theofanos, Sandra Spickard Prettyman, and Susanne Furman. Security fatigue. *IT Professional*, 2016.

[209] Keith Stevens, Philip Kegelmeyer, David Andrzejewski, and David Buttler. Exploring topic coherence over many models and many topics. In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 2012.

[210] Curtis Steward Jr, Luay A. Wahsheh, Aftab Ahmad, Jonathan M. Graham, Cheryl V. Hinds, Aurelia T. Williams, and Sandra J. DeLoatch. Software security: The dangerous afterthought. In *2012 Ninth International Conference on Information Technology-New Generations*, 2012.

[211] Elizabeth Stobert and Robert Biddle. A password manager that doesn't remember passwords. In *New Security Paradigms Workshop*, 2014.

[212] Elizabeth Stobert and Robert Biddle. The password life cycle: User behaviour in managing passwords. In *Symposium on Usable Privacy and Security (SOUPS)*. USENIX Association, July 2014.

[213] Karthik Subbian, Dhruv Sharma, Zhen Wen, and Jaideep Srivastava. Social capital: The power of influencers in networks. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, 2013.

[214] Melanie B. Tannenbaum, Justin Hepler, Rick S. Zimmerman, Lindsey Saul, Samantha Jacobs, Kristina Wilson, and Dolores Albarracín. Appealing to fear: A meta-analysis of fear appeal effectiveness and theories. *Psychological Bulletin*, 2015.

[215] Mike Thelwall, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas. Sentiment strength detection in short informal text. *Journal of the American society for information science and technology*, 2010.

[216] Kurt Thomas, Jennifer Pullman, Kevin Yeo, Ananth Raghunathan, Patrick Gage Kelley, Luca Invernizzi, Borbala Benko, Tadek Pietraszek, Sarvar Patel, Dan Boneh, and Elie Bursztein. Protecting accounts from credential stuffing with password breach alerting. In *USENIX Security Symposium*, 2019.

[217] Lynn Tsai, Primal Wijesekera, Joel Reardon, Irwin Reyes, Serge Egelman, David Wagner, Nathan Good, and Jung-Wei Chen. Turtle Guard: Helping android users apply contextual privacy preferences. In *Symposium on Usable Privacy and Security (SOUPS)*, 2017.

[218] Blase Ur, Jonathan Bees, Sean M. Segreti, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. Do users' perceptions of password security match reality? In *Conference on Human Factors in Computing Systems (CHI)*, 2016.

[219] Blase Ur, Felicia Alfieri, Maung Aung, Lujo Bauer, Nicolas Christin, Jessica Colnago,

Lorrie Faith Cranor, Henry Dixon, Pardis Emami Naeini, Hana Habib, Noah Johnson, and William Melicher. Design and evaluation of a data-driven password meter. In *Conference on Human Factors in Computing Systems (CHI)*, 2017.

[220] Wade M. Vagias. Likert-type scale response anchors. clemson international institute for tourism. *& Research Development, Department of Parks, Recreation and Tourism Management, Clemson University*, 2006.

[221] Kasisomayajula Viswanath, John R. Finnegan Jr., James Hertog, Phyllis Pirie, and David M. Murray. Community type and the diffusion of campaign information. *Gazette (Leiden, Netherlands)*, 1995.

[222] Rick Wash, Emilee Rader, Kami Vaniea, and Michelle Rizor. Out of the loop: How automated software updates cause unintended security consequences. In *Symposium on Usable Privacy and Security (SOUPS)*, 2014.

[223] Rick Wash, Emilee Rader, Ruthie Berman, and Zac Wellmer. Understanding password choices: How frequently entered passwords are re-used across websites. In *Symposium on Usable Privacy and Security (SOUPS)*, 2016.

[224] Rick Wash, Emilee Rader, and Chris Fennell. Can people self-report security accurately?: Agreement between self-report and behavioral measures. In *Conference on Human Factors in Computing Systems (CHI)*, 2017.

[225] Jane K. Winn. Are better security breach notification laws possible. *Berkeley tech. LJ*, 2009.

[226] Kim Witte and Mike Allen. A meta-analysis of fear appeals: Implications for effective public health campaigns. *Health education & behavior*, 2000.

[227] Yinglin Wu, Ling Xie, Shiang-Lin Huang, Ping Li, Zengwei Yuan, and Wenhua Liu. Using social media to strengthen public awareness of wildlife conservation. *Ocean & Coastal Management*, 153:76–83, 2018.

[228] Jason Yan. Security alert: User info breach. *Disqus Blog*, 2017. URL https://blog.disqus.com/security-alert-user-info-breach.

[229] Xinyue Ye, Bo Zhao, Thien Huu Nguyen, and Shaohua Wang. Social media and social awareness. *Manual of Digital Earth*, 2019.

[230] Ligita Zailskaitė-Jakštė, Armantas Ostreika, Adomas Jakštas, Evelina Stanevičienė, and Robertas Damaševičius. Brand communication in social media: The use of image colours in popular posts. In *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2017.

[231] Eric Zeng, Frank Li, Emily Stark, Adrienne Porter Felt, and Parisa Tabriz. Fixing https misconfigurations at scale: An experiment with security notifications. 2019.

[232] Kim Zetter. The year's 11 biggest hacks, from Ashley Madison to OPM. *Wired*, 2015. URL https://www.wired.com/2015/12/the-years-11-biggest-hacks-from-ashley-madison-to-opm/.

[233] Zhiguo Zhu. Discovering the influential users oriented to viral marketing based on online social networks. *Physica A: Statistical Mechanics and its Applications*, 2013.

[234] Yixin Zou and Florian Schaub. Beyond mandatory: Making data breach notifications useful for consumers. *IEEE Security & Privacy*, 2019.

[235] Yixin Zou, Abraham H. Mhaidli, Austin McCall, and Florian Schaub. "I've got nothing to lose": Consumers' risk perceptions and protective actions after the Equifax data breach. In *Symposium on Usable Privacy and Security (SOUPS)*, 2018.

[236] Yixin Zou, Shawn Danino, Kaiwen Sun, and Florian Schaub. You 'might' be affected: An empirical analysis of readability and usability issues in data breach notifications. In *Conference on Human Factors in Computing Systems (CHI)*, 2019.

Appendix

A Logistic regression assumptions for models in Chapter 3

Figure 1 shows the approximately linear relationships between each continuous feature and the log-odds of the outcome for the model studying the effects of the SeBIS feature set [128]. Similarly for the model studying browsing behavior, Figure 2 shows the approximately linear relationships between the continuous browsing features and the log-odds of the outcome. The linearity assumption for the demographics model is met by default due to all the features being categorical. The other two logistic regression assumptions, i.e., lack of influential observations and lack of multicollinearity, were satisfied for all models.

Table 3 depicts linearity assumptions for the logistic regression model studying the outcome of whether participants read about an incident and all four significant features from the first three models (Tables 3.4, 3.5, and 3.6 in Chapter 3). There were no influential observations or multicollinearity.

B Linear regression assumptions for models in Chapter 3

Figure 4 contains the plots showing that the linear regression model studying the number of actions in relation to browsing behavior approximately satisfies the assumptions of linearity, normality of residuals, and homogeneity of variance [127]. Figure 5 contains similar plots for the linear regression model for the number of actions in relation to participant demographics. For both models, the outcome variable was log-transformed and features were transformed as necessary to meet the model assumptions.

