RESPONSIBILITY, TRUST, AND MONITORING TOOLS FOR END-USER ACCOUNT SECURITY

by
Yomna Abdelaziz

A thesis submitted to
the Faculty of Graduate and Postdoctoral Affairs
in partial fulfillment of
the requirements for the degree of

MASTER OF APPLIED SCIENCE

Human Computer Interaction

at

CARLETON UNIVERSITY

Ottawa, Ontario
September, 2018

Abstract

Threats to online accounts are increasingly more sophisticated and proactive defences may be insufficient. We explore the role of users in security monitoring of their activity logs. First, we designed and prototyped an account monitoring app that allows users to monitor the activity of many accounts at once. We present the prototype account monitoring app and the results of a lab study with 15 participants to explore its usability. Besides usability results, we identified external factors influencing adoption relating to user trust of security tools and service providers. We next conducted a second study of 170 participants in an online survey to explore responsibility, trust, and monitoring for account security. We identified a mismatch in perceived responsibility between users and service providers, explored the trust cues participants use to trust their service providers, and explored end user activity log practices for account monitoring. We also designed and evaluated two updated activity log designs based on feedback from our first study.

Acknowledgements

All praise and thanks to God for blessing me with this opportunity and for sustaining me with His eternal Love and Kindness throughout my journey.

I want to acknowledge my awesome supervisor, Sonia Chiasson, for her mentorship and support. I am grateful for the opportunity you gave me to join your lab. Your commitment to the success of your students continues to inspire me. I am truly lucky to have been one of them.

Thank you, Robert Biddle, and Kasia Muldner for agreeing to be on my thesis committee. Thank you, Kasia, for helping me discover a love for statistics. Thank you, Lois Frankel for agreeing to chair my defense.

Thank you, Erenia Oliver, for helping me apply to the HCI program and for your invaluable advice. Thank you, Carleton University, for giving me the opportunity and funding to pursue this degree.

I acknowledge all the participants who graciously volunteered their time to take part in my studies.

I am grateful to all the members of the CHORUS lab and others for friendship and good times. Sana, Hala, Reham, Khadija, Daniela, Kalpana, Michael, Eric, Yosra, Mohamed, Mariam, Mostapha, thank you for pilot testing and feedback during earlier stages of my research. I wish you all success in your future endeavours.

To my parents, Somaia and Gamal, and my siblings, Mohamed, Yosra, Ahmed, Mostapha, Mariam: thank you for your continued support and prayers. I am eternally grateful for your love.

To my sister, Yosra: thank you for your unwavering support and kindness. This journey would not have been possible without you.

To my daughter, Khadeeja, and son, Hamza: thank you for your patience and for giving me the motivation I needed to succeed. This thesis is dedicated to you.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In this Chapter, we discuss our motivation behind our thesis and the research directions we explore. We define the scope of our research and identify our contributions.

1.1 Motivation

Threats to online accounts are increasingly more sophisticated and proactive defences may be insufficient. As an example, a phishing attack is the attempt to obtain sensitive information such as usernames, passwords, and credit card details (and money), by disguising as a trustworthy entity. This can take the form of a website designed to look the same as another well-known website. It can take the form of an email that appears to be from a friend. It is dangerous because it can be very convincing [5], and users are often tricked into disclosing personal information or credentials, resulting in the compromise of their online accounts [120]. Malware is software that is designed to damage computers or steal data [47]. Users can be tricked into downloading malware through phishing, spam emails, posts on social media, and other means. It is dangerous because users lose control of their computers or online accounts, suffering consequences such as lost data or a ruined reputation. Password cracking attacks and password guessing attacks [137] are dangerous because they involve sophisticated methods and are successful against a minority of user accounts [120]. The use of strong passwords alone is therefore not enough. End users are being advised to adopt other proactive security measures including two-factor authentication and the use of password managers [120]. However, if these proactive measures are bypassed by attackers, then reactive measures should be in place to detect potential compromises. Account monitoring by end users is one such reactive measure.

According to Dashlane[1], the average user has 90 online accounts [17]. Most service providers expect their users to play a role in protecting their accounts. For example, Facebook expects users to "not share [their] passwords, give access to [their] Facebook account to others, or transfer [their] account[s] to anyone else (without [Facebook's] permission)" [37]. Google expects users to "keep [their] password[s] confidential [...], try not to reuse [their] Google account password on third-party applications, [and] if [they] learn of any unauthorized use of [their] Google Account," then take steps to recover their account. Google holds users responsible for "the activity that happens on or through [their] Google account" [49]. Consequently, it is also important to understand how users perceive their responsibility for protective measures, in relation to their service providers.

If users must play a role in detecting unauthorized access and recovering their accounts, they need accessible reactive measures. A reasonable thing to do is pool their account activity so that they are not having to check 90 different sources. Such aggregation tools are available in other domains. Aires and Goncalves do this with their "Personal Information Dashboard," which allows users to combine several account activity feeds through plugins to enable users to see interesting patterns in their disclosed information. uTrack by Rodrigues *et al.* [96] allows users to track how the content they upload to social media sites diffuses throughout the web. However, pooling information from multiple sources for self-management is not common.

1.2 Research Directions and Scope

We take the following research directions:

- **RD1.** Exploring the use of combined activity logs for account security.
- **RD2.** Exploring attribution of responsibility and trust in account security.

For our first research direction, we limit our exploration of account activity logs to historical login data and the typical meta data made available by service providers. This meta data includes information such as the location of the account access events, and the type of device used.

[1] A password manager.

For our second research direction, we limit our exploration of responsibility and trust to two entities, the user and the service provider, and account security as it relates to user-facing threats. For service providers, we focus on Facebook, which has 2.23 Billion active users as of June 2018 [57], and Google, which has seven products with over 1 Billion active users each as of May 2017 [69]. To place these numbers into perspective, there are an estimated 4.1B internet users worldwide in 2018 [75].

1.3 Contributions

We list our main contributions as follows:

1. We designed, prototyped, and tested a combined activity log tool in a lab study with 15 participants. We identified usability and design issues with the app, and more specifically with the visual activity log contained therein. Based on these findings, we redesigned our visual activity log for testing in Study 2 with 134 participants.

2. We identified external factors contributing to user trust of security tools, and online service providers more generally. The absence of these factors caused our participants' lack of desire to adopt our app in Study 1. In Study 2, the same factors were reasons why our participants (N = 170) trusted their service providers, Facebook and Google.

3. We identified a mismatch in perceived responsibility between users and service providers. In Study 1, we found that a portion of our participants would not adopt the app because they believe that responsibility for account monitoring lies with the service provider. In Study 2, we found that participants (N = 170) identified clear roles with respect to primary responsibility for prevention, alerting/reporting (monitoring), and recovery from different types of user-facing attacks. We compare these findings with Facebook and Google's existing terms of use.

4. We provide a look into end users' activity log practices. In Study 1, we found that 80% of our participants (N = 15) do not check their account activity logs.

In Study 2, we found that 39% of the Facebook participants (n = 95) and 51% of the Google participants (n = 75) report checking their activity logs at least once a year. The main reason why participants do not check their activity logs is that they do not know that this information exists, however, they indicated a desire to check them.

1.4 Thesis Outline

In *Chapter 2*, we provide an overview of related work about trust in an online context, system-side and user-side account monitoring capabilities, expectations of responsibility surrounding account security, visualizations for security purposes, and relevant research that inspired our methodology.

In *Chapter 3*, we present the findings of *Study 1: Account Monitoring App* on participants' perception of our prototyped account monitoring tool and external factors impacting the potential for adoption of our tool.

In *Chapter 4*, we present findings of *Study2: Online Survey* on participants' perceived distribution of responsibility for their account security, the trust cues participants use to trust Facebook and Google, and the performance of two activity log designs.

In *Chapter 5*, we discuss the implications of our findings, future research directions, and limitations of our studies. We end the chapter with a conclusion.

Chapter 2

Background

We discuss what trust means in an online context, and the methods that people use to derive trust in entities. We cover account monitoring next, followed by the methods that systems and human can use for that purpose. We discuss the expectations that entities have regarding account monitoring, and the responsibility for doing so. This is followed by a brief overview of how visualizations are used as a monitoring tool in enterprise contexts, and how the same concept can be extended to account monitoring for end users. We finally give a brief overview of existing research that has inspired our study methodology.

2.1 Trust

In this section, we discuss how users trust computer systems and how this applies to Facebook and Google. For the purpose of our paper, an entity refers to a human (e.g., end-user) or non-human (e.g., service provider) agent who is involved in an online transaction. For example, a Facebook user is an entity who receives a service from another entity, Facebook, in exchange for her usage data and for being an audience to advertisers.

2.1.1 Definition and Beliefs

Trust is the willingness to be vulnerable to the actions of another entity [70] with the expectation that the other entity will not violate one's rights [56]. When it comes to interactions in the online world, trust is arguably more important due to a higher level of uncertainty. When an entity decides to trust another, her awareness of risk decreases and she continues engaging in transactions with that entity without re-calculating risk every time [43]. This is how phishing attacks work; by impersonating a person or website that the user already trusts [43]. Trust can manifest in different

beliefs and can apply to entities differently. Mayer and Schoorman [70] identify three beliefs about an entity which result in people trusting that entity: ability/competence, integrity, and benevolence. Gefen [48] applies this framework to participants' interaction with Amazon and finds that the three beliefs are highly correlated with one another, *i.e.*, participants are likely to hold the three beliefs simultaneously about Amazon. It is important to separate the three dimensions of trust however, because each dimension will determine what kinds of transactions people are willing to partake in with entities online [48]. For example, the competence of an online service provider (SP) is important for window-shopping, yet benevolence and integrity are more important for data sharing and purchasing products. This highlights that people trust entities for different reasons. For example, Jane trusts online Vendor A's competence and Vendor B's benevolence, so she reserves her research for Vendor A, and reserves her purchases for Vendor B. The beliefs that people hold about entities are not limited to Mayer and Schoorman's dimensions.

2.1.2 Trust Cues

People use different *trust cues* to arrive at a belief that an entity is trustworthy for one reason or another. One common cue is *transitive trust* [61]. Transitive trust is when trust can be transferred between entities, as illustrated by the following: "when Alice trusts Bob, and Bob trusts Claire, and Bob refers Claire to Alice, then Alice can derive a measure of trust in Claire based on Bob's referral combined with her trust in Bob" [61].

Another cue by which people decide to believe that an entity is trustworthy is *reputation*. Reputation, as defined by Artz and Gil [8], is "an assessment based on the history of interactions with or observations of an entity, either directly with the evaluator (personal experience) or as reported by others (recommendations or third party verification)." A good reputation is arguably a broader manifestation of transitive trust because it is a combination of people's testimonies, or reviews, which other people use as a basis for their trust decisions. Researchers have explored reputation score algorithms, such as those used on platforms like Ebay, that take different factors into account [59, 60, 77–80, 132], such as the weight of a rating. A

rating can be weighted by how old it is, the rater's trustworthiness [61], the platform on which the rating exists, etc.

In addition to transitive trust and reputation, *personal experience* is a third cue of deriving trust [61]. When an entity receives a service, they can assess the competence and integrity of the service provider more directly. Trust derived from personal experience can override a bad reputation [61, 126].

A fourth cue of deriving trust is *visual indicators or design factors*. Sillence *et al.* [107] found that users took design factors into account when making trust decisions about the competence and integrity of medical websites. A complex, busy layout, or in contrast, a boring design are two of the 11 design factors that caused their participants to reject websites. Stephens [113] found that page layout, navigation, style, graphics, and content significantly affected whether or not people would trust a website enough to book a hotel through it. Surprisingly, web seals did not impact trust.

Trust cues are not limited to the ones discussed here, and can take the form of other factors. It is also important to note that not all consumers of online services use trust cues in the same way, nor are trust cues used exclusively from one another. For example, one person may not use transitive trust for a particular service provider, but another person will. People may also combine visual indicators with reputation and other trust cues before they decide to trust a website's competence. A person's trust in a website's competence does not necessarily mean that they also trust its integrity. For the purpose of this thesis, we focus on a particular set of trust cues. We use the dimensions of trust [70] as applied by Gefen to online interactions [48], and the cues users employ in deriving trust [61, 107, 113] to provide a context for our research. In an online context, users can derive trust in a service provider's competence by personal experience (from using the service) as well as by transitive trust (i.e., recommendations from others) and reputation (i.e., online reviews). Integrity can be proven by reputation, as well as transitive trust and personal experience. As for the third dimension of trust, benevolence [70], it is out of scope for our research but can also be derived using trust cues.

It is difficult to apply the dimensions and cues of trust to Facebook and Google for two reasons. Firstly, the existing research on trust of online service providers is

incoherent due to a lack of a common trust model. Secondly, the reported perceptions of Facebook and Google's reputations mostly exist in non-academic literature and so the underlying causes of their reputation is not explored. With these considerations in mind, we report on the state of these two companies' reputation. Due to data leak allegations [102], the appearance of false news articles [101], and privacy concerns [13, 88, 114], it is our opinion that Facebook's reputation has seen a decline over recent years [119]. A 2018 study by Edison Research [34] finds that Facebook usage has dropped by 5%. Thousands of users have decided to disconnect, and some have indicated that they want to, but find it difficult to do so [13]. Harris Poll finds that Google's reputation has also declined over the years, but has generally maintained a higher reputation ranking that Facebook [119]. This is interesting given that Google is likely the largest data miner in the world, making the consequences of data leaks, false news articles, and privacy breaches more encompassing. However, Google differs fundamentally from Facebook as a service provider because it is more known for its search engine, email service, and productivity tools, not social media. In fact, Google's social network, Google+, failed to gain traction since it was introduced in 2011 [33]. Facebook, on the other hand, is the largest social network with 2.2 billion active users [62]. The way people perceive Google's service offerings could play a role in the state of its reputation as opposed to that of Facebook.

2.2 Account Monitoring

2.2.1 Existing Threats that Necessitate Monitoring

We provide a brief overview of existing threats to account security that necessitate account monitoring. A 2017 survey by Pew Research Centre finds that 16% of American Internet users have had an email account compromised, and 13% have had a social networking account compromised [87]. Gao *et al.* [47] discuss types of attacks in online social networks: phishing, impersonation or identity theft, de-anonymizing attacks, malware, and cross-site request forgeries [42]. One example of malware is Koobface [135], a worm which activates Facebook accounts using Gmail addresses, adds friends, then posts URLs on those friends' walls which link to the Koobface

loader component. Another threat comes from security flaws of the single sign-on functionality that allow attackers to gain access to their victims' accounts [128]. Due to scope, we do not explore all of these threats in Study 2 (Section 4.4.2), but we believe that account monitoring plays a crucial role in the prevention, detection, and recovery from these attacks. We explore the responsibility for preventing phishing, password stealing and password guessing.

Gao *et al.* [47] argue that defense mechanisms against phishing attacks are more reasonably implemented on the client side because end users are often willing to give out personal information. For example, browser toolbars can notify users of suspicious websites when those websites prompt them for credentials. However, Gao *et al.* [47] also say that "for the majority of security threats, if users don't take the initiative to protect their information, most server-end defenses would fail disastrously." Similarly, Thomas *et al.* [120] found that Google users are 400 times more likely to have their accounts compromised from a phishing attack than other types of attacks. We believe that providing accessible account monitoring methods would contribute to user awareness and early detection of attacks.

2.2.2 System-Side Monitoring

Brief Overview of Web Usage Mining

Account monitoring is carried out by means of web usage mining. Nina *et al.* [84] define web usage mining as "the discovery of meaningful patterns from data generated by client-server transactions on one or more Web localities." This process consists of three phases: (1) data preparation, (2) pattern discovery, and (3) pattern analysis and visualization [84]. They outline general algorithms for data preparation, user identification, and session identification. Many researchers have demonstrated different methods and algorithms of web usage mining [29, 36, 58, 64, 81, 116], including using pattern tree algorithms [50], website architecture analysis [103], graph theory [82], and user clustering algorithms [53]. A recent survey on web mining by Anitha and Isakki [6] shows that existing web mining techniques can be used to predict user behaviour.