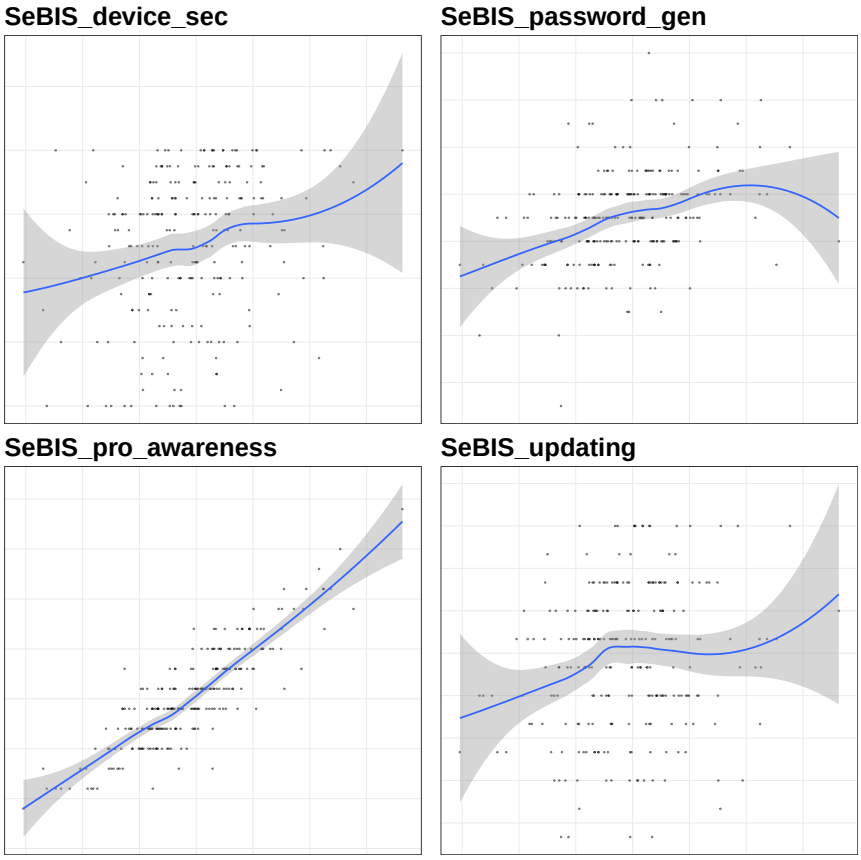


Figure 1: Scatterplots depicting approximate linearity between the four continuous SeBIS features and the log-odds of the outcome for the SeBIS model.

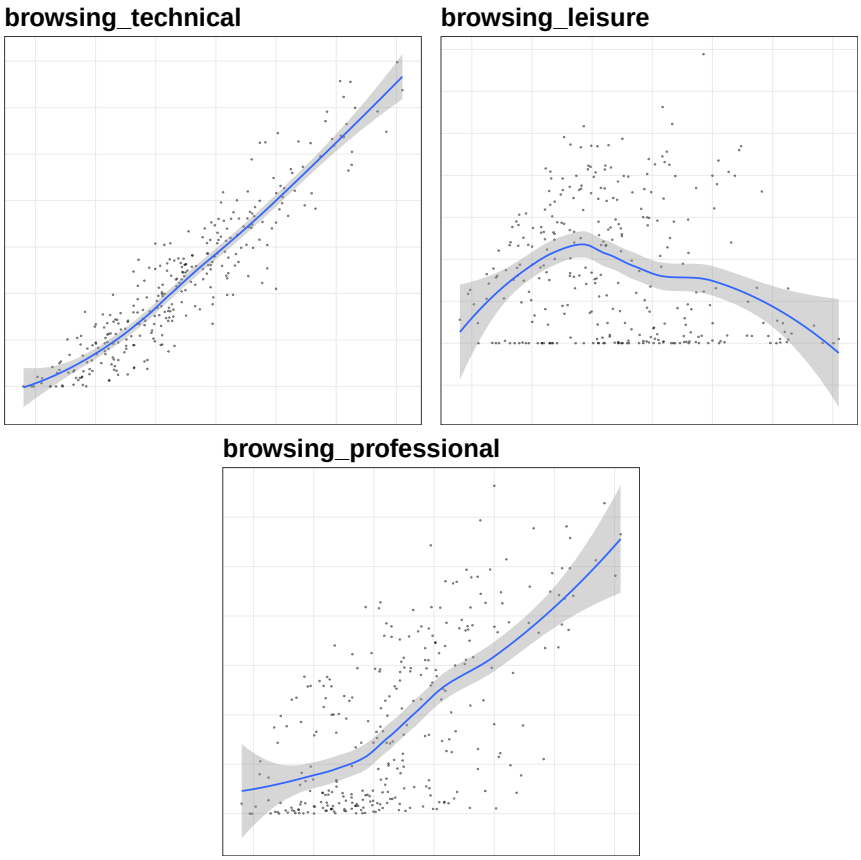


Figure 2: Scatterplots depicting approximate linearity between the three continuous browsing features and the log-odds of the outcome for the browsing behavior model.

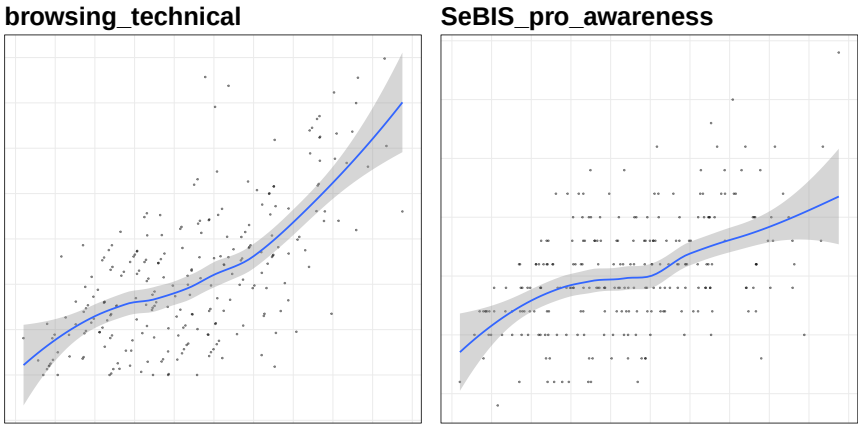
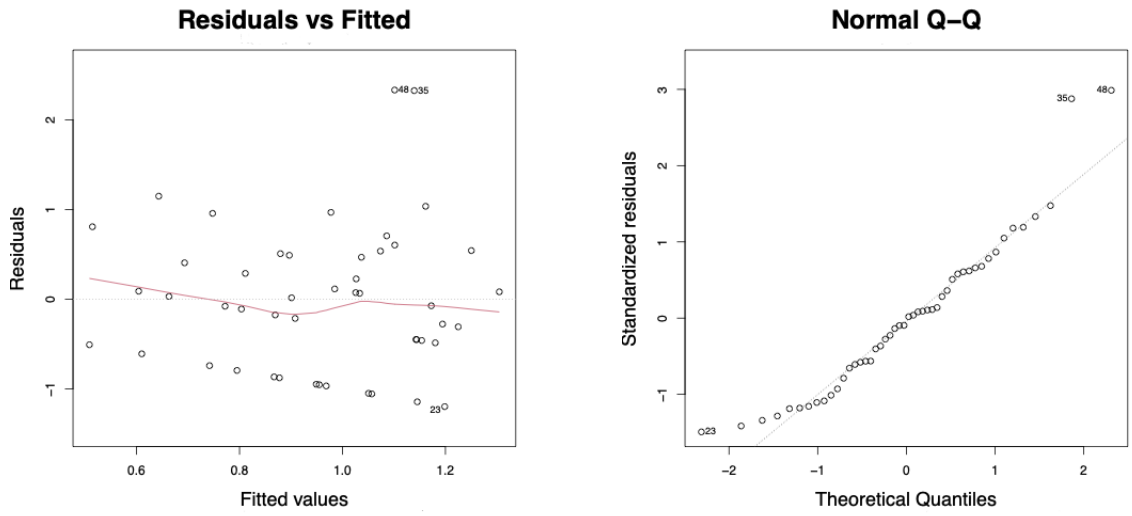
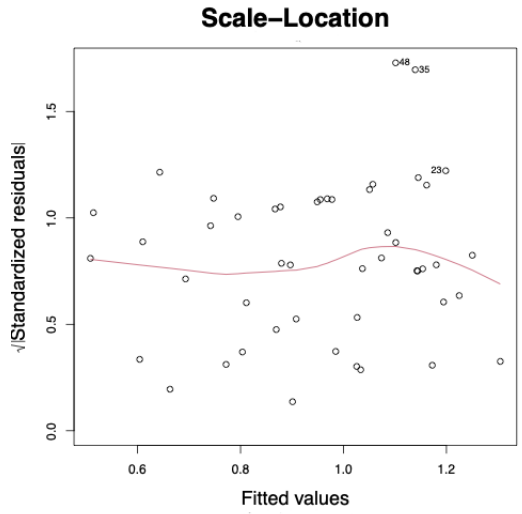


Figure 3: Scatterplots depicting approximate linearity between the continuous features and the log-odds of the outcome for the model studying the four significant features from the three feature sets.