Bhargav and Bhargav [14] demonstrate the notion of using weblogs to identify the

usage patterns of users, and in turn use that knowledge to make websites more usable. For example, if the usage patterns of a segment of users from a particular country indicate the occurrence of a special cultural holiday, search engines can then provide information and resources about that holiday for future site visitors. Intelligent tools exist to monitor the online activity of end users, yet this data is used by service providers, advertisers and other third parties to advance their interests. We propose that users should also be able to make use of their own activity logs to keep themselves safe by being aware of their own usage patterns.

Account Monitoring by Service Providers

Researchers at Google [120] have explained how they implement system-side monitoring to help protect user accounts. A similar process is employed by Facebook. They do this by building a risk profile of each user. This risk profile contains a user's historical access patterns, locations of access, and the devices they use. When a login attempt does not match a user's risk profile, Google blocks the login or requires additional authentication information. In the event of the account being hijacked, Google not only locks out any new access attempts, but also severs all existing sessions. The rightful user must then prove her ownership of the account either by providing a code that she received on another email account or via text message on her phone, answering a fallback authentication question [1], or identifying her prior account access times. This lockout strategy is generally effective against password guessing and stealing attacks, yet not as effective against phishing.

In the event that a successful login takes place which does not match the user's risk profile, Google alerts the user by sending her an email or notification on her phone. In this alert, the user is instructed to take steps to secure her account if the activity is not hers. Figures 2.1 and 2.2 show examples of the security alerts generated by Google and Facebook.

Priambodo *et al.* [92] demonstrate that Facebook users' account activity patterns

[1] a fallback authentication question [93] asks the user a question about her personal information that is presumably private to her, for example, "what is your mother's maiden name?" These fallback authentication questions are chosen by the user at the time of signing up for her account. Rabkin [93] discusses the security limitations of authentication questions.

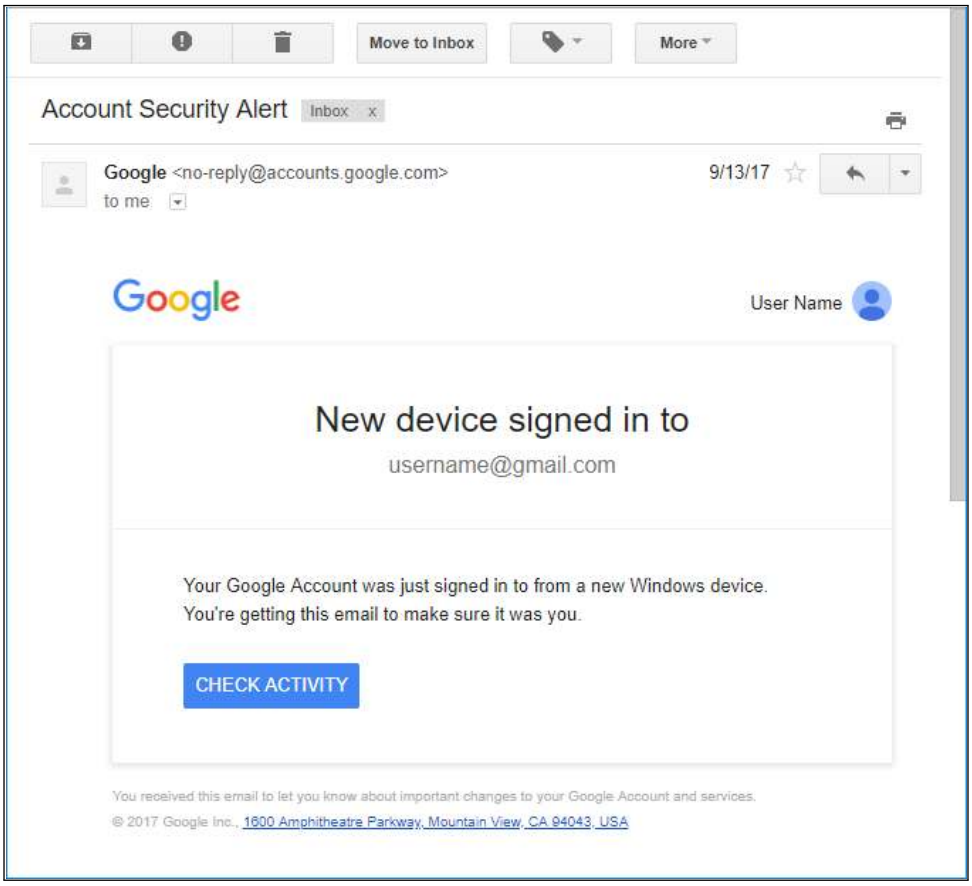


Figure 2.1: Existing Gmail security alert.

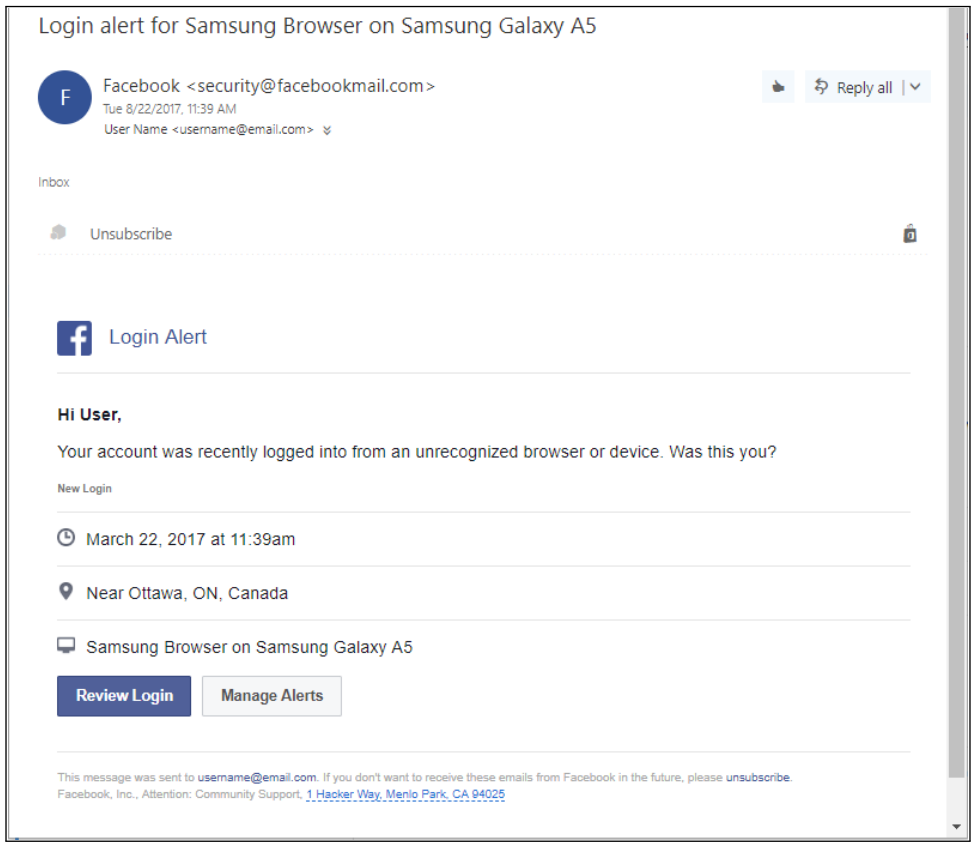


Figure 2.2: Existing Facebook security alert.

differ between mobile and desktop access. They analyzed the transaction and activity logs of 515 usernames and 293 mobile devices actively connected to Facebook, over a two-month period. They find that mobile usage activity by Facebook users differs from desktop in that multiple sessions are opened at once, sessions can last longer (even more than seven days), and that the same account can be accessed through active sessions from more than one mobile device. These findings hold important implications for account monitoring because it points to the need for an intelligent and encompassing monitoring system that can aggregate all of the account activity logs that are generated by the same user on multiple devices.

2.2.3 User-Side Monitoring

In the case of Facebook and Google, users are able to monitor their accounts by accessing the activity logs that are made available to them. Not all service providers make account activity logs available, however. To access their activity logs, Facebook users must first download their data from the Settings section under "Your Facebook Information" and ensure that "Security and Login Information" is selected. After downloading their Facebook data, users can then view their activity logs by opening the index.html file (which can be viewed in a browser like a website) and clicking "Security." Figure 2.3 shows an example of a Facebook activity log.

Google also makes account activity logs available to end users, which they can access by clicking "Details" at the bottom of their Gmail inbox (Figure 2.4). Security alerts for unusual activity are on by default, but users can disable them, as depicted in the bottom of Figure 2.4[2]. For a high-level listing of account activity across Google services (such as Android, Play Store, Chrome, etc), users can click "Google Account," then "Go to my activity" under "Personal info and privacy." Alternatively, users can directly access this high-level activity through "myactivity.google.com." This activity log, however, does not show the login history for Gmail, nor the IP addresses or locations of the access events.

In our opinion, Facebook and Google's existing account activity logs suffer from usability issues. Users may not know where to find them, let alone that they exist.

[2]https://support.google.com/mail/answer/45938?hl=en

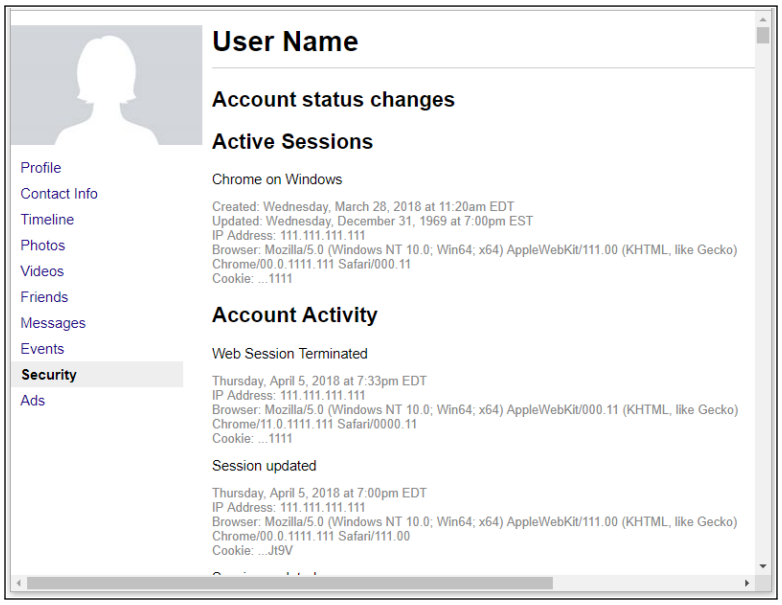
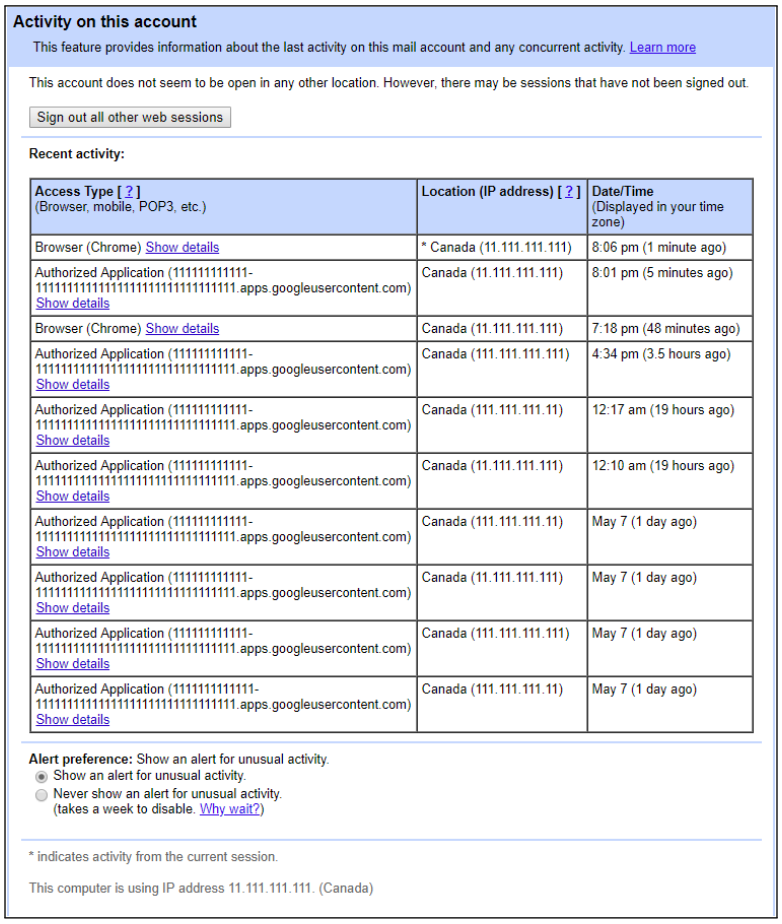


Figure 2.3: Existing Facebook activity log.

If they do find them, the content may be confusing. In addition, users do not want to be notified of every account activity event unless it requires action on their part. This is partly because it is tiresome and boring [43]. Furthermore, too many alerts from the service provider can cause warning fatigue, or habituation [138].

User-side Prevention

An important approach to account security is to help users prevent unwanted account access. Password managers [11, 24, 71, 110] and two-factor authentication mechanisms [127] [28,65] can be used for that purpose. In addition, researchers have demonstrated how educating users by way of persuasive media [73, 139] and design [44] can help users make secure decisions that would protect their accounts. For example, Zhang-Kennedy *et al.* [140] provide updated password advice to better support end users in keeping their accounts secure. Hayashi *et al.* [55] propose a device-level security mechanism that controls access to certain accounts or apps when the device is unlocked. They argue that this method is more usable than all-or-nothing device access control because it will enable device sharing while reducing the risk of unauthorized account access, whether accidentally or maliciously.



Activity on this account

This feature provides information about the last activity on this mail account and any concurrent activity. Learn more

This account does not seem to be open in any other location. However, there may be sessions that have not been signed out.

Sign out all other web sessions

Recent activity:

Access Type [?] (Browser, mobile, POP3, etc.)	**Location (IP address) [?]**	**Date/Time** (Displayed in your time zone)
Browser (Chrome) Show details	* Canada (11.111.111.111)	8:06 pm (1 minute ago)
Authorized Application (111111111111-11111111111111111111111111111111.apps.googleusercontent.com) Show details	Canada (11.111.111.111)	8:01 pm (5 minutes ago)
Browser (Chrome) Show details	Canada (11.111.111.111)	7:18 pm (48 minutes ago)
Authorized Application (111111111111-11111111111111111111111111111111.apps.googleusercontent.com) Show details	Canada (111.111.111.111)	4:34 pm (3.5 hours ago)
Authorized Application (111111111111-11111111111111111111111111111111.apps.googleusercontent.com) Show details	Canada (111.111.111.11)	12:17 am (19 hours ago)
Authorized Application (111111111111-11111111111111111111111111111111.apps.googleusercontent.com) Show details	Canada (111.111.111.111)	12:10 am (19 hours ago)
Authorized Application (111111111111-11111111111111111111111111111111.apps.googleusercontent.com) Show details	Canada (111.111.111.11)	May 7 (1 day ago)
Authorized Application (111111111111-11111111111111111111111111111111.apps.googleusercontent.com) Show details	Canada (111.111.111.111)	May 7 (1 day ago)
Authorized Application (111111111111-11111111111111111111111111111111.apps.googleusercontent.com) Show details	Canada (111.111.111.111)	May 7 (1 day ago)
Authorized Application (111111111111-11111111111111111111111111111111.apps.googleusercontent.com) Show details	Canada (111.111.111.11)	May 7 (1 day ago)

Alert preference: Show an alert for unusual activity.
- (•) Show an alert for unusual activity.
- ( ) Never show an alert for unusual activity. (takes a week to disable. Why wait?)

* indicates activity from the current session.

This computer is using IP address 11.111.111.111. (Canada)

Figure 2.4: Existing Gmail activity log.