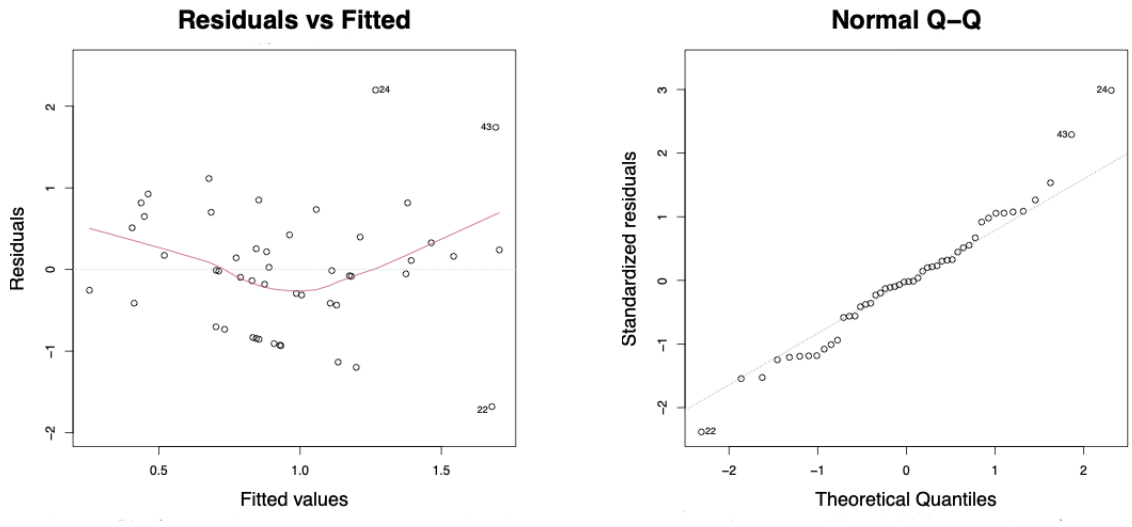
(a)

(b)



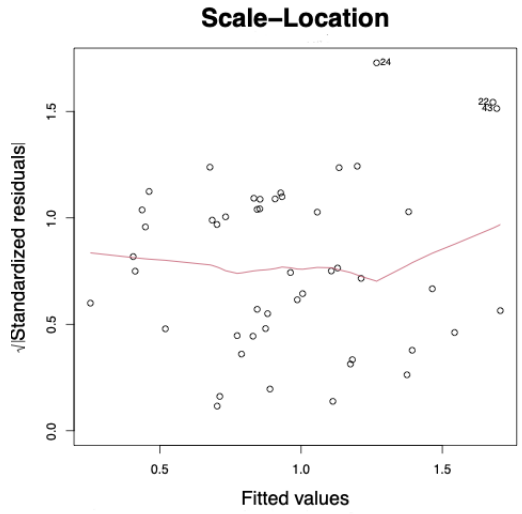
(c)

Figure 4: Scatterplots depicting approximate linearity, normality of residuals, and homogeneity of variance for the model studying actions in relation to browsing behavior. The linearity plot (a) shows an approximately horizontal distribution of the points scattered around the red line with no particular pattern. The normality of residuals plot (b) shows that most of the points approximately fall along the diagonal line. The homogeneity of variance plot (c) shows an approximately horizontal line with the points evenly scattered around it.



(a)

(b)



(c)

Figure 5: Scatterplots depicting approximate linearity, normality of residuals, and homogeneity of variance for the model studying actions in relation to demographics. The linearity plot (a) shows an approximately horizontal distribution of the points scattered around the red line with no particular pattern. The normality of residuals plot (b) shows that most of the points approximately fall along the diagonal line. The homogeneity of variance plot (c) shows an approximately horizontal line with the points evenly scattered around it.

C Confirmatory study survey from Chapter 3

The following survey contains questions about your computer usage and other behaviors. In some questions, we are specifically asking about the computer on which you have installed the SBO software, which we refer to as "**SBO computer**" throughout.

1. For each of the following events, please indicate whether you are familiar with the event. [*1=Not at all familiar*, *2=Slightly familiar*, *3=Somewhat familiar*, *4=Moderately familiar*, *5=Extremely familiar*]
   (a) Hurricane Katrina
   (b) Yahoo! passwords breach
   (c) Airbnb social security number breach
   (d) Russia meddling in the 2016 presidential elections
   (e) Equifax data breach
   (f) The 2018 Royal wedding
   (g) WannaCry ransomware attack
   (h) Panama papers leak
   (i) 2018 Soccer World Cup
2. (If answer to 1.e >= Somewhat familiar) Was your personal information leaked during the **Equifax data breach** (i.e., was your data stolen)? [*Yes*, *No*, *Not sure*]
3. (If answer to 1.e >= Slightly familiar) Did you take any of the following actions following the **Equifax data breach**? (check as many as apply)
   (a) Can't remember
   (b) Didn't do much/didn't do anything
   (c) Read more about it online
   (d) Read more about it somewhere else
   (e) Visited the Equifax website
   (f) Called Equifax
   (g) Informed myself about the breach in another way
   (h) Froze my credit report
   (i) Other: _____
4. When you **read about the Equifax data breach online** or **visited the Equifax site**, did you do so on your SBO computer or on another device (i.e., any other laptop/desktop/mobile/tablet)? [*Yes*, *No*, *Not sure*]
   (a) On your SBO computer
   (b) On another device
5. (If answer to 1.b >= Somewhat familiar) Was your password stolen in the **Yahoo! data breach**? [*Yes*, *No*, *Not sure*]

6. (If answer to 1.b >= Slightly familiar) Did you take any of the following actions following the **Yahoo! data breach**? (check as many as apply)
   (a) Can't remember
   (b) Didn't do much/didn't do anything
   (c) Read more about it online
   (d) Read more about it somewhere else
   (e) Informed myself about the breach in another way
   (f) Changed my Yahoo! password
   (g) Other: _____
7. (If answer to 1.g >= Slightly familiar) Did you take any of the following actions following the **WannaCry attack**?
   (a) Can't remember
   (b) Didn't do much/didn't do anything
   (c) Read more about it online
   (d) Read more about it somewhere else
   (e) Informed myself about the breach in another way
   (f) Paid ransom
   (g) Downloaded software patch
   (h) Other: _____
8. Have you ever been affected by some data breach or computer attack ***other than*** the Equifax breach, the Yahoo! passwords breach, or WannaCry? [*Yes*, *No*, *Not sure*]
9. In general when you **read web pages** (e.g., news articles, links you clicked on) **about data breaches** (e.g., Equifax, Yahoo! passwords breach, Ashley Madison breach, Target credit card data breach), how often do you read them on your SBO computer or on another device (i.e., any other laptop/desktop/mobile/tablet)? [*Never*, *Rarely*, *Sometimes*, *Often*, *Always*]
   (a) On your SBO computer
   (b) On another device
10. More generally, over ***all*** of your web browsing, what percentage of it do you on your SBO computer vs. on any other device (i.e., any other laptop/desktop/mobile/tablet)? [*0%/25%/50%/75%/100% on your SBO computer*]

D Quantile regression models from Chapter 5

Tables 1, 2, 3, 4, 5, 6, 7, and 8 describe the results of the quantile regression models studying the relationships between social media interactions and security behavior outcomes.

feature_name	*baseline*	*coef.*	*std. err.*	*t*	*p*
(Intercept)		-0.979	0.524	-1.867	0.070
gender: male	female	-0.268	0.282	-0.950	0.349
knows_prog_lang: true	false	-0.175	0.281	-0.625	0.536
is_student: true	**false**	**0.672**	**0.321**	**2.091**	**0.044**
age		0.008	0.010	0.808	0.425
social_factor_1		-0.121	0.066	-1.841	0.074
social_factor_2		-0.098	0.069	-1.424	0.163
social_factor_3		10×10^{-5}	0.157	0.001	0.999
social_factor_4		-0.119	0.119	-0.997	0.326
social_factor_5		-0.060	0.085	-0.706	0.485
social_factor_6		0.030	0.085	0.353	0.726

Table 1: Quantile regression model for the 25th percentile studying the first security behavior factor.