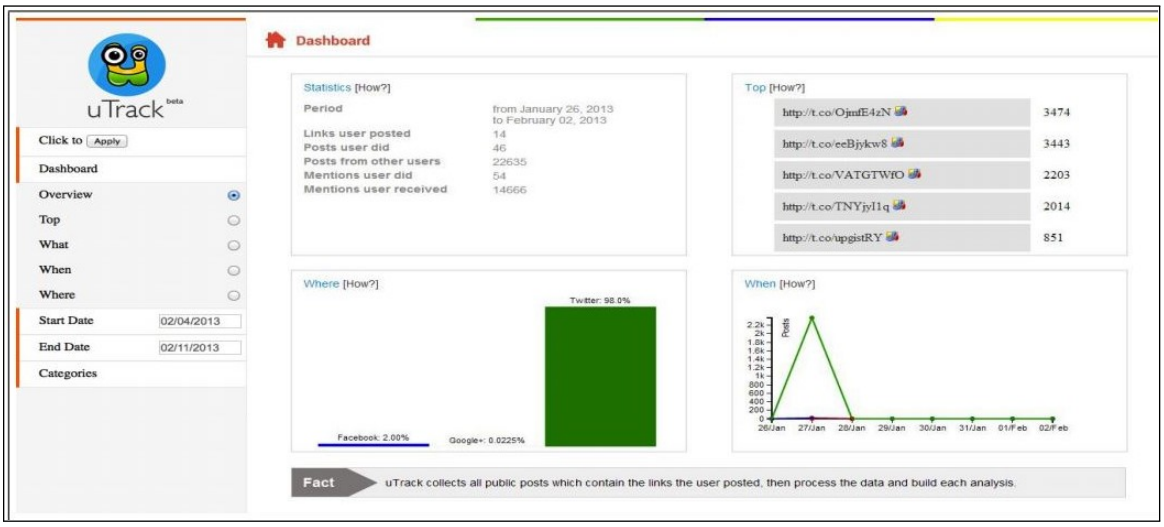


Figure 2.5: "uTrack" by Rodrigues *et al.* [96] allows users to track the diffusion of their social media data.

Related Tools

Other than the account activity logs that some service providers produce, we are unaware of any existing tools that enable end users to monitor the security of their accounts. There are tools, however, that allow users to pull in their data from multiple accounts for other purposes. The "uTrack" application (Figure 2.5) by Rodrigues *et al.* [96] is a dashboard application that displays visualizations of information diffusion across social networks. Users can track how and where the content they post on social media sites is shared. Aires and Goncalves [1] demonstrate the use of plugins to allow users to pull in data from multiple accounts, including Gmail, Twitter, Flickr[3], and Panoramio[4]. They explain that the motivation behind their "Personal Information Dashboard" (Figure 2.6) application is to combine several sources of personal information to show users "interesting facets and patterns" of their lives. They argue that although applications that combine different information sources exist [32, 68, 74, 121–124], none of them do so in a cohesive manner like their tool.

[3]https://www.flickr.com
[4]https://www.cnet.com/news/google-panoramio-kills-photo-sharing-users-angry/

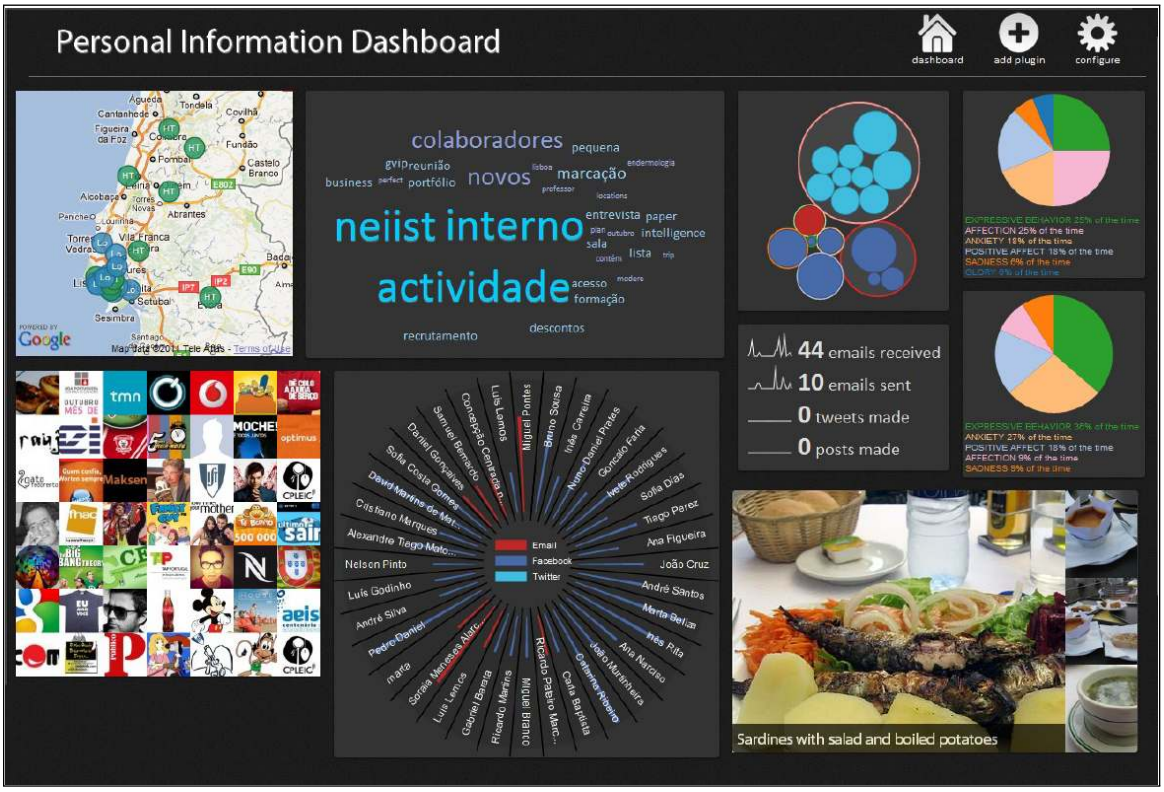


Figure 2.6: "Personal Information Dashboard" by Aires and Goncalves [1] allows users to pull data in from several accounts.

2.2.4 Expectations and Responsibility

The Problem of Responsibility

When it comes to monitoring accounts, it can be unclear where the service provider's responsibility ends and where the user's begins. Responsibility, as defined in the Merriam-Webster dictionary, means "liable to be called on to answer [...]; liable to be called to account as the primary cause, motive, or agent [...]; being the cause or explanation." Nissenbaum [85] identifies four challenges to responsibility in the digital age: (1) the problem of "many hands", *i.e.*, it takes many people to build software or an online service (2) the acceptance of software bugs as an inherent part of applications and systems, (3) the use of computers as scapegoats, e.g., they are too complex and one cannot always control the outcome and (4) ownership without liability. Nissenbaum argues that companies tend to claim ownership of their software, but dismiss the corresponding responsibilities that traditionally come with ownership. For example, software license agreements can assert the manufacturer's ownership of the software, but deny responsibility for damages resulting from defects in the software[5]. One can argue that (3) the use of computers as scapegoats and (4) ownership without liability can make it difficult to ascribe responsibility when things go wrong from the end user's side as well. The issue of responsibility in the digital realm is constantly debated [26, 27, 83, 108]. We will not add to this philosophical debate, rather we are interested in how users perceive responsibility.

User's Perception of Responsibility

In an online survey, Shay *et al.* [104] found that most participants either attributed responsibility to themselves (37%), or shared the responsibility with the service provider for preventing account compromises (40%), regardless of whether they had experienced an account compromise. Shay *et al.* take this as a favourable pretext for system design: "these high rates at which participants acknowledged some responsibility suggest [...] that at least one barrier [to adopting security tools] user attitudes may be overcome." For the purpose of this thesis, we focus on responsibility, not

[5]This is true of Facebook and Google's terms of use. See https://www.facebook.com/terms, https://policies.google.com/terms

accountability because the former implies a subjective, moral meaning, whereas the latter implies an objective legal meaning. We believe that it is easier for end users to provide their perspective on responsibility rather than accountability.

Service Provider's Perception of Responsibility

In the case of Facebook, Nadon *et al.* [83] argue that users are held responsible for their decision to use its service. To act responsibly, users must first know to what they are agreeing. Although all the necessary information is made available by Facebook for its users, the availability of this information does not necessarily result in responsible decision-making by the end users [108, 109]. The terms of service are difficult to comprehend and change often. Nadon *et al.* [83] perceive that end users are frustrated with "unrealistic responsibilities of acquiring encyclopaedist knowledge, in order to engage in informed consent [86]."

End-User License Agreements (EULAs)

Mannan and van Oorschot [67] identify a gap between what online banking providers expect in their terms of use and users' practices. In their survey of 123 technically advanced users, they found that many security requirements are too difficult for users to meet. They argue that the assurances that banks provide about account security are misleading because most users would be unable to meet the security requirements that the banks expect. Not only do users not read the EULA, but if they were to do so, they are unlikely to understand it. Researchers have tried to design better EULAs [125], and many observe that it takes an enormous amount of effort to make each decision an informed one [15,67,72]. McDonald and Cranor estimate that typical users would need 201 hours every year to fully read and understand their EULAs [72], based on the estimate that people visit 112 unique websites per month.

2.3 Data Visualization

Information visualization is the graphical representation of data to aid in its cognition [21]. The process of graphically presenting data is typically computer-supported.

Visualizations function as "cognitive tools," artifacts that support decision making [129]. For the purpose of this thesis, we define visual encodings as graphical characteristics (such as shape, placement, etc) that represent variables (such as location, time, etc.).

Gestalt laws are rules that describe the innate human ability to perceive patterns in visual stimuli. They play an important role in the design of visualizations [129]. We describe four Gestalt laws here. *Proximity* indicates that closely-placed objects are perceived to be of the same group. For example, data points that cluster in a scatterplot convey a close association with one another. *Similarity* indicates that stimuli that resemble each other in appearance are seen as part of the same group. For example, data points represented by visual encoding A appear as a distinct group from those represented by visual encoding B. *Closure and common region* indicates the ability to infer the complete form of a shape, even when parts of its contour are missing. It also means that the human eye can infer distinct contours within overlapping shapes. Closure is important for defining which data belongs inside, and which is outside. For example, a Euler diagram shows that entities can simultaneously be members of multiple groups, but not of others. *Figure-ground* indicates that objects are perceived to be in the foreground, and whatever lies behind is the background. Other Gestalt laws, such as *closure*, make an object distinct from its background. For example, bars in a bar graph that have highly saturated colours make them distinct from their white background.

When choosing visual encodings in visualizations, Ware [129] outlines several principles to guide design. *Preattentive processing* "determines what visual objects are offered up to our attention and easy to find in the next fixation" [41,129]. This makes *preattentive cues* faster to find than *non-preattentive* ones. Common visual encodings such as shape, colour, and orientation are preattentively processed.

Redundant coding [129] can be leveraged to allow the user to search for data points by several types of characteristics. For example, data points can be made distinguishable not only by colour, but also by shape and spatial placement. An advantage of redundant coding is that it can allow for *analytic processing* [129]. This means that people can infer information from each characteristic, independently from

others. For example, in data point A, colour represents one variable, shape represents another, and placement represents a third variable.

Such design principles are involved in the design of glyphs. A glyph is "a small independent visual object that depicts attributes of a data [point]" [16]. In their survey on glyphs, Borgo *et al.* [16] found that using glyphs depends on the type of data being visualized. For example, Fuchs *et al.* [46] found that line glyphs (enclosed line shapes) are a good choice for detecting temporal trends and peaks, whereas radial glyphs performed best when participants needed to find a particular point in time. Feng *et al.* [39] found that 3D sphere glyphs performed better than to superquadric glyphs (which map four variables) in identifying correlations and estimating data values.

2.3.1 Security Visualizations

System administrators deal with large data sets for the purpose of monitoring and intrusion detection. Many data aggregation methods have been developed to enable security-relevant analysis by technical users. One such method, visual data mining, combines graphical representation of complex data sets with a user interface to manipulate the data in search of patterns. In 1997, Cox *et al.* [30] describe NicheWorks, a visualization interface depicting a high-level summary of calling patterns and usage by Bell telephone subscribers. Since then, a diverse number of visualization interfaces have been developed for security analysts and system administrators. For example, visualizations exist to show internet anomalies [118], network intrusions [76], fake social network accounts [136], web server attacks [4], malicious log-ins in enterprise networks [106], and source code vulnerabilities [9].

When it comes to non-technical end users, however, the study of visualizations specifically for account security is lacking. Of the little research we found, Trust Neighbourhoods by Elmqvist and Tsigas is a visualized system depicting trust relationships in a distributed file sharing system [35]. In 2014, Thomas Steiner describes an application that visualizes server activity logs to inform end users, in real-time, of the amount of Wikipedia edits done by humans and those by done by bots [112].

2.3.2 Visualizations with Account Monitoring Implications

Fu *et al.* [45] demonstrate the utility of "T-Cal," a calendar-based visualization they designed to depict online conversation activity. During two case studies with four expert users, *T-Cal* was used to depict the activity on Slack channels in an enterprise context and Slack channels in an educational context. "T-Cal" interactively displays data on the year, month, week, day, and text corpus levels. This fine-grained exploration of data can fulfill different goals such as identifying problems during the development lifecycle of a product or understanding the communication strategies that students use in collaborating for a course project. Fu *et al.* point to the applications of *T-Cal* outside of team messaging platforms, and we believe that their design can extend to online account monitoring.

Fuchs *et al.* [46] evaluated four types of glyphs that visualize large-scale time-series data. The glyphs encoded the position of the datapoint in time and its quantitative value. They argue that line glyphs (enclosed line shapes) are a good choice for detecting temporal trends and peaks in the data. Radial glyphs performed best when participants needed to find a particular point in time. Similar to the idea behind behind clocks, radial glyphs use a circular layout to encode time. In this thesis, we also explore the use of glyphs for time-series account activity log data, but our interest has to do with finding patterns, not trends.

Rodrigues *et al.* [96] present "uTrack" (Figure 2.5), a dashboard application that visualizes information diffusion across social networks. It uses APIs to connect to a user's social media accounts and then collects all the URLs of the user's content. It then searches for posts containing those URLs in other social networking sites. This information is then analysed and displayed to the user in the form of visualizations and statistics. Rodrigues *et al.* do not mention if they conducted a user study of "uTrack," so we are unaware if users would trust this third-party application with access to their accounts. Furthermore, the effectiveness of the visualizations needs to be investigated.

Aires and Goncalves [1] define *personal information* as "information from or to someone not always owned or controlled by the subject." This includes artifacts such as the files one creates, the content one posts on social media, and the emails one

sends. They are dispersed across different websites and platforms but generate a holistic narrative when brought together. Aires and Goncalves propose their "Personal Information Dashboard" as a cohesive way of tracking one's personal information. It is a web application that uses plugins to retrieve data from different sources and combine it into one interface. It is customizable and contains details on demand in the form of tooltips[6]. The Personal Information Dashboard is also enabled with *plugin intercommunication*, "an action in one visualization can affect other visualizations, for instance, clicking on a word in one visualization will fade out all the other words in [the] other plugins." Aires and Goncalves tested their application with usability tests ($N = 16$), in which the participants were supplied with data, and case studies ($N = 5$) in which the participants used their own data. Findings from the usability tests were generally positive, but led them to theorize that interactive visualizations have less usability than the static ones. The case studies were generally positive, revealing that users gained insight into their data that they did not notice before. There is no discussion of the privacy or security concerns that participants may have about this application accessing their accounts.

2.4 Methodological Approach

In this section, we discuss studies relevant to our methodology.

2.4.1 Scenario-Based Studies

A simulation [51] involves using props, such as prototypes, often in the lab to simulate aspects of use in the real world. It combines the use of scenarios [22] or role-playing to enable researchers and participants to simulate a use scenario. Simulation studies are used when natural settings cannot be sufficiently accessed for the purpose of evaluation, or when a use context does not yet exist. We take a scenario-based approach for our first study involving a prototyped account monitoring app for mobile devices. Within the field of usable security, scenarios can be particularly useful in studies involving privacy [90] or threats to security [99] [66] because it is challenging

[6]A tooltip is a message that appears when the user positions the mouse cursor over an element such as an icon, image or hyperlink. [117]

to ethically study these scenarios in the wild. For example, Woodruff *et al.* [133] used scenarios to gain a better understanding of people's privacy attitudes in relation to their practices. One scenario was, "A marketing company offers you $1000 and free genetic testing in exchange for the rights to all your current and future medical records. They will have the right to resell or publish your data (anonymously or with information that could identify you, at their discretion)." They found that people who were concerned about their privacy (based on a pre-measurement of their attitudes using the Westin Privacy Segmentation index [131]) do not necessarily give answers indicative of their attitude. Harbach *et al* [54] present a field study and online survey on user phone locking behaviour. In their online survey, they used scenarios such as, "Please rate how serious you find the following: [...] someone being able to access my data when my phone is unattended." One finding was that some of their participants often use physical means of protecting their phones instead of locking it. One reason for leaving their phone unlocked is to make it easier for someone who finds their phone to access their contacts to notify them of the lost phone.

2.4.2 Online Surveys

Online surveys are an established method of studying human attitudes and administering stimuli. They are helpful in accessing larger, more diverse participant pools. Buhrmester *et al.* [19] find that Amazon's online micro-task platform, Mechanical Turk, is a good source of high-quality data. It is widely used for online data collection in usable security studies [23,31,63,105,133]. Similarly, researchers use Qualtrics, an online survey software to recruit [130] and administer surveys [3,7]. Online panels of potential participants, such as those provided by Qualtrics, are more representative of the general population than social media and crowd-sourced samples [95]. Most panel members, however, choose not to respond to online surveys and this may lead to sampling bias [95].

2.4.3 Assessing Visualizations

In an online survey, Bravo-Lillo *et al.* [18] investigated the effect of *inhibitive attractors*: interactive, visual, or temporal features designed to slow the user down from

clicking through a dialog box mindlessly by drawing their attention to salient information. The aim was to determine at what level participants were willing to ignore critical security-relevant information within installation and permission-request dialogs. Their participants answered comprehension questions about the security-critical content that they saw in the dialog boxes. Bravo-Lillo *et al.* found that participants correctly answered more comprehension questions in the inhibitive attractor conditions than within the control condition. We draw on this study for inspiration for our methodology. We use comprehension questions in our online survey to evaluate our experimental activity log which draws attention to salient information against our control which does not.

2.5 Summary

Although existing literature discusses responsibility for account security, end users' perception of it is unclear. We address this research gap by exploring how users allocate responsibility between themselves and their service providers. Existing literature provides models of trust in an online context. We add to this research area by exploring the use of trust cues by participants for their service providers. We do not know enough about end users' practices with respect to account monitoring. We address this research gap by gauging participants' beliefs and existing practices. Similarly, we do not know enough about the usability of account monitoring tools for end users. We address this research gap by designing, prototyping, and testing account monitoring tools.

Chapter 3

Study 1: Account Monitoring App

3.1 Introduction

In Study 1, we explored the concept of an account activity app, *Account Sentinel*, (Figure 3.1) that would monitor all of a user's main accounts and display the account activity within one interface. To this end, we designed a medium-fidelity prototype that generated an interactive visualization of the user's activity across 14 different online accounts during a one-week time period. This activity log is depicted in Figure 3.2. We conducted a user study to collect data about participants' perceptions of this visual interactive activity log, their perceptions of the app itself, and their attitudes toward account monitoring.

Since we could not readily collect data about the participants' own account activity without invading their privacy, we asked them to imagine themselves as a hypothetical user named Jane Doe while they used the prototype. Jane Doe's hypothetical accounts were mainly based on the Top 50 domains accessed in Canada [2]. We included a Carleton account for Jane Doe to facilitate empathy with her and, consequently, the role-playing mindset that we asked from our participants.

3.1.1 Research Questions

The following research questions formed the basis of our motivation for Study 1:

- **RQ1.** Will users understand the combined account activity logs from the proposed visualization?
- **RQ2.** Will users discover the unusual events in the proposed visualization?
- **RQ3.** Are users likely to adopt the (prototyped) account monitoring app?

Our hypotheses were:

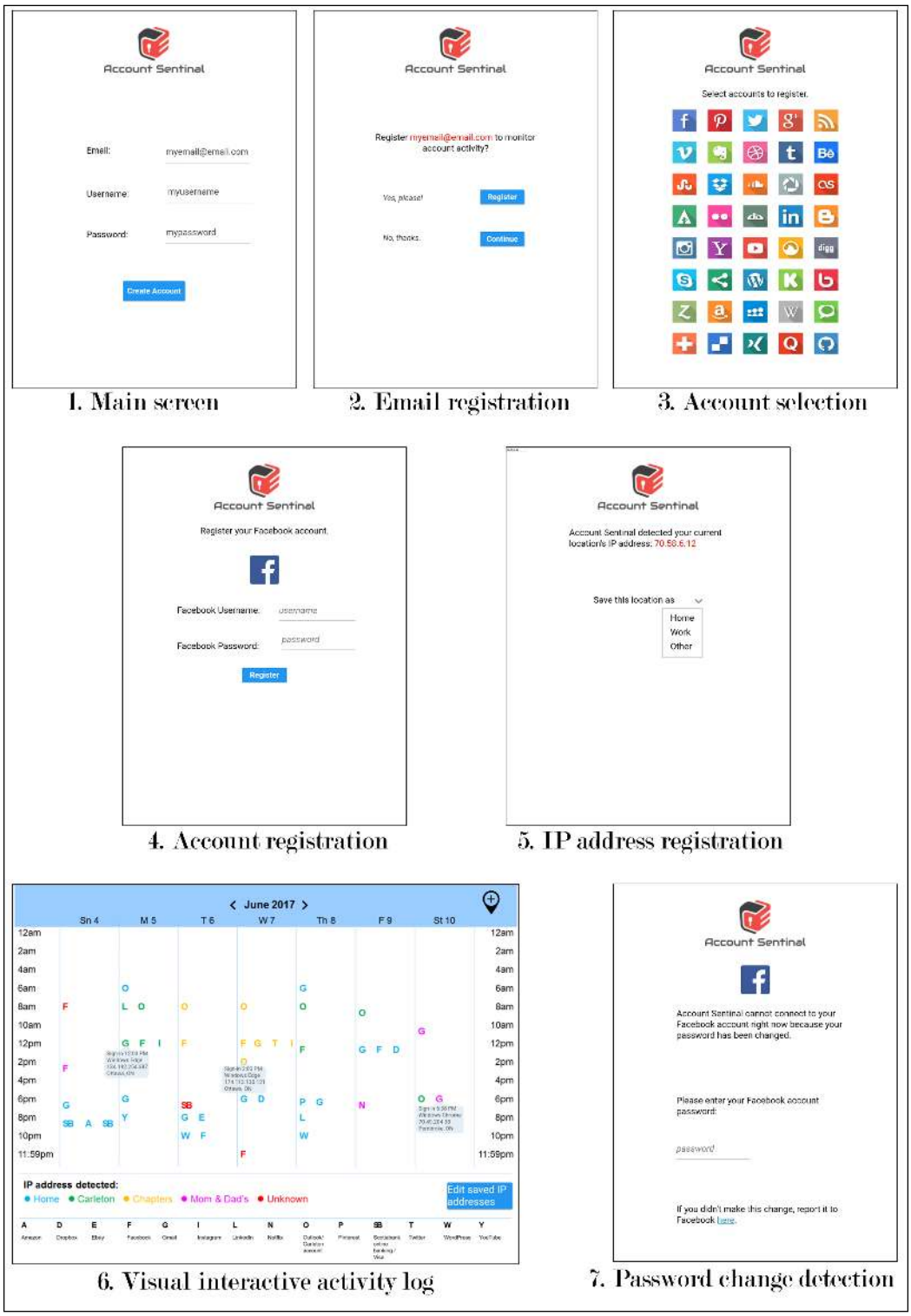


Figure 3.1: Prototype screens

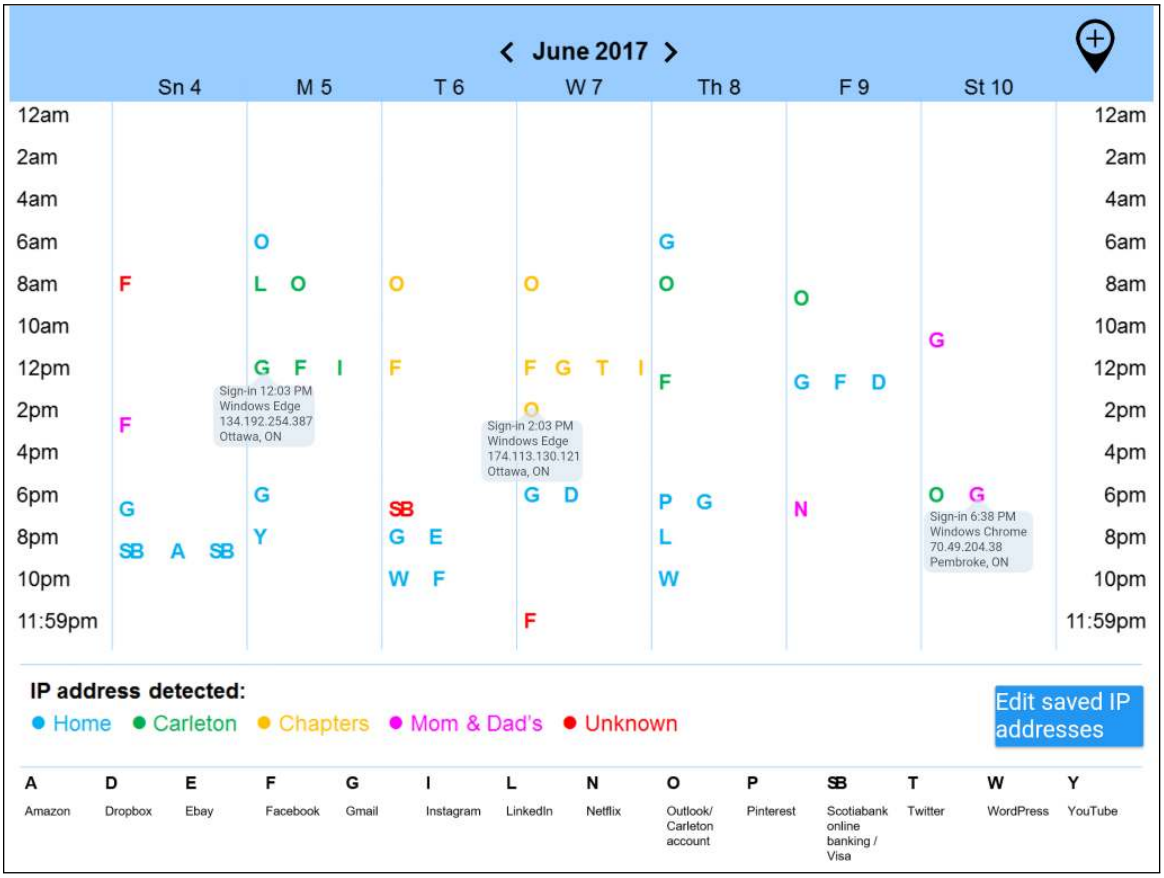


Figure 3.2: Screen 6: Interactive visual account activity log depicting a hypothetical user's activity over a one-week time period.

- **H1.** The visual account activity log will be understood by participants. This was evaluated by analyzing the participants' answers to comprehension questions regarding the visualization.

- **H2.** The unusual events will be interpreted correctly. We measured this by noting which events were characterized as unusual by participants and why.

- **H3.** The app will be perceived favourably by the participants and they would want to use it for the monitoring of their own accounts. This hypothesis was measured by aggregating the participants' responses to the corresponding questions in the interview.

3.1.2 Terms Used

We use "the app," "our app," "account monitoring app," and "*Account Sentinel*" to refer to our prototype.

3.2 Methodology

We recruited 15 participants through word of mouth, advertising, and the researcher's circle of acquaintances. Participants were asked to answer questionnaires, test a prototype, and participate in a semi-structured interview. Testing was conducted through one-on-one sessions with the experimenter. We reimbursed them $10 for their time. This study was reviewed and cleared by the Carleton University Research Ethics Board.

We collected qualitative and quantitative data through pre-test and post-test questionnaires administered through Qualtrics, usability testing of the prototype, and a semi-structured interview. The researcher took notes of the participants' interaction with the prototype and audio-recorded the semi-structured interviews.

3.2.1 Session

The sessions took place in Carleton's Human-Oriented Research in Usable Security (CHORUS) lab or off-site at a mutually convenient location of the participant's choice.

Sessions lasted between 20 and 60 minutes, depending on how much time each participant took in answering the questionnaires and testing the prototype. Participants read and signed a consent form and proceeded to participate in the following:

1. **Demographic and pre-test questionnaire:** This questionnaire was administered using Qualtrics (a survey software) on a provided laptop. Besides basic demographic questions, the demographic questionnaire included a question about whether participants have colour blindness. All participants indicated that they do not have colour blindness. The pre-test questionnaire included questions to gauge the participants' perceived importance of their accounts, how concerned they are about the security of those accounts, and whether they check the activity logs of those accounts. See Appendix A for the complete questionnaires.

2. **Usability test of prototype and semi-structured interview:** Participants used the prototype on an 8-inch touch screen tablet computer (Samsung Galaxy Tab A) and were asked to proceed through the screens as if they were the hypothetical user Jane Doe (Figure 3.1). When they reached the activity log screen (Figure 3.2), the researcher read aloud the hypothetical profile of Jane Doe (see Section 3.2.3) to illustrate where and when she accesses her 14 accounts. Participants were given a printout of Jane Doe's profile to refer to as they were looking at the activity log screen. In line with our second research question (in Section 3.1.1), the researcher asked the participants whether they spotted any events that appear unusual for Jane Doe. Each participant was given as much time as they needed to look at the visualization and answer the question. The last screen in the prototype was the password change detection screen (Section 3.4.3), included to gauge the participants' reactions to it.

3. **Post-test questionnaire:** This questionnaire was administered in the same way as the pre-test questionnaire. Participants answered questions about whether they plan to check the activity logs of their accounts in the future, whether they preferred checking their account activity using a visual activity log or existing textual logs, and how easy it was to understand the visual activity log presented

to them.

3.2.2 Design Process

Our design process for the visualization started by experimenting with Google Chart tools[1] and D3[2], a JavaScript library for visualizing data with HTML, SVG, and CSS. We did not find existing templates to depict activity log data in a way that represented multiple variables at once. We decided to design our own visualization using a calendar layout for two reasons: (1) the calendar layout is ubiquitous in both digital and print media and is therefore likely to be understood by end users, and (2) the calendar layout can visualize time. We iterated on the design of the calendar visualization and the rest of the app screens through discussions within the research team. We generated a medium-fidelity app for user-testing.

3.2.3 Materials

Hypothetical User Profile: The following is the hypothetical user profile we read to the participants. A paper printout of this user profile was also made available for participants to refer to as they looked at the visual activity log.

Jane Doe is a full-time student at Carleton University. She is on campus three days a week on Mondays, Thursdays and Fridays. On Tuesdays and Wednesdays she works at Chapters from 9am to 5pm. Every Friday night she drives up north to Pembroke to spend the weekend with her parents. Jane is constantly checking her Gmail and Carleton accounts on her iPhone. On Friday nights, she usually watches a movie on her Netflix account with her parents. This particular week, she went out for breakfast with her parents on Sunday morning. Jane has an Ebay account that she uses occasionally. She is currently looking for a personalized gift for her best friend's 21st birthday, but has not purchased it yet. She prefers Amazon for domestic purchases. This week, she purchased a set of headphones and a case for her tablet in one transaction on Amazon using her Scotiabank Visa credit card. Jane is also an avid blogger; she regularly posts to her Wordpress blog before going to bed at 11pm.