feature_name	*baseline*	*coef.*	*std. err.*	*t*	*p*
(Intercept)		0.072	0.502	0.143	0.887
gender: male	female	0.049	0.236	0.210	0.835
knows_prog_lang: true	false	-0.098	0.327	-0.298	0.767
is_student: true	false	-0.064	0.374	-0.171	0.865
age		-0.011	0.010	-1.189	0.243
social_factor_1		0.109	0.065	1.677	0.102
social_factor_2		**0.258**	**0.064**	**4.059**	**<0.01**
social_factor_3		0.058	0.135	0.434	0.667
social_factor_4		0.007	0.108	0.064	0.950
social_factor_5		0.150	0.141	1.068	0.293
social_factor_6		-0.127	0.102	-1.239	0.223

Table 2: Quantile regression model for the 25th percentile studying the second security behavior factor.

feature_name	*baseline*	*coef.*	*std. err.*	*t*	*p*
(Intercept)		-0.710	0.588	-1.207	0.235
gender: male	female	-0.191	0.315	-0.608	0.547
knows_prog_lang: true	false	-0.703	0.368	-1.911	0.064
is_student: true	**false**	**0.904**	**0.411**	**2.198**	**0.035**
age		0.012	0.011	1.039	0.306
social_factor_1		-0.100	0.106	-0.949	0.349
social_factor_2		-0.117	0.124	-0.942	0.353
social_factor_3		0.041	0.126	0.323	0.748
social_factor_4		-0.115	0.128	-0.901	0.374
social_factor_5		-0.250	0.126	-1.982	0.055
social_factor_6		0.070	0.119	0.592	0.557

Table 3: Quantile regression model for the 50th percentile studying the first security behavior factor.

feature_name	*baseline*	*coef.*	*std. err.*	*t*	*p*
(Intercept)		0.067	0.336	0.198	0.844
gender: male	female	0.185	0.180	1.026	0.312
knows_prog_lang: true	false	0.012	0.211	0.054	0.957
is_student: true	false	-0.365	0.235	-1.551	0.130
age		-0.001	0.006	-0.182	0.857
social_factor_1		**0.152**	**0.060**	**2.509**	**0.017**
social_factor_2		**0.251**	**0.071**	**3.523**	**0.001**
social_factor_3		0.038	0.072	0.524	0.604
social_factor_4		0.014	0.073	0.184	0.855
social_factor_5		0.067	0.072	0.925	0.361
social_factor_6		**-0.147**	**0.068**	**-2.160**	**0.038**

Table 4: Quantile regression model for the 50th percentile studying the second security behavior factor.

feature_name	*baseline*	*coef.*	*std. err.*	*t*	*p*
(Intercept)		0.374	0.539	0.695	0.491
gender: male	female	-0.162	0.337	-0.479	0.635
knows_prog_lang: true	false	-0.809	0.474	-1.708	0.097
is_student: true	false	0.368	0.492	0.748	0.460
age		0.008	0.010	0.730	0.470
social_factor_1		-0.029	0.122	-0.235	0.815
social_factor_2		-0.281	0.213	-1.322	0.195
social_factor_3		**0.518**	**0.114**	**4.551**	**<0.01**
social_factor_4		-0.103	0.125	-0.825	0.415
social_factor_5		-0.245	0.202	-1.217	0.232
social_factor_6		0.169	0.151	1.120	0.270

Table 5: Quantile regression model for the 75th percentile studying the first security behavior factor.

feature_name	*baseline*	*coef.*	*std. err.*	*t*	*p*
(Intercept)		0.005	0.431	0.012	0.991
gender: male	female	0.151	0.219	0.690	0.495
knows_prog_lang: true	false	0.235	0.280	0.837	0.408
is_student: true	false	-0.351	0.301	-1.165	0.252
age		0.005	0.008	0.639	0.527
social_factor_1		0.151	0.075	2.024	0.051
social_factor_2		0.191	0.131	1.451	0.156
social_factor_3		0.069	0.073	0.940	0.353
social_factor_4		-0.007	0.075	-0.091	0.928
social_factor_5		0.025	0.069	0.356	0.724
social_factor_6		-0.067	0.088	-0.763	0.451

Table 6: Quantile regression model for the 75th percentile studying the second security behavior factor.

feature_name	*baseline*	*coef.*	*std. err.*	*t*	*p*
(Intercept)		-0.262	0.885	-0.296	0.769
gender: male	female	0.071	0.474	0.150	0.881
knows_prog_lang: true	false	-1.227	0.865	-1.419	0.165
is_student: true	false	1.041	0.930	1.120	0.270
age		**0.036**	**0.018**	**2.053**	**0.048**
social_factor_1		0.176	0.188	0.938	0.354
social_factor_2		-0.534	0.461	-1.159	0.254
social_factor_3		**1.111**	**0.190**	**5.845**	**<0.01**
social_factor_4		0.260	0.187	1.390	0.173
social_factor_5		-0.422	0.440	-0.961	0.343
social_factor_6		0.207	0.254	0.815	0.420

Table 7: Quantile regression model for the 90th percentile studying the first security behavior factor.

feature_name	*baseline*	*coef.*	*std. err.*	*t*	*p*
(Intercept)		-1.509	1.721	-0.876	0.387
gender: male	female	0.301	0.643	0.469	0.642
knows_prog_lang: true	false	1.130	0.821	1.376	0.178
is_student: true	false	-0.234	0.922	-0.254	0.801
age		0.053	0.030	1.744	0.090
social_factor_1		-0.010	0.205	-0.050	0.960
social_factor_2		0.177	0.515	0.342	0.734
social_factor_3		-0.144	0.507	-0.284	0.778
social_factor_4		-0.262	0.199	-1.310	0.199
social_factor_5		0.265	0.209	1.269	0.213
social_factor_6		-0.231	0.303	-0.763	0.451

Table 8: Quantile regression model for the 90th percentile studying the second security behavior factor.

E Quantile regression models from Chapter 6

Tables 9, 10, 11, 12, 13, 14, and 15 describe the results of the quantile regression models studying the popularity of security and privacy posts. Tables 16, 17, 18, 19, 20, 21, 22, and 23 describe the results of the quantile regression models studying the popularity of baseline posts.

feature_name	*baseline*	*coef.*	*std. err.*	*t*	*p*
Intercept		**1.524**	**0.255**	**5.976**	**<0.01**
has_link: true	**false**	**-1.090**	**0.077**	**-14.127**	**<0.01**
has_image: true	false	0.040	0.086	0.468	0.640
only_link_no_text: true	**false**	**-1.888**	**0.088**	**-21.465**	**<0.01**
has_author_flair: true	**false**	**0.195**	**0.077**	**2.530**	**0.011**
has_link_flair: true	**false**	**2.687**	**0.051**	**52.206**	**<0.01**
is_crosspost: true	false	-0.105	0.144	-0.732	0.464
num_emojis		**0.062**	**0.015**	**4.018**	**<0.01**
poster_is_gold: true	false	0.018	0.111	0.161	0.872
poster_num_submissions		**-0.033**	**0.003**	**-10.255**	**<0.01**
poster_total_karma		**-1.312×10^{-6}**	**2.76×10^{-8}**	**-47.556**	**<0.01**
poster_comment_karma		-2.912×10^{-8}	1.29×10^{-7}	-0.226	0.821
poster_avg_score		**0.011**	**2.6×10^{-5}**	**408.953**	**<0.01**
poster_avg_upvote_ratios		**1.025**	**0.268**	**3.828**	**<0.01**
poster_awardee_karma		**2×10^{-4}**	**4.56×10^{-7}**	**341.643**	**<0.01**
poster_awarder_karma		**4.542×10^{-5}**	**1.15×10^{-5}**	**3.936**	**<0.01**
readability_flesch		**0.002**	**0.000**	**12.832**	**<0.01**
text_num_characters		**8×10^{-4}**	**7.28×10^{-6}**	**110.234**	**<0.01**
sentiment_pos		**0.683**	**0.045**	**15.186**	**<0.01**
sentiment_neg		**-0.176**	**0.032**	**-5.488**	**<0.01**
tone_anger		0.082	0.199	0.410	0.682
tone_fear		0.036	0.167	0.212	0.832
tone_joy		**1.035**	**0.107**	**9.714**	**<0.01**
tone_sadness		**0.442**	**0.106**	**4.163**	**<0.01**
tone_analytical		-0.026	0.063	-0.413	0.680
tone_confident		-0.003	0.117	-0.023	0.981
tone_tentative		**2.656**	**0.070**	**37.779**	**<0.01**