[1] https://developers.google.com/chart/
[2] https://d3js.org/

She is not a big fan of social media though, and usually accesses her Facebook, Twitter, and Instagram accounts as a way of passing time and keeping up with her friends. Jane created an account on Pinterest to post pictures of her hobbies and creative pastimes. She has a YouTube account that she uses occasionally. Jane also has a Dropbox account she uses to access pictures that her mother shares with her. Jane uses her LinkedIn account to stay up-to-date on potential job openings that would be more relevant to her after graduation.

Visualization: Multivariate Time-Series Data Using Proto.io[3], a prototyping software, we created an interactive prototype consisting of 7 screens depicted in Figure 3.1. The prototype included an interactive *Calendar visualization* depicting the account activity logs of 14 accounts belonging to a hypothetical user named Jane Doe. This activity log spanned a time period of one week, as illustrated in Figure 3.2.

The visualization uses a calendar layout; the vertical axis is a continuous scale for the time of day and the horizontal axis is an ordinal scale for the day of the week. The fictional dataset used for this visualization consisted of ten variables. We mapped four variables to the visualization:

1. a glyph identified the account domain, e.g., "F" = Facebook
2. the vertical placement signalled the hour of day
3. the horizontal placement indicated the day of week
4. colour identified the source IP address

The remaining variables were not visually depicted, but appeared in text form within pop-overs. A pop-over is a message that appears when the user clicks an element such as an icon, image or hyperlink. Unlike tooltips, pop-overs stay open until they are clicked or tapped again. The following variables were contained in the pop-overs that appeared when participants tapped each glyph:

5. The exact time of the event, e.g., 12:03 PM

[3]https://proto.io/

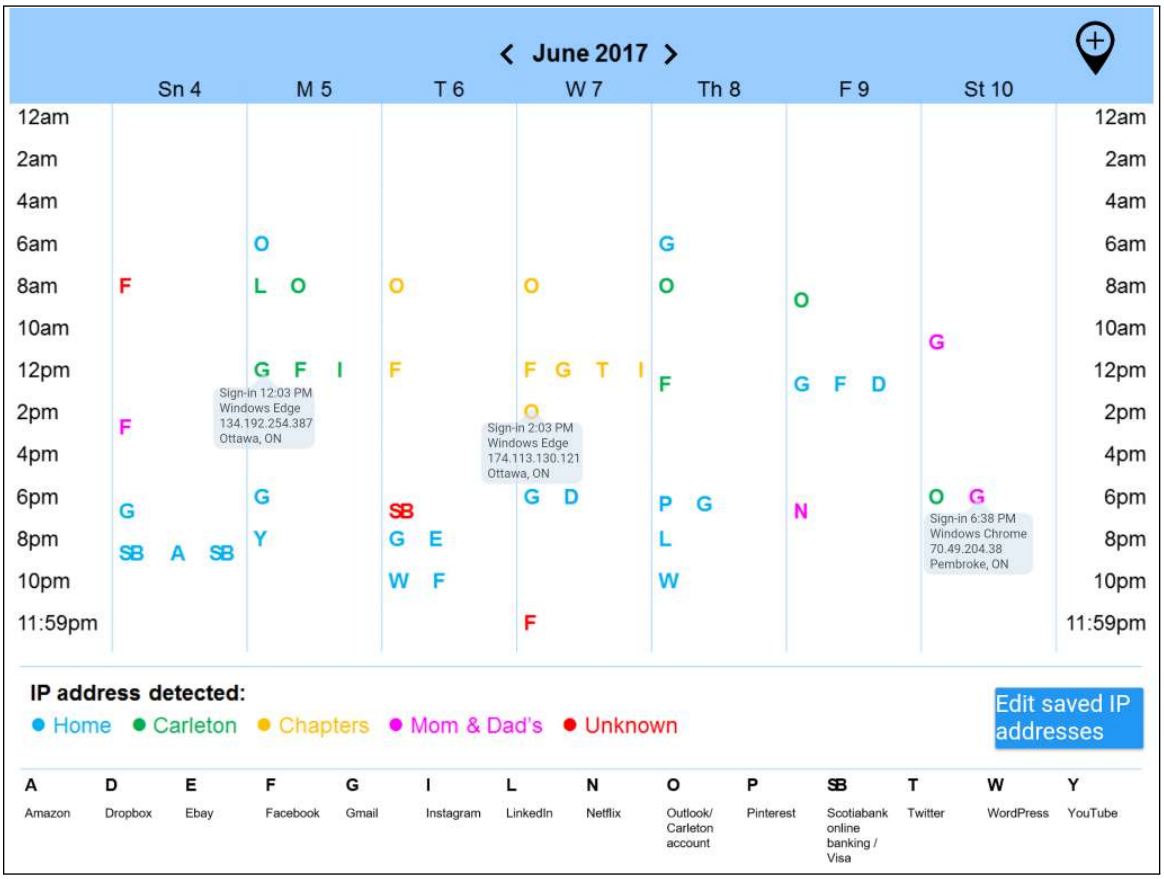


Figure 3.3: Four variables were mapped to the visualization and five variables were available in text form within pop-overs that appeared when participants tapped the glyphs.

6. Event type, e.g., sign-in, attempted sign-in, purchase
7. Operating system used to access the account, e.g., Windows
8. Browser used, e.g., Chrome
9. Source IP address, e.g., 174.113.130.121
10. City (geographical location of access), e.g., Ottawa

In keeping with design principles and Gestalt laws, we used saturated colours on a white background to maintain contrast [129]. We also used horizontal proximity [129] to indicate grouping by hour of day. This is because we wanted participants to be able to infer the hourly usage patterns across the week.

Interactivity and Tapping: Medium-fidelity interactivity was enabled within the prototype. Each glyph was tappable, and a pop-over appeared with additional information for each event. The pop-over stayed open until the participant tapped the same glyph again. Multiple pop-overs could be opened at once to enable participants to compare events.

Activity Log Events: We visualized a total of 45 events (Figure 3.2) depicting access to Jane's accounts: 43 sign-ins to her Facebook, Gmail, Scotiabank, Amazon, Outlook, LinkedIn, Instagram, YouTube, Ebay, Wordpress, Twitter, Dropbox, Pinterest, and Netflix accounts, one debit from her bank account, and one purchase from her Amazon account. We designed four of these events to appear unusual or suspicious, as if they represented unauthorized access to Jane's accounts. For example, in Figure 3.2, we see a red F at 8am on Sunday, representing a Facebook login from an unknown location at that time.

The four unusual events we injected into the activity log are summarized in Table 3.1 and highlighted in Figure 3.5.

3.2.4 Participants

We recruited 15 participants; 8 female and 7 male. Seven of the participants were undergraduate students, one was a graduate student, one was a post-doctoral fellow, three were stay-at-home parents, and three were full-time employees as an administrative assistant, elementary school teacher, and software developer. Eight of the 15 participants indicated that they have worked in a computer, computer security, or information technology-related field. The participants' educational backgrounds were composed of computer science (6) electrical engineering (1), biomedical engineering (1), anthropology (1), and psychology (1). Five participants did not indicate their educational backgrounds.

Table 3.1: The four unusual events. The classification was not revealed to the participants, but only mentioned here for clarity.

Event	**Day**	**IP address label (location)**	**Classification**
Event 1 F (in red)	Sun Jun 4 9:28 AM	Unknown (Pembroke, ON)	Non-malicious: as indicated in her profile, Jane went out for breakfast with her parents in Pembroke on Sunday morning. The location from which she accessed her Facebook account was unknown to the app at the time.
Event 2 SB (in red)	Tues Jun 6 6:42 PM	Unknown (Ottawa, ON)	Malicious: Janes credit card was stolen and used for a purchase. Her profile indicates that she made only one purchase that week.
Event 3 F (in red)	Wed Jun 7 11:57 PM	Unknown (London, UK)	Malicious: Someone has compromised Jane's Facebook account. This access occurred from London, UK, while Jane's profile indicates she was in Ottawa the whole week.
Event 4 O (in green)	Sat Jun 10 6:22 PM	Carleton (Ottawa, ON)	Potentially malicious: Unauthorized access to Jane's account. Her profile indicates that she was not at Carleton University, but still in Pembroke for the weekend. This is apparent by the magenta-coloured event depicting access to her Gmail account from her parents' house, within the same time frame.

3.3 Data Analysis Process

3.3.1 Qualitative Analysis

The audio recordings of the semi-structured interviews were selectively transcribed. Using the qualitative analysis software AtlasTi [10], we conducted a thematic analysis [20] on the interview transcripts. The researcher started by reading the provided answers several times to understand the data as a collective body. She then derived codes directly from the words that conveyed key concepts. Depending on the level of detail that participants provided, the researcher grouped codes into categories, and merged the ones with the same underlying concepts. This was done until a code applied to the selected quotations. Table 3.2 lists examples of data extracts and the codes applied to them. We analyzed the results of the questionnaires by aggregating them and comparing the pre-test and post-test answers.

For brevity, the high-level codes associated with our research questions are listed here. The remaining codes are listed in Appendix B.

1. Will users understand the consolidated account activity logs in the form of the proposed visualization?

    - Information used for deciding suspiciousness of events (9 sub-codes): The cues and indications that participants used to decide whether or not an event was unusual or suspicious to them.

2. Will users discover the unusual events in the proposed visualization?

    - Events missed (4 sub-codes): The unusual events that participants misattributed as benign and any reasons they mentioned for their misattributions.
    - Events spotted (18 sub-codes): The unusual events that participants attributed as such and their reasons for doing so.

3. Are users likely to adopt the (prototyped) account monitoring app?

    - Drawbacks of the app ("cons") (14 sub-codes): The functions and features,

Table 3.2: Sample data extracts and the codes applied to them.

Data extract	**Codes applied**
"If a lot of people use [the app], then I'll be like, 'Oh, it seems like it's fine,' but if I'm the only one or like I know I'm the first few I would definitely not put my bank account on this."	• Decision to trust an app: used by many people • Decision to trust an app: been around for a long time • Perceived account value: high - bank
"Of course this app is transferring data back and forth from all accounts [...] so the consolidation is happening on a database server which is not my phone. If whoever created the app today [...] ran away with everybody's bank accounts [...] because they have that information on the database there [...] no matter how much encryption, they are the ones who encrypted it, they can decrypt it. So there is a huge security risk that can occur." -P2	• Decision to mistrust app: external server • Drawback of app: threat of compromise from its own developers
"If there was a lightweight app that [...] didn't have any cons or significant setup time, [...] I could say yes access my account information and just do it automatically [...] and pop up notifications with the right balance, I'd definitely download that and use that. I wouldn't search it out because I don't have problems with security, but for financial [accounts] I have a higher standard..." -P8	• Desired attribute of app: lightweight • Desired attribute of app: short setup time • Desired function: alerts • Desired function: customization • Perceived account value: high - bank
"I will trust Google, Netflix, Dropbox, Amazon, most of these guys, I would trust them to send me an email of my anomalous logs more than I would trust this app." -P12	• User attitude: burden of security is upon the service providers • Decision to trust service providers because of their reputation

or lack thereof, that participants believed made the app a poor candidate for adoption.

- Desired changes to the app (33 sub-codes): What participants indicated would make the app a better candidate for use, or increase their likelihood of adoption.

A critical theme emerged from our data and we explored it in Study 2 (See Section 4.4.3):

1. Trust (35 sub-codes): The reasons participants gave for whether or not they would trust *Account Sentinel* with monitoring their accounts, and the attitudes around trusting mobile apps/online service providers in general.

3.3.2 Quantitative Analysis

During the usability test of the app, we asked participants if they spotted any unusual events for Jane. We counted how many events were discovered, and how many were correctly interpreted. The results of this stimulus test were aggregated and compared between participants.

3.4 Results

In our graphs, we use red and green for negative and positive results or responses, grey and orange for pre-test and post test results, and blue for reporting frequencies and percentages.

3.4.1 Pre-Test Questionnaire

The results of the pre-test questionnaire establish a baseline of the participants' understanding and behaviours with regards to their account activity logs. The top accounts that the participants want to protect are their bank accounts (8 of 15 participants), their email accounts (5), and their Carleton University accounts (2). The main reasons participants cited are to prevent financial, identity, and information asset theft, and to prevent privacy threats.

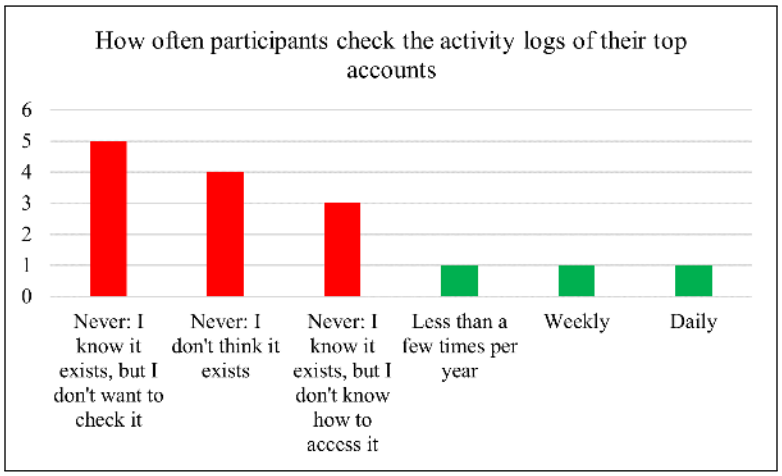


Figure 3.4: The pre-test responses to how often participants check the activity logs of their top accounts.

The top accounts for daily use are Gmail (12 of 15 participants), Facebook (10), WhatsApp(9), Carleton University account (8), YouTube (5), and Netflix (4).

Eleven of 15 participants reported having at least one account compromised or almost compromised. Table 3.3 specifies the accounts and reported contexts of the compromises. This is an indication of the vulnerability of users and serves as a motivation for us to develop a tool to aid them in the maintenance of their account security.

In the pre-test questionnaire, twelve participants reported never checking their account activity logs, and three reported having checked them (see Figure 3.4).

3.4.2 Stimulus Test: Which events did they discover?

The majority of participants identified and correctly interpreted Events 1, 2, and 3 (see Table 3.1). This success rate was reversed for Event 4: only 5 of 15 participants correctly identified it. Table 3.4 details the results of the stimulus test. Figure 3.6 shows a comparison of the events.

Event 4 is depicted as a green O. The colour green symbolizes the IP address of Carleton University, as labelled by Jane. However, Jane is out of town visiting her parents on the Sunday afternoon. Despite the participants having heard the experimenter reading Jane's profile aloud and having the printout to refer to, most of

Table 3.3: Eleven participants' reports of account compromise or attempted compromise.

Participant ID	**Account**	**Context**	**Action(s) Taken**
P1	Hotmail	Unauthorized access of her account when she was 13 years old. Felt scared and vulnerable.	Created new account Limited who she shares email address with
P4	*Not specified*	*Not specified*	Changed password
P5	Email	Activity log revealed login from overseas.	Changed password Limited use of said account Created new account
P6	Email and "all accounts."	*Not specified*	Changed password Retaliated against the attacker
P7	*Not specified*	*Not specified*	Changed password Provided backup email
P8	*Not specified*	After logging in through hotel computer. Suspected the use of a keylogger.	Reset password Created stronger password
P9	Bank account	Unauthorized attempted use of bank card.	Replaced the bank card.
P10	Facebook	Facebook alert that someone overseas was attempting login.	Changed password Logged in again on other devices using the new password
P11	Hotmail	*Not specified*	Changed password
P12	ICQ	1997 ICQ account hacked via a Trojan Horse program mistakenly installed on victim's computer.	Lost account Stopped using service
P15	Facebook	Facebook alert that someone overseas was attempting login.	Changed password

Table 3.4: Stimulus test results showing the number of participants who correctly identified the unusual events.