Table 9: Quantile regression model studying the popularity of security and privacy posts for the 0.50 quantile for the num_all_comments outcome.

feature_name	*baseline*	*coef.*	*std. err.*	*t*	*p*
Intercept		1.054	1.507	0.700	0.484
has_link: true	false	0.440	0.461	0.955	0.340
has_image: true	false	0.129	0.530	0.244	0.808
only_link_no_text: true	false	0.261	0.499	0.523	0.601
has_author_flair: true	false	0.147	0.493	0.298	0.766
has_link_flair: true	**false**	**0.684**	**0.314**	**2.180**	**0.029**
is_crosspost: true	false	0.015	0.711	0.021	0.983
num_emojis		-0.012	0.107	-0.114	0.909
poster_is_gold: true	false	0.004	0.683	0.006	0.996
poster_num_submissions		-0.026	0.020	-1.277	0.202
poster_total_karma		**-3.116×10^{-6}**	**1.56×10^{-7}**	**-19.981**	**<0.01**
poster_comment_karma		**-1.666×10^{-6}**	**6.56×10^{-7}**	**-2.541**	**0.011**
poster_avg_score		**0.066**	**0.000**	**465.833**	**<0.01**
poster_avg_upvote_ratios		0.844	1.555	0.543	0.587
poster_awardee_karma		**3×10^{-4}**	**2.42×10^{-6}**	**142.273**	**<0.01**
poster_awarder_karma		**9×10^{-4}**	**5.07×10^{-5}**	**18.115**	**<0.01**
readability_flesch		0.001	0.001	1.308	0.191
text_num_characters		**0.003**	**3.74×10^{-5}**	**90.059**	**<0.01**
sentiment_pos		**1.805**	**0.260**	**6.942**	**<0.01**
sentiment_neg		**-2.084**	**0.183**	**-11.391**	**<0.01**
tone_anger		0.025	1.265	0.020	0.984
tone_fear		0.148	0.942	0.157	0.875
tone_joy		0.275	0.647	0.424	0.671
tone_sadness		0.254	0.660	0.385	0.700
tone_analytical		0.420	0.371	1.134	0.257
tone_confident		0.084	0.662	0.126	0.900
tone_tentative		0.271	0.425	0.636	0.525

Table 10: Quantile regression model studying the popularity of security and privacy posts for the 0.75 quantile for the num_all_comments outcome.

feature_name	*baseline*	*coef.*	*std. err.*	*t*	*p*
Intercept		0.190	3.268	0.058	0.954
has_link: true	false	-0.181	0.832	-0.218	0.828
has_image: true	false	-0.073	1.023	-0.071	0.943
only_link_no_text: true	false	-0.101	0.951	-0.106	0.915
has_author_flair: true	false	-0.023	0.944	-0.024	0.981
has_link_flair: true	**false**	**1.184**	**0.598**	**1.982**	**0.047**
is_crosspost: true	false	-0.014	1.171	-0.012	0.991
num_emojis		**1.576**	**0.217**	**7.261**	**<0.01**
poster_is_gold: true	false	-2.132×10^{-6}	1.251	-1.7×10^{-6}	1.000
poster_num_submissions		**0.686**	**0.045**	**15.182**	**<0.01**
poster_total_karma		$\mathbf{3.839\times10^{-6}}$	$\mathbf{2.44\times10^{-7}}$	**15.721**	**<0.01**
poster_comment_karma		$\mathbf{-1.868\times10^{-5}}$	$\mathbf{1.08\times10^{-6}}$	**-17.293**	**<0.01**
poster_avg_score		**0.267**	**0.000**	**1373.469**	**<0.01**
poster_avg_upvote_ratios		0.059	3.323	0.018	0.986
poster_awardee_karma		$\mathbf{4\times10^{-4}}$	$\mathbf{3.8\times10^{-6}}$	**105.994**	**<0.01**
poster_awarder_karma		**0.004**	$\mathbf{7.46\times10^{-5}}$	**51.170**	**<0.01**
readability_flesch		0.002	0.002	0.960	0.337
text_num_characters		**0.010**	$\mathbf{6.13\times10^{-5}}$	**167.938**	**<0.01**
sentiment_pos		-0.144	0.493	-0.292	0.771
sentiment_neg		-0.433	0.363	-1.194	0.233
tone_anger		0.005	2.516	0.002	0.998
tone_fear		-0.006	1.715	-0.004	0.997
tone_joy		-0.074	1.203	-0.062	0.951
tone_sadness		0.177	1.257	0.140	0.888
tone_analytical		0.033	0.693	0.048	0.962
tone_confident		0.036	1.267	0.028	0.978
tone_tentative		0.069	0.809	0.085	0.932

Table 11: Quantile regression model studying the popularity of security and privacy posts for the 0.90 quantile for the num_all_comments outcome.

feature_name	*baseline*	*coef.*	*std. err.*	*t*	*p*
Intercept		**4.947**	**0.424**	**11.681**	**<0.01**
has_link: true	**false**	**-0.450**	**0.117**	**-3.830**	**<0.01**
has_image: true	**false**	**2.516**	**0.132**	**19.110**	**<0.01**
only_link_no_text: true	false	0.128	0.133	0.967	0.333
has_author_flair: true	**false**	**4.548**	**0.119**	**38.159**	**<0.01**
has_link_flair: true	**false**	**1.024**	**0.079**	**12.907**	**<0.01**
is_crosspost: true	false	-0.115	0.217	-0.531	0.596
num_emojis		**0.330**	**0.022**	**15.318**	**<0.01**
poster_is_gold: true	false	0.175	0.163	1.074	0.283
poster_num_submissions		**-0.015**	**0.005**	**-3.000**	**<0.01**
poster_total_karma		**6.068×10^{-7}**	**4.1×10^{-8}**	**14.787**	**<0.01**
poster_comment_karma		**-1.088×10^{-6}**	**2.1×10^{-7}**	**-5.176**	**<0.01**
poster_avg_score		**0.014**	**3.84×10^{-5}**	**370.109**	**<0.01**
poster_avg_upvote_ratios		**-3.647**	**0.449**	**-8.130**	**<0.01**
poster_awardee_karma		**7.233×10^{-6}**	**7.44×10^{-7}**	**9.726**	**<0.01**
poster_awarder_karma		**2×10^{-4}**	**1.77×10^{-5}**	**12.165**	**<0.01**
readability_flesch		**8×10^{-4}**	**0.000**	**4.205**	**<0.01**
text_num_characters		**5×10^{-4}**	**1.11×10^{-5}**	**47.556**	**<0.01**
sentiment_pos		0.011	0.068	0.157	0.876
sentiment_neg		-0.009	0.047	-0.192	0.848
tone_anger		0.225	0.299	0.754	0.451
tone_fear		0.214	0.260	0.824	0.410
tone_joy		0.268	0.164	1.635	0.102
tone_sadness		**0.441**	**0.166**	**2.657**	**<0.01**
tone_analytical		0.008	0.095	0.079	0.937
tone_confident		0.007	0.174	0.040	0.968
tone_tentative		0.187	0.110	1.693	0.091

Table 12: Quantile regression model studying the popularity of security and privacy posts for the 0.25 quantile for the total_votes_estimate outcome.