Event	**Classification**	**Identified event**	**Perception**	**Reasons cited by participants**
Event 1: F (in red)	Not malicious	11 of 15 participants	9 of 11 correctly assumed it was not malicious	Red colour; IP address; Timing
Event 2: SB (in red)	Malicious	13 of 15 participants	12 of 13 correctly assumed it was malicious	Too large a purchase; Does not match the profile because Jane made only one purchase that week; Not accompanied by a sign-in to Amazon; Location; IP address
Event 3: F (in red)	Malicious	15 of 15 participants	15 of 15 correctly assumed it was malicious	Location; Red colour; Timing
Event 4: O (in green)	Potentially malicious	5 of 15 participants	5 of 5 correctly assumed it was potentially malicious	Location (because Jane was at a different location at the time of access, according to her profile.)

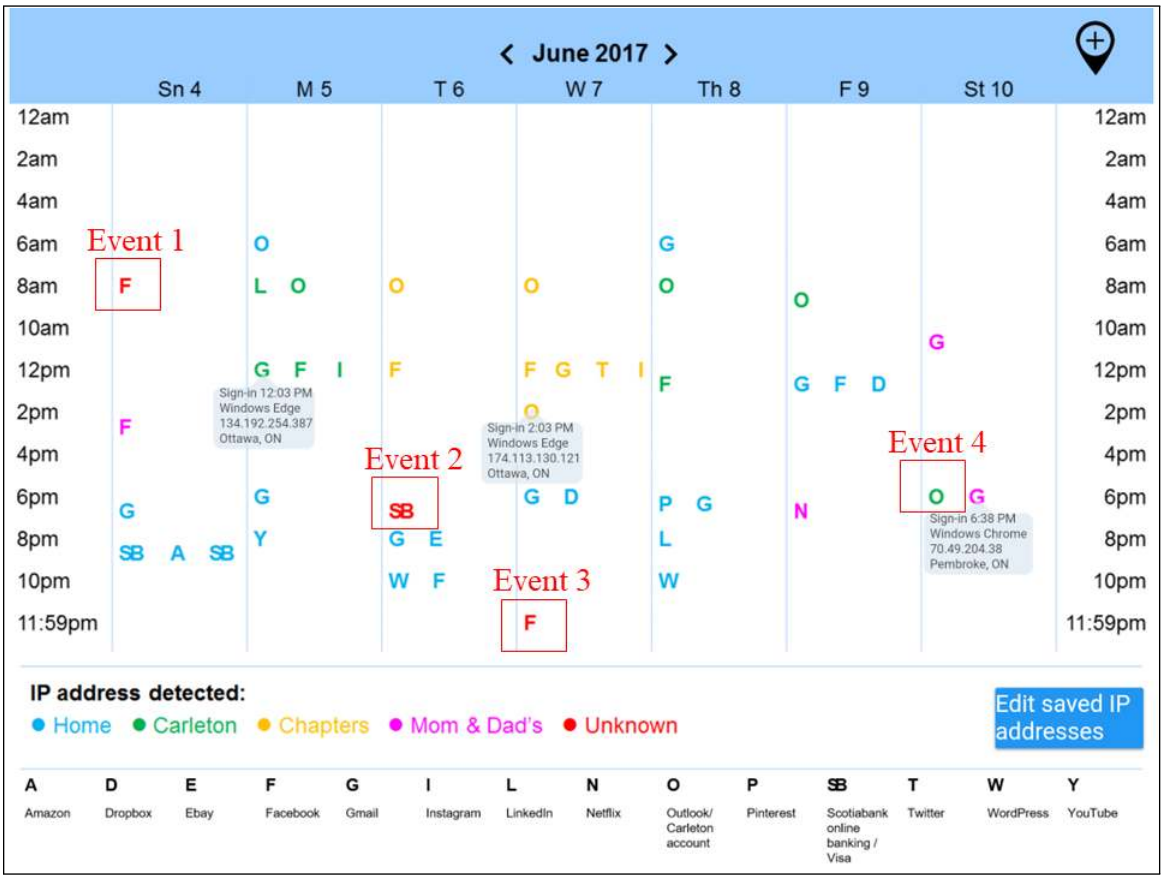


Figure 3.5: The four unusual events in the visual activity log. They were not disclosed in the prototype, but artificially labelled here in red for clarity.

them did not notice the discrepancy between Jane's actual location and the location of the access depicted by Event 4. We postulate that the false negative attributions of Event 4 were due to its colour being green. Similarly, we argue that the false positive attributions of Event 1 were due to its colour being red. This is because people tend to associate safety with green [25] and danger with red [89].

During the interview, we asked, *Which information is more important to you in deciding whether or not a particular event is unusual or suspicious: the device information, the IP address, the location, or something else?* Participants reported using location, IP address, and time of access most often. The question posed was a leading one, however, and therefore we are unsure of what information participants would use outside the context of this study. Table 3.5 aggregates the responses.

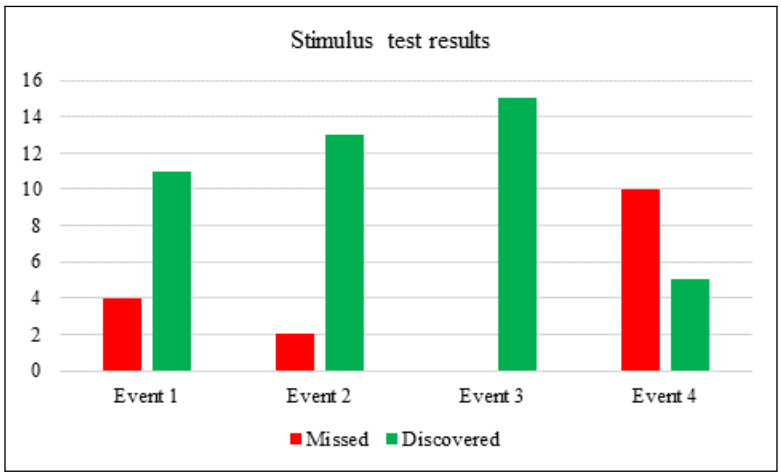


Figure 3.6: A side-by-side comparison of the discovery rate for each event.

Table 3.5: The information participants would use in deciding whether an event is unusual.

Information	**Number of participants**
Location	8
IP address	4
Time of access	3
Time of access + IP address	1
Operating system	1
Account being accessed	1
Street-level location	1
Device	1
Browser	1

Table 3.6: Participants' reactions to the password change detection screen.

Behaviour	**Number of participants out of 15**
Try source login (external to the app)	8
Follow app instructions; report it	7
Try the same password in the app again	2
Change password	2
Check account activity	1
Would not put password in the app	1
Perception	**Number of participants out of 15**
Someone hacked my account	7
Virus or error in the app	5
Syncing problem	1

3.4.3 Usability Test and Interview

Password Change Detection Screen The password change detection screen (Figure 3.7) was included in the prototype as system feedback in the event that the app could not access the fictional user's Facebook account. This scenario would occur if the user had changed their Facebook password and did not update the password in Account Sentinel. The participants' perceptions of the password change detection screen varied. Table 3.6 lists the participants' answers in response to the question, "What would you do if you saw this? What does it mean?" There is some overlap in perception across participants, and not all participants provided their perception. Interestingly, half of the participants took this to mean that their Facebook account would have been hacked, yet a third of the participants perceived this as the app itself having been compromised or flawed. This points to the sense of unease that most participants had about the app.

Usability and Perceptions of the App The top features of the app that participants found to be useful are the colour coding of the visual activity log (8 participants), the red colour within the visual activity log for what they sometimes mistakenly perceived to be suspicious events (4), and the availability of the source IP address or location of each event within the pop-overs (5).

The app was perceived as being easy to use by 11 of 15 participants. P9 explains: "It was a nice [...] quick graphical representation. I was expecting more of a list of things which I wouldn't want to go through." Five participants found the colours

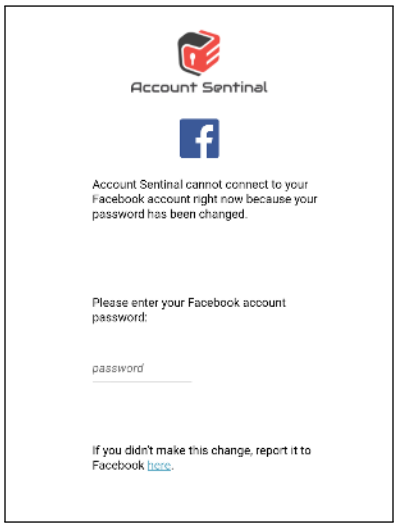


Figure 3.7: Password change detection screen.

confusing, and three participants felt that the activity log was confusing overall. P12 found the activity log to be unintuitive. Two participants felt that the activity log is cluttered or has the potential to be cluttered with more activity.

Despite the perceived ease of use, the likelihood of adoption was very low; only three participants indicated they would use the app if they were sure of its security (which they were not). The main reason behind a lack of adoption was the mistrust of an app that can access all of their accounts. Figure 3.8 shows the potential advantages and drawbacks that participants cited.

Improvements Suggested by Participants Participants suggested a variety of improvements. The consolidated responses are listed below by topic.

Security and Privacy:

- Facilitate prevention of attacks through blacklists and user education
- Enable password management of monitored accounts
- Automatically identify and only show unusual or suspicious events
- Include a clear privacy policy

Visualization:

- More clearly demarcate the hours of the day

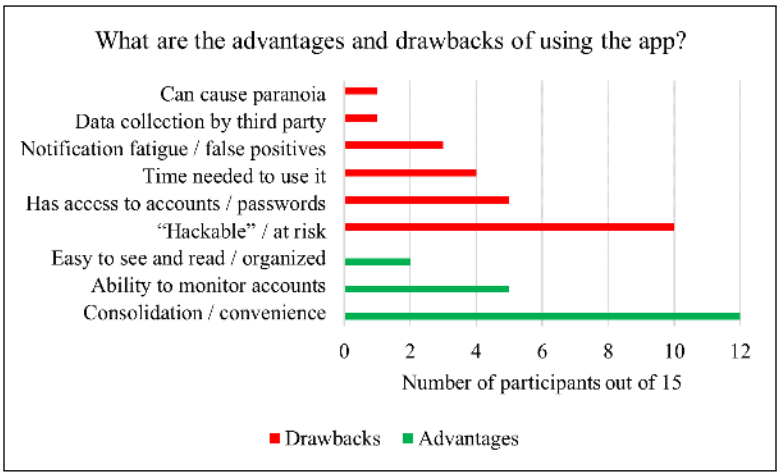


Figure 3.8: Advantages and drawbacks of using the app as cited by participants. Multiple responses were allowed per participant.

- Use icons instead of letters to represent the accounts
- Use shape coding instead of colour for location

Interaction:

- Allow filtering of events by particular criteria such as time, date, and event type
- Include customized alerts or notifications for activity on accounts chosen by the user
- Allow customization of visual encoding
- Enable zooming

Security Practices Pertaining to Activity Logs Prior to the study, most participants reported never checking activity logs. In the post-test questionnaire, they were asked how often they intend on doing so in the future. Figure 3.9 shows the pre-test and post-test responses side-by-side. It is interesting that the number of participants who never check the activity logs of their accounts (12) is reversed for those who do intend on checking them in the future. This is possibly due to the participants being primed by the topic of account security.

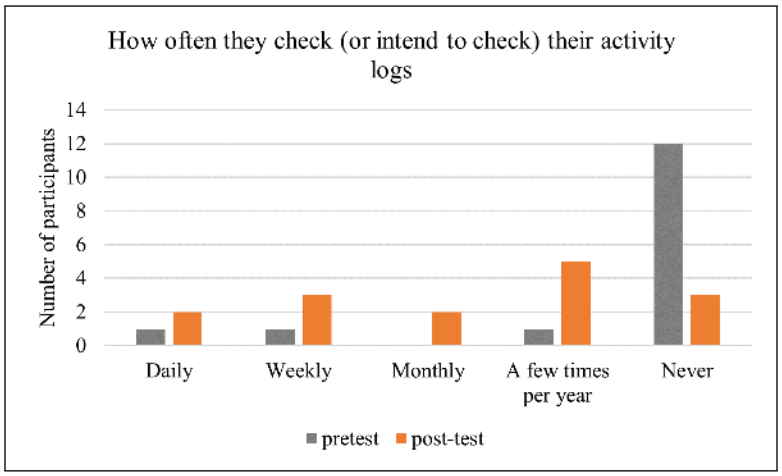


Figure 3.9: The answers to, "How often do you check (pre-test)/intend on checking (post-test) the activity logs of your top accounts?"

Trust

A critical theme that emerged from the interview data was trust. As mentioned previously, the main reason that participants cited for lack of adoption was their mistrust of the app either because it was collecting all of their account passwords and data, or because of the high stakes involved should it become compromised. Fourteen of 15 participants indicated that they would not trust *Account Sentinel* enough to use it, and one participant was unsure.

P3 and P11 felt that it was risky to put all their passwords in one app. P5 was under the impression that the account security app would share her metadata, putting her at risk. P10 said, "it's a little scary, one account that can see all of my activities, so I feel like it's a place for somebody who would want to see what's going on in my life, they could just access this account and they would know what's happening."

One participant was concerned about the risk of compromise from the app's own developers: "Of course this app is transferring data back and forth from all accounts [...] so the consolidation is happening on a database server which is not my phone. If whoever created the app today [...] ran away with everybody's bank accounts [...] because they have that information on the database there [...] no matter how much encryption, they are the ones who encrypted it, they can decrypt it. So there is a huge security risk that can occur (P2)."

P12 pointed out that the same APIs that connect a users' accounts with *Account Sentinel* would have to decrypt those account passwords, revealing them in plain text and storing them. He also believed that such single sign-on schemes are more vulnerable to attacks than separate sign-on for each account. Furthermore, the server used by *Account Sentinel* would be a third party attempting access to a user's accounts, thereby potentially causing denial of service.

P7 said that Account Sentinel's "design is bland," and that he would not trust apps that look like ours enough to download them. P9, P11, and P15 also indicated that they do not like the visual appearance of our app. In retrospect, using a low-fidelity paper prototype would have been more effective for extracting users' perspectives on the monitoring tool itself rather than external or design factors.

Trust Cues: When asked how the app can win their trust, participants gave the following *trust cues*. Several participants cited more than one way the app can win their trust:

1. If it is highly reviewed in positive online ratings or through word-of-mouth (P6, P7, P8, P9, P10, P15)
2. If it comes from a reputable organization or company or can be trusted (P7, P8, P9, P11, P12, P15)
3. If it is used by many people (P1, P2, P6, P4, P11)
4. If it has robust security architecture (P2, P4, P5, P12, P14)
5. If it is endorsed by experts (P3, P11)
6. If it has been in existence for a long time (P1)
7. If it ensures privacy (P14)
8. If it is a native app (P3) or partner with my bank (P13)
9. If it was a password manager (P3)

The semi-structured format of the interview allowed for follow-up questions about how they trust the apps currently installed on their mobile devices. P2 said that she

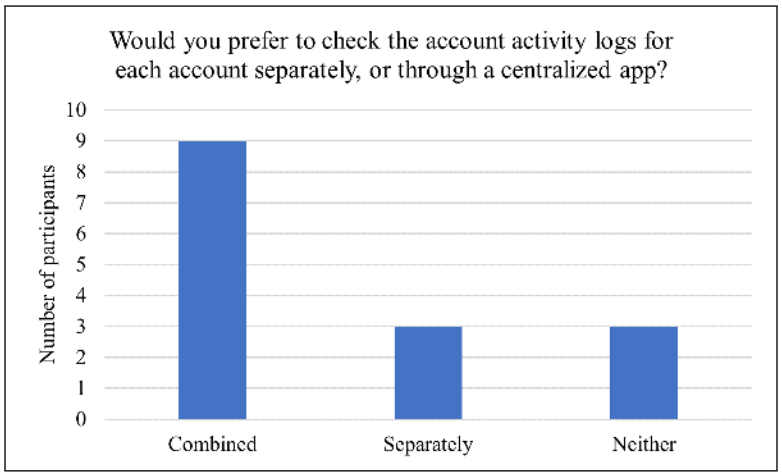


Figure 3.10: Participants' preferences with regards to how they would like their account activity logs to be presented.

trusts her apps if they are not third-party. Interestingly, P3 uses a third-party app, Mint, that tracks her financial activity by connecting to her bank account and trusts the app because it does not have her bank password. P4 said that his apps do not store his passwords, and P8 trusts his currently installed apps because of word-of-mouth recommendations. P12 trusts his apps, including a password manager, because of their policies, reputation, and security architecture.