feature_name	*baseline*	*coef.*	*std. err.*	*t*	*p*
Intercept		**30.345**	**0.604**	**50.211**	**<0.01**
has_link: true	**false**	**0.881**	**0.169**	**5.198**	**<0.01**
has_image: true	**false**	**14.127**	**0.186**	**76.032**	**<0.01**
only_link_no_text: true	**false**	**2.705**	**0.192**	**14.101**	**<0.01**
has_author_flair: true	**false**	**11.147**	**0.166**	**67.134**	**<0.01**
has_link_flair: true	**false**	**2.882**	**0.112**	**25.689**	**<0.01**
is_crosspost: true	**false**	**-0.808**	**0.309**	**-2.611**	**<0.01**
num_emojis		0.039	0.033	1.194	0.232
poster_is_gold: true	**false**	**6.003**	**0.238**	**25.259**	**<0.01**
poster_num_submissions		**-0.120**	**0.007**	**-16.885**	**<0.01**
poster_total_karma		**2.523×10^{-6}**	**5.88×10^{-8}**	**42.886**	**<0.01**
poster_comment_karma		**-6.41×10^{-6}**	**2.75×10^{-7}**	**-23.274**	**<0.01**
poster_avg_score		**0.1335**	**5.55×10^{-5}**	**2406.328**	**<0.01**
poster_avg_upvote_ratios		**-30.551**	**0.630**	**-48.491**	**<0.01**
poster_awardee_karma		**-4.059×10^{-5}**	**9.66×10^{-7}**	**-42.002**	**<0.01**
poster_awarder_karma		**4×10^{-4}**	**2.46×10^{-5}**	**17.657**	**<0.01**
readability_flesch		3×10^{-4}	0.000	0.959	0.338
text_num_characters		**0.003**	**1.57×10^{-5}**	**167.404**	**<0.01**
sentiment_pos		0.099	0.098	1.016	0.309
sentiment_neg		**-0.157**	**0.070**	**-2.249**	**0.025**
tone_anger		**2.110**	**0.433**	**4.876**	**<0.01**
tone_fear		0.099	0.363	0.273	0.785
tone_joy		**1.208**	**0.232**	**5.199**	**<0.01**
tone_sadness		0.047	0.233	0.202	0.840
tone_analytical		-0.019	0.138	-0.135	0.892
tone_confident		0.131	0.255	0.512	0.608
tone_tentative		0.072	0.154	0.470	0.639

Table 13: Quantile regression model studying the popularity of security and privacy posts for the 0.50 quantile for the total_votes_estimate outcome.

feature_name	*baseline*	*coef.*	*std. err.*	*t*	*p*
Intercept		-0.920	0.966	-0.953	0.341
has_link: true	**false**	**0.696**	**0.251**	**2.771**	**<0.01**
has_image: true	**false**	**5.266**	**0.291**	**18.095**	**<0.01**
only_link_no_text: true	**false**	**1.721**	**0.292**	**5.899**	**<0.01**
has_author_flair: true	**false**	**12.142**	**0.242**	**50.112**	**<0.01**
has_link_flair: true	**false**	**4.530**	**0.166**	**27.303**	**<0.01**
is_crosspost: true	false	-0.045	0.456	-0.099	0.921
num_emojis		**0.443**	**0.048**	**9.215**	**<0.01**
poster_is_gold: true	**false**	**3.524**	**0.354**	**9.942**	**<0.01**
poster_num_submissions		**-0.110**	**0.011**	**-10.384**	**<0.01**
poster_total_karma		**1.271×10^{-5}**	**8.67×10^{-8}**	**146.603**	**<0.01**
poster_comment_karma		**-1.917×10^{-5}**	**3.69×10^{-7}**	**-51.965**	**<0.01**
poster_avg_score		**1.249**	**9.27×10^{-5}**	**1.35×10^{4}**	**<0.01**
poster_avg_upvote_ratios		-1.430	0.998	-1.433	0.152
poster_awardee_karma		**-4×10^{-4}**	**1.31×10^{-6}**	**-290.209**	**<0.01**
poster_awarder_karma		**0.002**	**3.4×10^{-5}**	**67.717**	**<0.01**
readability_flesch		7×10^{-4}	0.001	1.242	0.214
text_num_characters		**0.005**	**2.28×10^{-5}**	**228.722**	**<0.01**
sentiment_pos		**0.780**	**0.144**	**5.406**	**<0.01**
sentiment_neg		**-1.468**	**0.103**	**-14.312**	**<0.01**
tone_anger		0.208	0.663	0.313	0.754
tone_fear		-0.001	0.526	-0.003	0.998
tone_joy		**1.166**	**0.341**	**3.422**	**<0.01**
tone_sadness		0.282	0.343	0.823	0.411
tone_analytical		-0.026	0.207	-0.126	0.900
tone_confident		0.008	0.379	0.021	0.983
tone_tentative		**0.848**	**0.228**	**3.720**	**<0.01**

Table 14: Quantile regression model studying the popularity of security and privacy posts for the 0.75 quantile for the total_votes_estimate outcome.

feature_name	*baseline*	*coef.*	*std. err.*	*t*	*p*
Intercept		0.004	2.515	0.002	0.999
has_link: true	false	-0.002	0.597	-0.003	0.998
has_image: true	false	-0.003	0.815	-0.004	0.997
only_link_no_text: true	false	-0.002	0.751	-0.003	0.998
has_author_flair: true	false	0.003	0.651	0.005	0.996
has_link_flair: true	false	0.006	0.427	0.014	0.989
is_crosspost: true	false	-5×10^{-4}	1.089	-0.001	1.000
num_emojis		0.001	0.116	0.010	0.992
poster_is_gold: true	false	0.002	0.939	0.002	0.999
poster_num_submissions		**-0.045**	**0.022**	**-2.040**	**0.041**
poster_total_karma		$\mathbf{2\times10^{-4}}$	$\mathbf{1.94\times10^{-7}}$	**1190.238**	**<0.01**
poster_comment_karma		$\mathbf{-3\times10^{-4}}$	$\mathbf{8.74\times10^{-7}}$	**-302.694**	**<0.01**
poster_avg_score		**4.435**	**0.000**	$\mathbf{2.69\times10^{4}}$	**<0.01**
poster_avg_upvote_ratios		0.004	2.618	0.002	0.999
poster_awardee_karma		**-0.002**	$\mathbf{3.29\times10^{-6}}$	**-470.341**	**<0.01**
poster_awarder_karma		**0.002**	$\mathbf{6.72\times10^{-5}}$	**26.737**	**<0.01**
readability_flesch		0.002	0.002	1.011	0.312
text_num_characters		**0.010**	$\mathbf{5.66\times10^{-5}}$	**182.989**	**<0.01**
sentiment_pos		-7×10^{-4}	0.366	-0.002	0.998
sentiment_neg		-0.012	0.258	-0.048	0.962
tone_anger		6×10^{-4}	1.732	0.000	1.000
tone_fear		6×10^{-4}	1.222	0.000	1.000
tone_joy		0.003	0.825	0.003	0.998
tone_sadness		0.002	0.840	0.002	0.998
tone_analytical		0.003	0.493	0.005	0.996
tone_confident		8×10^{-4}	0.924	0.001	0.999
tone_tentative		0.002	0.544	0.004	0.996

Table 15: Quantile regression model studying the popularity of security and privacy posts for the 0.90 quantile for the total_votes_estimate outcome.

feature_name	*baseline*	*coef.*	*std. err.*	*t*	*p*
Intercept		0.353	0.310	1.138	0.255
only_link_no_text: true	**false**	**-1.156**	**0.098**	**-11.814**	**<0.01**
only_image_no_text: true	**false**	**0.948**	**0.087**	**10.874**	**<0.01**
has_author_flair: true	**false**	**1.072**	**0.089**	**12.024**	**<0.01**
has_link_flair: true	**false**	**0.351**	**0.060**	**5.867**	**<0.01**
is_crosspost: true	false	-0.085	0.208	-0.410	0.682
num_emojis		-0.001	0.001	-0.663	0.507
poster_is_gold: true	false	0.027	0.139	0.195	0.846
poster_num_submissions		0.003	0.003	0.942	0.346
poster_total_karma		**-5.721×10^{-7}**	**5.33×10^{-8}**	**-10.741**	**<0.01**
poster_comment_karma		**2.157×10^{-6}**	**3.3×10^{-7}**	**6.540**	**<0.01**
poster_avg_score		**0.002**	**2.83×10^{-5}**	**63.157**	**<0.01**
poster_avg_upvote_ratios		0.406	0.341	1.189	0.235
poster_awardee_karma		**3.274×10^{-6}**	**2.68×10^{-7}**	**12.210**	**<0.01**
poster_awarder_karma		5.118×10^{-6}	$2.56e\times10^{-5}$	0.200	0.842
readability_flesch		**0.002**	**0.000**	**5.546**	**<0.01**
text_num_characters		**3×10^{-4}**	**3.28×10^{-5}**	**7.899**	**<0.01**
sentiment_pos		**0.128**	**0.053**	**2.422**	**0.015**
sentiment_neg		**-0.496**	**0.048**	**-10.297**	**<0.01**
tone_anger		0.088	0.214	0.412	0.680
tone_fear		0.037	0.250	0.146	0.884
tone_joy		0.133	0.107	1.240	0.215
tone_sadness		0.159	0.124	1.278	0.201
tone_analytical		0.111	0.083	1.350	0.177
tone_confident		0.164	0.150	1.096	0.273
tone_tentative		**0.404**	**0.081**	**4.992**	**<0.01**