3.4.4 Post-Test Questionnaire

Participants were asked a series of questions in the post-test questionnaire to gauge their perceptions of the app and preferences around account activity logs. Most participants (nine) indicated that they preferred checking their activity logs in a combined format. The main reason they cited was convenience. Participants who did not want to check their activity logs neither separately nor in a combined format indicated that it is the online service provider's responsibility to alert them in case of unusual activity. One participant wanted to check her accounts separately because she deemed that using an app is not safe. Another participant wanted to check her accounts separately simply because she did not want to download another app. Figure 3.10 aggregates the responses.

Overall, participants rated the app favourably in how easy and how fast it was

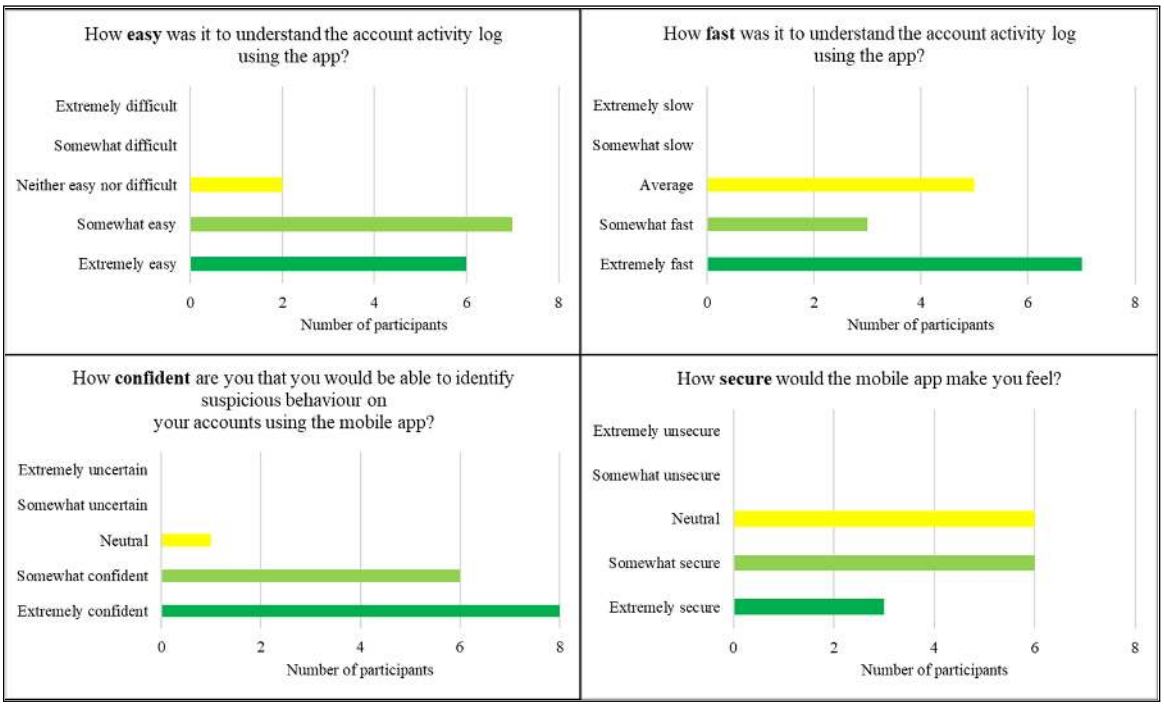


Figure 3.11: Participant responses on a 5-point Likert scale to questions about the app and activity log.

for them to understand the activity log, as well as how confident it would make them feel in monitoring their accounts. These rating are reversed for how secure the app would make them feel. Figure 3.11 shows the participant responses.

3.5 Discussion

Our first research question was, ***RQ1.*** *Will users understand the combined account activity logs from the proposed visualization?* We hypothesized that the visual account activity log will be understood by participants. Our findings partially confirm our H1 hypothesis.

Our second research question was, ***RQ2.*** *Will users discover the unusual events in the proposed visualization?* We hypothesized that the unusual events will be interpreted correctly. Our results partially confirm our H2 hypothesis.

Our third research question was, ***RQ3.*** *Are users likely to adopt the (prototyped) account monitoring app?* We hypothesized that the app will be perceived favourably by the participants and they would want to use it for the monitoring of their own

accounts. Our findings oppose our hypothesis and show that users had reservations about adopting the app. We, therefore, reject H3.

In Study 1, we expected to test the visual activity log. However, external factors surrounding the visualization played a bigger role. Those external factors were the cause for *H3* being rejected. Participants would not trust our app enough to use it although they liked the idea of being able to monitor all their accounts in one place, in a combined format. Participants identified *trust cues* they would look for in a new app, or that they use for the apps and online service providers to which they currently subscribe. One trust cue that led to negative perception was the visual appearance and design of our app. The design of our app was also a cause for some usability issues. The lesson we learned from this was that we should have used a low-fidelity or paper prototype because it would have been more effective for gauging users' perceptions of the monitoring tool and the visual activity log. Higher-fidelity prototypes are known to distract users from its main purpose or cause hesitation to give honest feedback out of respect for the effort that went into designing it [98].

Another important lesson we derived was that our design process was weak. We should have conducted iterative design of the app and visualization using pilot sessions before testing with participants. This would have reduced the usability issues that arised.

Some participants not only rejected *Account Sentinel* because of mistrust, they also did not believe that they should be monitoring their accounts because it was their service provider's responsibility to do so. The trust and responsibility surrounding account monitoring, and by extension, account security, became necessary for us to investigate. We therefore expanded our research direction in Study 2 to explore the attribution of trust and responsibility in account security.

The role of *trust cues* that appeared in our study confirms existing literature on trust. Previous research has found that people use *transitive trust* [61], *reputation* [8], *personal experience* [61], and *visual appearance or design factors* [107] [113] in deciding to trust their service providers. Because these trust cues were absent for *Account Sentinel*, our participants would not adopt it in real life. We take this research direction further in Study 2 by exploring which trust cues are used for two service

providers in particular, Facebook and Google.

Existing literature on responsibility captures an ongoing discussion debating where responsibility lies with respect to account security [27] [108] [26] [83]. This discussion is out of scope for this thesis, however, we became interested in how end users perceive such responsibility. Previous research [104] found that most participants either attributed responsibility to themselves, or shared it with the service provider for preventing account compromises. We take this research direction further in Study 2 by exploring responsibility for different types of user-facing attacks.

Finally, we address the usability issues uncovered in Study 1 by implementing design improvements to a second iteration of our visual activity log. Due to the limitations of using a mobile screen, we redesigned our visual activity log for desktop or laptop screens. We test this new activity log, *Diagram*, in Study 2.

3.6 Limitations

We discuss limitations of the prototype itself and of the methodology for our user study.

One usability issue that arose was that participants were not used to the behaviour of pop-overs when they expanded them by tapping individual events. This is because pop-overs are not meant to close in the same manner as tooltips, which close upon hovering away from the active interface element, or in the case of mobile screens, by tapping anywhere else. We purposefully designed the app this way to enable simultaneous comparisons of multiple events, but this confused participants.

One participant faced difficulty in tapping glyphs. She indicated that this was a problem she faced with other touchscreens as well: "My fingerprints are somehow not read well by screens" (P2). This may be a result of the target tapping area of the glyphs being too small in comparison to the size of her fingertips, or the sensitivity of the touchscreen being too limited. This participant resorted to using the tablet computer's native stylus for the remainder of the session.

A methodological limitation of our study is the lack of ecological validity. Participants were role-playing a fictional character and linking her profile to the activity log in order to make conclusions about unusual activity. The fictional dataset may

not be representative of real life activity because some users may have considerably more activity. In addition, role-playing may have had an effect on the participants' motivation to inspect the activity log closely [100]. The lab setting is also likely to have affected how participants interacted with the prototype [111].

In retrospect, pilot testing with users earlier in the design process would have been helpful. Furthermore, we presented a medium-fidelity app to our participants when we should have presented a lower-fidelity or paper prototype instead. This is because higher-fidelity prototypes are known to distract users from its main purpose or cause hesitation to give honest feedback out of respect for the effort that went into designing it. For future studies, researchers wishing to test a prototype or visualization should follow the interaction design process established in HCI literature [98].

Chapter 4

Study 2: Online Survey

4.1 Introduction

Our second study was an online survey and it built upon our first study in several ways. In our first study, the top accounts that our participants reported to using daily were Facebook and Google. We recruited Facebook and Google users for our survey due to the popularity of those two platforms. We divided them into two groups to compare any effects that the platform has on the participants' answers. In our first study, we also learned about existing security attitudes and practices with respect to account security and monitoring. We asked the same questions in our survey to gauge how these attitudes and practices extend to a larger sample.

A key finding from our first study is that the users who do not want to monitor their accounts believe it should be their service provider's responsibility to do so. We explore the concept of responsibility on a more nuanced level in our survey. Trust of the service provider also emerged as a critical theme in our first study. In our survey, we asked participants what makes their service providers, Facebook and Google, trustworthy.

We learned from our first study that a third-party account monitoring application is very unlikely to be adopted by users due to it not being trusted by the users. In this study, we shifted focus from the concept of a third-party application to the idea of the combined activity log itself. As in our first study, we allowed participants to take as much time as they wanted to look at the activity log. The objective was not to test speed, but accuracy and comprehension. We introduced a second condition in our survey to compare the performance of a visual activity log against a more traditional text format.

Based on these findings, we pursued the following research questions.

4.1.1 Research Questions

We defined the following research questions to guide our study:

- **RQ1.** Who do end users perceive is responsible for (a) preventing attacks, (b) alerting the user or reporting to the service provider of unusual activity, and (c) for recovering the account after an attack? We measured this with 5-point Likert scale questions.

- **RQ2.** Why do users trust (or mistrust) Facebook and Google? We measured this with multiple choice and open-ended questions.

- **RQ3.** How does the visual activity log (*Diagram*) compare with the text activity log (*Textlog*) in (a) how effective it is in identifying unusual activity, and (b) how comprehensible it is? We measured this by (a) using a stimulus test that revealed which events the participants clicked in response to our question, *Which events are unusual?* and by (b) asking three close-ended comprehension questions.

Our hypotheses were the following:

- **H1.** Users will (a) assign responsibility of prevention to Google and Facebook, (b) share the responsibility of alerting/reporting, and (c) assign most of the responsibility to Google and Facebook for recovery.

- **H2.** Users trust both Google and Facebook mainly because they are well-known service providers and because they use secure technology.

- **H3.** The visual activity log will be more (a) comprehensible and (b) more effective in identifying unusual activity than the text activity log.

4.1.2 Terms Used

We refer to the end user, i.e., the participant, as an *entity*. We also refer to the service provider, i.e., Facebook or Google, as an *entity*. We sometimes abbreviate service provider as *SP*.

4.2 Methodology

4.2.1 Pilot Testing

We conducted two iterations of pilot testing on our initial survey. The first iteration of testing was done with five participants. This iteration revealed that the meaning of some questions was confusing, and that some features of the visual activity log needed adjustment, among several other issues. After implementing improvements, we conducted a second pilot test with another five participants which revealed more questions in need of re-wording. After this process, we adjusted the survey questions to be more accessible and the survey length to be shorter.

4.2.2 Study Design

This study was reviewed and cleared by the Carleton University Research Ethics Board. We conducted an online survey through Qualtrics to explore our research questions. We collected the data in June 2018. We used a 2×2 study design: Half of the participants were assigned to the Facebook group and half to Google, based on the screening question, *Which account do you use regularly? -Facebook -Google/Gmail -both Facebook and Google/Gmail*. Participants who used both accounts could be placed in either group. Each group was asked questions relating to their respective platform (Facebook or Google). Within each group, half were randomly assigned to either the *Diagram* or *Textlog* condition. We used a "survey flow" algorithm provided by Qualtrics to evenly distribute participants across the two groups (Facebook and Google) and to randomly distribute them across the two conditions (*Diagram* and *Textlog*). The survey questions are in Appendix C.

To help control for the integrity of the responses, we set a minimum time of 6 minutes and 30 seconds for participants to complete the survey. They were allowed to use as much time as they wanted beyond that. Participants took between 6 minutes, 30 seconds and 34 minutes, 18 seconds to complete the survey, with an average time of 12 minutes, 30 seconds. Once participants progressed through a page of questions, they could not go back to a previous page.

4.2.3 Survey Design

The survey consisted of 49 questions in total, and a stimulus test. It had 24 Likert scales, 9 open-ended, and 16 close-ended questions. The survey is included in Appendix C. The survey could only be answered on desktop or laptop. Participants who accessed the survey link using a mobile device were told that they did not qualify for it. We set this requirement due to the large size of the activity logs. The survey consisted of the following components:

1. Screening question
2. Demographic questions (3 open-ended and 3 close-ended questions)
3. Security questions (either regarding their Facebook or Google account) about existing security practices and attitudes, opinions on the entities responsible for their account security, how they would like their service provider to monitor their account, the reasons they trust their service provider, and whether they believe that their service provider is able to maintain the security of their data. (34 questions: 22 Likert scales, 4 open-ended, and 9 close-ended.)
4. Activity log (either *Diagram* or *Textlog*) task to select unusual events and corresponding comprehension questions. Participants also answered questions on their perceptions and preferences for the activity log. (Stimulus test and 8 questions: 2 Likert scales, 2 open-ended, and 4 close-ended.)

4.2.4 Activity Log Design

The *Diagram* and *Textlog* conditions explored two different activity log representations. In *Diagram*, we used the same calendar layout in our first study with design improvements that we extracted from usability testing. In our control condition, we tested *Textlog*, a tabular text form of the activity log. It contained the same information as *Diagram*.

Participants were shown an activity log, completed a corresponding task and answered corresponding questions. The task was to identify the events within the activity log that look like they may not be from the owner of the accounts. To this end,

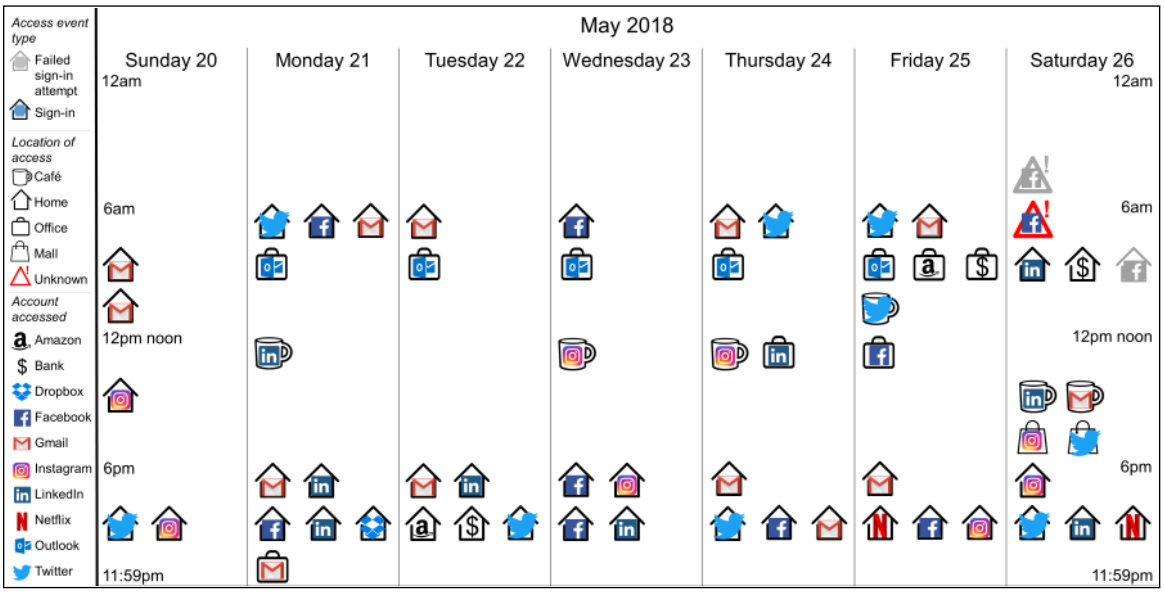


Figure 4.1: The visual activity log presented in the *Diagram* condition.

the activity logs included four events that we designed to appear as unusual. The questions were designed to evaluate their comprehension of the activity log. Both activity logs appeared as static images within the survey, but we placed an invisible hotspot over each event to enable us to identify which events the participants clicked.