Table 16: Quantile regression model studying the popularity of baseline posts for the 0.25 quantile for the num_all_comments outcome.

feature_name	*baseline*	*coef.*	*std. err.*	*t*	*p*
Intercept		**2.513**	**0.652**	**3.855**	**<0.01**
only_link_no_text: true	false	-0.150	0.216	-0.695	0.487
only_image_no_text: true	**false**	**0.805**	**0.186**	**4.331**	**<0.01**
has_author_flair: true	**false**	**1.811**	**0.188**	**9.650**	**<0.01**
has_link_flair: true	**false**	**0.384**	**0.127**	**3.009**	**<0.01**
is_crosspost: true	**false**	**-1.392**	**0.421**	**-3.302**	**<0.01**
num_emojis		**-0.006**	**0.003**	**-2.055**	**0.040**
poster_is_gold: true	**false**	**2.346**	**0.315**	**7.447**	**<0.01**
poster_num_submissions		0.008	0.007	1.058	0.290
poster_total_karma		**-1.875×10^{-6}**	**1.17×10^{-7}**	**-16.075**	**<0.01**
poster_comment_karma		**4.848×10^{-6}**	**6.11×10^{-7}**	**7.934**	**<0.01**
poster_avg_score		**0.007**	**6.38×10^{-5}**	**102.840**	**<0.01**
poster_avg_upvote_ratios		-0.850	0.705	-1.205	0.228
poster_awardee_karma		**1.566×10^{-5}**	**5.82×10^{-7}**	**26.911**	**<0.01**
poster_awarder_karma		**4×10^{-4}**	**5.3×10^{-5}**	**6.745**	**<0.01**
readability_flesch		**0.002**	**0.001**	**2.424**	**0.015**
text_num_characters		**0.001**	**6.19×10^{-5}**	**21.272**	**<0.01**
sentiment_pos		**0.458**	**0.111**	**4.136**	**<0.01**
sentiment_neg		**-1.300**	**0.101**	**-12.909**	**<0.01**
tone_anger		**1.851**	**0.467**	**3.964**	**<0.01**
tone_fear		-0.146	0.542	-0.269	0.788
tone_joy		0.081	0.225	0.360	0.719
tone_sadness		-0.120	0.263	-0.458	0.647
tone_analytical		0.216	0.177	1.223	0.221
tone_confident		**1.043**	**0.326**	**3.204**	**<0.01**
tone_tentative		**0.742**	**0.173**	**4.285**	**<0.01**

Table 17: Quantile regression model studying the popularity of baseline posts for the 0.50 quantile for the num_all_comments outcome.

feature_name	*baseline*	*coef.*	*std. err.*	*t*	*p*
Intercept		1.160	1.640	0.707	0.480
only_link_no_text: true	false	0.295	0.593	0.497	0.619
only_image_no_text: true	false	0.348	0.477	0.731	0.465
has_author_flair: true	false	0.086	0.503	0.172	0.863
has_link_flair: true	**false**	**1.175**	**0.322**	**3.652**	**<0.01**
is_crosspost: true	false	4×10^{-4}	1.026	0.000	1.000
num_emojis		-0.017	0.012	-1.457	0.145
poster_is_gold: true	false	-0.005	0.876	-0.005	0.996
poster_num_submissions		**0.147**	**0.018**	**7.991**	**<0.01**
poster_total_karma		**-3.978×10^{-6}**	**2.96×10^{-7}**	**-13.423**	**<0.01**
poster_comment_karma		**1.475×10^{-5}**	**1.22×10^{-6}**	**12.122**	**<0.01**
poster_avg_score		**0.025**	**0.000**	**164.406**	**<0.01**
poster_avg_upvote_ratios		1.024	1.757	0.582	0.560
poster_awardee_karma		**7.444×10^{-5}**	**1.51×10^{-6}**	**49.418**	**<0.01**
poster_awarder_karma		**0.001**	**0.000**	**10.136**	**<0.01**
readability_flesch		**0.010**	**0.002**	**4.134**	**<0.01**
text_num_characters		**0.005**	**0.000**	**33.734**	**<0.01**
sentiment_pos		**1.796**	**0.273**	**6.582**	**<0.01**
sentiment_neg		**-2.200**	**0.262**	**-8.400**	**<0.01**
tone_anger		0.201	1.258	0.160	0.873
tone_fear		0.024	1.371	0.018	0.986
tone_joy		0.253	0.562	0.450	0.653
tone_sadness		0.140	0.656	0.213	0.831
tone_analytical		0.207	0.449	0.460	0.646
tone_confident		0.125	0.858	0.146	0.884
tone_tentative		0.350	0.439	0.796	0.426

Table 18: Quantile regression model studying the popularity of baseline posts for the 0.75 quantile for the num_all_comments outcome.

feature_name	*baseline*	*coef.*	*std. err.*	*t*	*p*
Intercept		0.042	3.788	0.011	0.991
only_link_no_text: true	false	0.021	1.435	0.015	0.988
only_image_no_text: true	false	0.007	0.950	0.007	0.994
has_author_flair: true	false	0.014	1.013	0.013	0.989
has_link_flair: true	false	0.022	0.655	0.033	0.973
is_crosspost: true	false	-5×10^{-4}	1.926	-0.000	1.000
num_emojis		-0.054	0.037	-1.469	0.142
poster_is_gold: true	false	0.010	1.779	0.006	0.996
poster_num_submissions		**1.061**	**0.042**	**25.106**	**<0.01**
poster_total_karma		-1.009×10^{-6}	5.54×10^{-7}	-1.819	0.069
poster_comment_karma		$\mathbf{2.756\times10^{-5}}$	$\mathbf{1.79\times10^{-6}}$	**15.366**	**<0.01**
poster_avg_score		**0.106**	**0.000**	**402.494**	**<0.01**
poster_avg_upvote_ratios		0.035	4.036	0.009	0.993
poster_awardee_karma		$\mathbf{1\times10^{-4}}$	$\mathbf{2.97\times10^{-6}}$	**45.684**	**<0.01**
poster_awarder_karma		**0.003**	**0.000**	**16.387**	**<0.01**
readability_flesch		**0.038**	**0.007**	**5.088**	**<0.01**
text_num_characters		**0.017**	**0.000**	**74.963**	**<0.01**
sentiment_pos		0.060	0.566	0.107	0.915
sentiment_neg		-0.052	0.589	-0.088	0.930
tone_anger		6×10^{-4}	2.625	0.000	1.000
tone_fear		3×10^{-4}	2.785	0.000	1.000
tone_joy		0.004	1.114	0.003	0.997
tone_sadness		0.006	1.330	0.004	0.997
tone_analytical		0.010	0.915	0.011	0.991
tone_confident		0.002	1.794	0.001	0.999
tone_tentative		0.016	0.907	0.017	0.986

Table 19: Quantile regression model studying the popularity of baseline posts for the 0.90 quantile for the num_all_comments outcome.

feature_name	*baseline*	*coef.*	*std. err.*	*t*	*p*
Intercept		**2.704**	**1.036**	**2.611**	**<0.01**
only_link_no_text: true	**false**	**8.182**	**0.309**	**26.437**	**<0.01**
only_image_no_text: true	**false**	**11.480**	**0.248**	**46.362**	**<0.01**
has_author_flair: true	**false**	**3.266**	**0.254**	**12.839**	**<0.01**
has_link_flair: true	false	0.177	0.174	1.017	0.309
is_crosspost: true	false	-1.016	0.563	-1.806	0.071
num_emojis		-4×10^{-4}	0.003	-0.132	0.895
poster_is_gold: true	**false**	**2.181**	**0.412**	**5.295**	**<0.01**
poster_num_submissions		-1.161×10^{-5}	0.010	-0.001	0.999
poster_total_karma		$\mathbf{5.413\times10^{-6}}$	$\mathbf{1.54\times10^{-7}}$	**35.255**	**<0.01**
poster_comment_karma		$\mathbf{3.16\times10^{-5}}$	$\mathbf{7.79\times10^{-7}}$	**40.541**	**<0.01**
poster_avg_score		**0.019**	$\mathbf{7.76\times10^{-5}}$	**241.204**	**<0.01**
poster_avg_upvote_ratios		-1.319	1.129	-1.168	0.243
poster_awardee_karma		$\mathbf{-8.046\times10^{-6}}$	$\mathbf{7.7\times10^{-7}}$	**-10.455**	**<0.01**
poster_awarder_karma		1×10^{-4}	6.12×10^{-5}	1.770	0.077
readability_flesch		-0.001	0.001	-0.494	0.621
text_num_characters		**0.001**	$\mathbf{8.75\times10^{-5}}$	**10.613**	**<0.01**
sentiment_pos		**0.324**	**0.150**	**2.161**	**0.031**
sentiment_neg		**-0.504**	**0.135**	**-3.722**	**<0.01**
tone_anger		**2.069**	**0.645**	**3.210**	**<0.01**
tone_fear		-0.236	0.729	-0.324	0.746
tone_joy		**1.303**	**0.306**	**4.257**	**<0.01**
tone_sadness		-0.518	0.361	-1.435	0.151
tone_analytical		**-0.747**	**0.243**	**-3.069**	**<0.01**
tone_confident		0.480	0.437	1.074	0.283
tone_tentative		**-1.309**	**0.246**	**-5.325**	**<0.01**

Table 20: Quantile regression model studying the popularity of baseline posts for the 0.25 quantile for the total_votes_estimate outcome.

feature_name	*baseline*	*coef.*	*std. err.*	*t*	*p*
Intercept		1.659	1.530	1.084	0.278
only_link_no_text: true	**false**	**16.427**	**0.438**	**37.484**	**<0.01**
only_image_no_text: true	**false**	**15.418**	**0.364**	**42.355**	**<0.01**
has_author_flair: true	**false**	**3.529**	**0.370**	**9.551**	**<0.01**
has_link_flair: true	**false**	**0.632**	**0.255**	**2.484**	**0.013**
is_crosspost: true	false	0.228	0.827	0.276	0.783
num_emojis		-0.006	0.006	-1.098	0.272
poster_is_gold: true	**false**	**9.723**	**0.614**	**15.823**	**<0.01**
poster_num_submissions		**-0.090**	**0.015**	**-6.175**	**<0.01**
poster_total_karma		**-2.575×10^{-6}**	**2.24×10^{-7}**	**-11.505**	**<0.01**
poster_comment_karma		**1×10^{-4}**	**1.17×10^{-6}**	**103.930**	**<0.01**
poster_avg_score		**0.192**	**0.000**	**1558.560**	**<0.01**
poster_avg_upvote_ratios		-0.728	1.643	-0.443	0.658
poster_awardee_karma		**-3.262×10^{-5}**	**1.12×10^{-6}**	**-29.023**	**<0.01**
poster_awarder_karma		-2×10^{-4}	0.000	-1.547	0.122
readability_flesch		0.002	0.001	1.396	0.163
text_num_characters		**0.003**	**0.000**	**24.478**	**<0.01**
sentiment_pos		**1.396**	**0.220**	**6.350**	**<0.01**
sentiment_neg		**-2.038**	**0.200**	**-10.173**	**<0.01**
tone_anger		**6.257**	**0.941**	**6.652**	**<0.01**
tone_fear		**-3.023**	**1.076**	**-2.809**	**<0.01**
tone_joy		**1.877**	**0.449**	**4.184**	**<0.01**
tone_sadness		**-1.149**	**0.529**	**-2.172**	**0.030**
tone_analytical		**-1.637**	**0.354**	**-4.619**	**<0.01**
tone_confident		0.260	0.647	0.402	0.688
tone_tentative		**-2.096**	**0.347**	**-6.036**	**<0.01**

Table 21: Quantile regression model studying the popularity of baseline posts for the 0.50 quantile for the total_votes_estimate outcome.

feature_name	*baseline*	*coef.*	*std. err.*	*t*	*p*
Intercept		0.006	4.101	0.001	0.999
only_link_no_text: true	false	3×10^{-4}	1.210	0.000	1.000
only_image_no_text: true	false	2×10^{-4}	1.001	0.000	1.000
has_author_flair: true	false	3×10^{-4}	0.967	0.000	1.000
has_link_flair: true	false	0.004	0.644	0.006	0.996
is_crosspost: true	false	3.722×10^{-6}	2.321	1.6×10^{-6}	1.000
num_emojis		-0.003	0.023	-0.129	0.897
poster_is_gold: true	false	3.495×10^{-5}	1.661	2.1×10^{-5}	1.000
poster_num_submissions		**0.140**	**0.036**	**3.846**	**<0.01**
poster_total_karma		**8.589×10^{-5}**	**6.15×10^{-7}**	**139.648**	**<0.01**
poster_comment_karma		**-2×10^{-4}**	**2.82×10^{-6}**	**-55.145**	**<0.01**
poster_avg_score		**1.214**	**0.000**	**3173.673**	**<0.01**
poster_avg_upvote_ratios		0.005	4.339	0.001	0.999
poster_awardee_karma		**-3×10^{-4}**	**2.92×10^{-6}**	**-99.412**	**<0.01**
poster_awarder_karma		**0.004**	**0.000**	**12.620**	**<0.01**
readability_flesch		**0.076**	**0.005**	**13.970**	**<0.01**
text_num_characters		**0.004**	**0.000**	**12.001**	**<0.01**
sentiment_pos		0.006	0.578	0.010	0.992
sentiment_neg		-0.006	0.536	-0.011	0.992
tone_anger		0.001	2.522	0.000	1.000
tone_fear		6.69×10^{-6}	2.665	2.51×10^{-6}	1.000
tone_joy		1.175×10^{-6}	1.146	1.03×10^{-6}	1.000
tone_sadness		-7.451×10^{-5}	1.288	-5.78×10^{-5}	1.000
tone_analytical		-9.395×10^{-7}	0.883	-1.06×10^{-6}	1.000
tone_confident		2.314×10^{-5}	1.719	1.35×10^{-5}	1.000
tone_tentative		7.226×10^{-5}	0.864	8.37×10^{-5}	1.000

Table 22: Quantile regression model studying the popularity of baseline posts for the 0.75 quantile for the total_votes_estimate outcome.

feature_name	*baseline*	*coef.*	*std. err.*	*t*	*p*
Intercept		-2.137	7.312	-0.292	0.770
only_link_no_text: true	**false**	**5.420**	**1.640**	**3.304**	**<0.01**
only_image_no_text: true	**false**	**4.451**	**1.369**	**3.250**	**<0.01**
has_author_flair: true	false	-0.778	1.257	-0.619	0.536
has_link_flair: true	**false**	**-3.072**	**0.905**	**-3.393**	**<0.01**
is_crosspost: true	false	0.444	3.307	0.134	0.893
num_emojis		-0.007	0.012	-0.611	0.541
poster_is_gold: true	false	0.062	2.364	0.026	0.979
poster_num_submissions		**0.872**	**0.055**	**15.854**	**<0.01**
poster_total_karma		**8×10^{-4}**	**9.07×10^{-7}**	**886.985**	**<0.01**
poster_comment_karma		**-0.001**	**2.76×10^{-6}**	**-435.719**	**<0.01**
poster_avg_score		**3.596**	**0.000**	**9343.059**	**<0.01**
poster_avg_upvote_ratios		-1.507	7.700	-0.196	0.845
poster_awardee_karma		**-0.001**	**4.05×10^{-6}**	**-351.866**	**<0.01**
poster_awarder_karma		**0.017**	**0.001**	**25.224**	**<0.01**
readability_flesch		**0.023**	**0.008**	**3.095**	**<0.01**
text_num_characters		**0.002**	**0.000**	**3.560**	**<0.01**
sentiment_pos		**1.610**	**0.791**	**2.036**	**0.042**
sentiment_neg		**1.673**	**0.783**	**2.136**	**0.033**
tone_anger		-0.429	3.714	-0.115	0.908
tone_fear		0.306	3.978	0.077	0.939
tone_joy		0.895	1.604	0.558	0.577
tone_sadness		-0.156	1.921	-0.081	0.935
tone_analytical		-0.880	1.248	-0.705	0.481
tone_confident		-0.388	2.429	-0.160	0.873
tone_tentative		-1.747	1.212	-1.442	0.149

Table 23: Quantile regression model studying the popularity of baseline posts for the 0.90 quantile for the total_votes_estimate outcome.