The visual activity log, *Diagram*, is originally 4488 x 2250 pixels in size, but was automatically adjusted to the participant's screen. Similarly, the text activity log, *Textlog*, is 1145 x 2296 pixels in size. The activity logs are depicted in Figures 4.1 and 4.2.

Fictional Dataset Used

The fictional dataset we used in the activity logs consisted of 63 account access events: 61 sign-ins, and two failed sign-in attempts. The account access events occurred from 10 different accounts: Amazon, online banking, Dropbox, Facebook, Gmail, Instagram, LinkedIn, Netflix, Outlook, and Twitter. Those events occurred from five possible locations: *Cafe*, *Home*, *Office*, *Shopping mall*, and *Unknown*. We designed the dataset to depict a person with a typical nine-to-five work week. Her account access events usually start at 6am at home. By around 8am, the events usually occur from her office, and by 6pm the events occur from home.

Sunday May 20 to Saturday May 26, 2018

Access event type	Account accessed	Location of access	Day	Hour block
Sign-in	Gmail	Home	Sunday	8:00 am
Sign-in	Gmail	Home	Sunday	10:00 am
Sign-in	Instagram	Home	Sunday	2:00 pm
Sign-in	Twitter	Home	Sunday	8:00 pm
Sign-in	Instagram	Home	Sunday	8:00 pm
Sign-in	Twitter	Home	Monday	6:00 am
Sign-in	Facebook	Home	Monday	6:00 am
Sign-in	Gmail	Home	Monday	7:00 am
Sign-in	Outlook	Office	Monday	8:00 am
Sign-in	LinkedIn	Cafe	Monday	12:00 pm
Sign-in	Gmail	Home	Monday	6:00 pm
Sign-in	LinkedIn	Home	Monday	7:00 pm
Sign-in	Facebook	Home	Monday	8:00 pm
Sign-in	LinkedIn	Home	Monday	8:00 pm
Sign-in	Dropbox	Home	Monday	9:00 pm
Sign-in	Gmail	Office	Monday	11:00 pm
Sign-in	Gmail	Home	Tuesday	6:00 am
Sign-in	Outlook	Office	Tuesday	8:00 am
Sign-in	Gmail	Home	Tuesday	6:00 pm
Sign-in	LinkedIn	Home	Tuesday	6:00 pm
Sign-in	Amazon	Home	Tuesday	8:00 pm
Sign-in	Bank	Home	Tuesday	8:00 pm
Sign-in	Twitter	Home	Tuesday	9:00 pm
Sign-in	Facebook	Home	Wednesday	6:00 am
Sign-in	Outlook	Office	Wednesday	8:00 am
Sign-in	Instagram	Café	Wednesday	12:00 pm
Sign-in	Facebook	Home	Wednesday	6:00 pm
Sign-in	Instagram	Home	Wednesday	7:00 pm
Sign-in	Facebook	Home	Wednesday	8:00 pm
Sign-in	LinkedIn	Home	Wednesday	8:00 pm
Sign-in	Gmail	Home	Thursday	6:00 am
Sign-in	Twitter	Home	Thursday	6:00 am
Sign-in	Outlook	Office	Thursday	8:00 am
Sign-in	Instagram	Café	Thursday	12:00 pm
Sign-in	LinkedIn	Office	Thursday	1:00 pm
Sign-in	Gmail	Home	Thursday	6:00 pm
Sign-in	Twitter	Home	Thursday	8:00 pm
Sign-in	Facebook	Home	Thursday	8:00 pm
Sign-in	Gmail	Home	Thursday	9:00 pm
Sign-in	Twitter	Home	Friday	6:00 am
Sign-in	Gmail	Home	Friday	6:00 am
Sign-in	Outlook	Office	Friday	8:00 am
Sign-in	Amazon	Office	Friday	9:00 am
Sign-in	Bank	Office	Friday	9:00 am
Sign-in	Twitter	Café	Friday	10:00 am
Sign-in	Facebook	Office	Friday	12:00 pm
Sign-in	Gmail	Home	Friday	6:00 pm
Sign-in	Netflix	Home	Friday	8:00 pm
Sign-in	Facebook	Home	Friday	9:00 pm
Sign-in	Instagram	Home	Friday	9:00 pm
Failed sign-in attempt	Facebook	Unknown	Saturday	5:00 am
Sign-in	Facebook	Unknown	Saturday	6:00 am
Sign-in	LinkedIn	Home	Saturday	8:00 am
Sign-in	Bank	Home	Saturday	9:00 am
Failed sign-in attempt	Facebook	Home	Saturday	9:00 am
Sign-in	LinkedIn	Café	Saturday	2:00 pm
Sign-in	Gmail	Café	Saturday	2:00 pm
Sign-in	Instagram	Mall	Saturday	4:00 pm
Sign-in	Twitter	Mall	Saturday	4:00 pm
Sign-in	Instagram	Home	Saturday	6:00 pm
Sign-in	Twitter	Home	Saturday	8:00 pm
Sign-in	LinkedIn	Home	Saturday	8:00 pm
Sign-in	Netflix	Home	Saturday	9:00 pm

Figure 4.2: The activity log presented in the *Textlog* condition. Due to its length, participants had to scroll down to see all 63 rows of events, however, the size has been adjusted for this document.

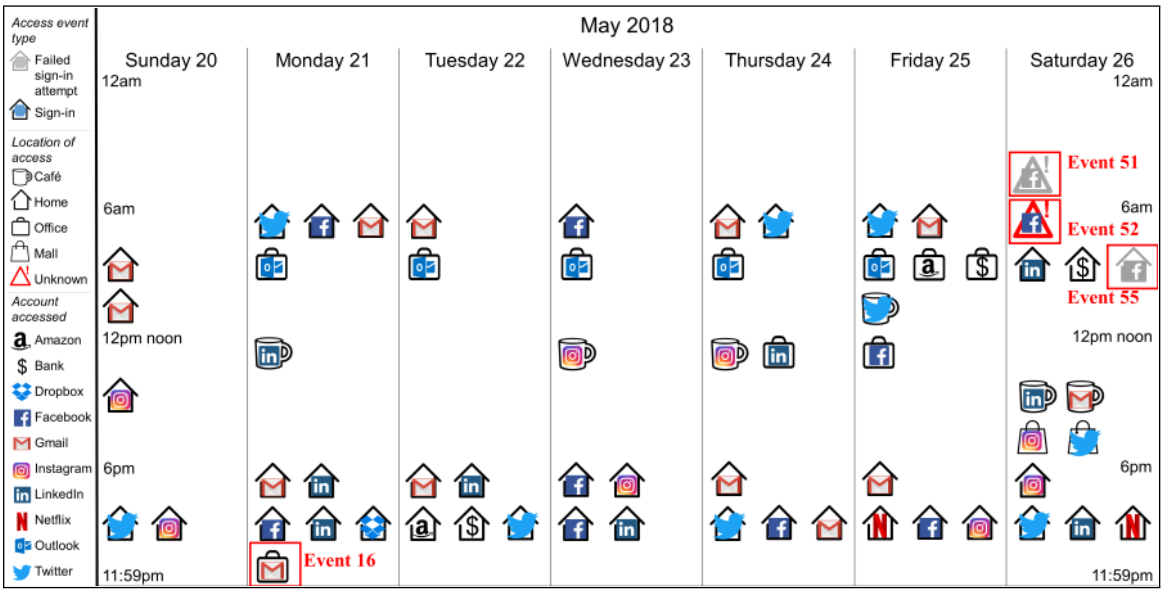
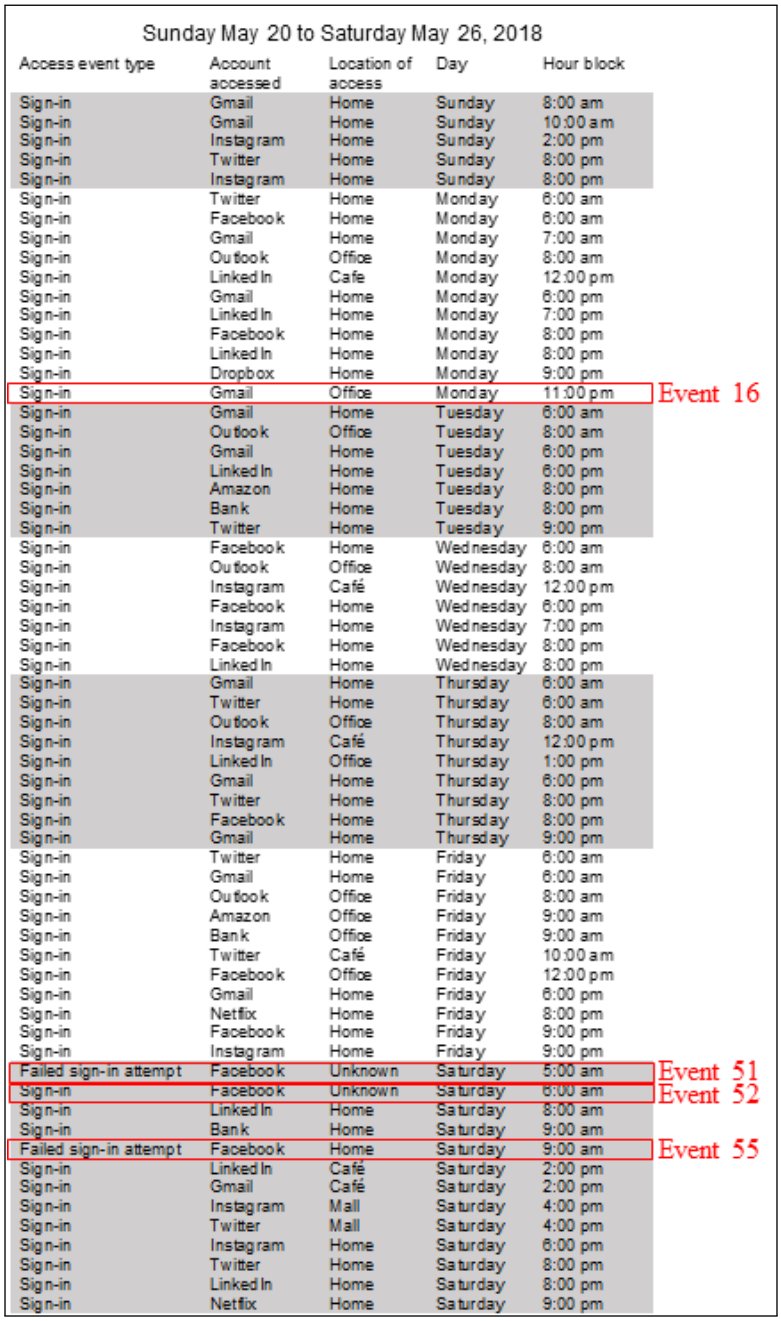


Figure 4.3: The four unusual events in *Diagram*. They were not disclosed in the survey, but artificially labelled here for clarity using red boxes.

Although participants could perceive any events to be unusual for a variety of subjective reasons, we included four *unusual events* that broke the pattern. We refer to the events in their chronological order. Event 16 depicts access to the user's Gmail account from *Office* at a time when she is usually at home, and after several events occurred from *Home*. This could suggest that she either went back to the office late at night to access her Gmail account, used a VPN to access her account, or, based on the access patterns of the rest of the week, that someone else accessed it at that time. Event 51 is a failed sign-in attempt to her Facebook account from an unknown location, followed by a successful sign-in (Event 52) from an unknown location, and then a failed sign-in attempt from home (Event 55). Those three events were designed to look like an account hijacking scenario: someone tried to access the user's Facebook account, got in successfully, then changed her password. Figures 4.3 and 4.4 show the four events that we designed to be unusual. They were not disclosed in the survey, but highlighted here for clarity.



Sunday May 20 to Saturday May 26, 2018

Access event type	Account accessed	Location of access	Day	Hour block	
Sign-in	Gmail	Home	Sunday	8:00 am	
Sign-in	Gmail	Home	Sunday	10:00 am	
Sign-in	Instagram	Home	Sunday	2:00 pm	
Sign-in	Twitter	Home	Sunday	8:00 pm	
Sign-in	Instagram	Home	Sunday	8:00 pm	
Sign-in	Twitter	Home	Monday	6:00 am	
Sign-in	Facebook	Home	Monday	6:00 am	
Sign-in	Gmail	Home	Monday	7:00 am	
Sign-in	Outlook	Office	Monday	8:00 am	
Sign-in	LinkedIn	Cafe	Monday	12:00 pm	
Sign-in	Gmail	Home	Monday	6:00 pm	
Sign-in	LinkedIn	Home	Monday	7:00 pm	
Sign-in	Facebook	Home	Monday	8:00 pm	
Sign-in	LinkedIn	Home	Monday	8:00 pm	
Sign-in	Dropbox	Home	Monday	9:00 pm	
Sign-in	Gmail	Office	Monday	11:00 pm	Event 16
Sign-in	Gmail	Home	Tuesday	6:00 am	
Sign-in	Outlook	Office	Tuesday	8:00 am	
Sign-in	Gmail	Home	Tuesday	6:00 pm	
Sign-in	LinkedIn	Home	Tuesday	6:00 pm	
Sign-in	Amazon	Home	Tuesday	8:00 pm	
Sign-in	Bank	Home	Tuesday	8:00 pm	
Sign-in	Twitter	Home	Tuesday	9:00 pm	
Sign-in	Facebook	Home	Wednesday	6:00 am	
Sign-in	Outlook	Office	Wednesday	8:00 am	
Sign-in	Instagram	Café	Wednesday	12:00 pm	
Sign-in	Facebook	Home	Wednesday	6:00 pm	
Sign-in	Instagram	Home	Wednesday	7:00 pm	
Sign-in	Facebook	Home	Wednesday	8:00 pm	
Sign-in	LinkedIn	Home	Wednesday	8:00 pm	
Sign-in	Gmail	Home	Thursday	6:00 am	
Sign-in	Twitter	Home	Thursday	6:00 am	
Sign-in	Outlook	Office	Thursday	8:00 am	
Sign-in	Instagram	Café	Thursday	12:00 pm	
Sign-in	LinkedIn	Office	Thursday	1:00 pm	
Sign-in	Gmail	Home	Thursday	6:00 pm	
Sign-in	Twitter	Home	Thursday	8:00 pm	
Sign-in	Facebook	Home	Thursday	8:00 pm	
Sign-in	Gmail	Home	Thursday	9:00 pm	
Sign-in	Twitter	Home	Friday	6:00 am	
Sign-in	Gmail	Home	Friday	6:00 am	
Sign-in	Outlook	Office	Friday	8:00 am	
Sign-in	Amazon	Office	Friday	9:00 am	
Sign-in	Bank	Office	Friday	9:00 am	
Sign-in	Twitter	Café	Friday	10:00 am	
Sign-in	Facebook	Office	Friday	12:00 pm	
Sign-in	Gmail	Home	Friday	6:00 pm	
Sign-in	Netflix	Home	Friday	8:00 pm	
Sign-in	Facebook	Home	Friday	9:00 pm	
Sign-in	Instagram	Home	Friday	9:00 pm	
Failed sign-in attempt	Facebook	Unknown	Saturday	5:00 am	Event 51
Sign-in	Facebook	Unknown	Saturday	6:00 am	Event 52
Sign-in	LinkedIn	Home	Saturday	8:00 am	
Sign-in	Bank	Home	Saturday	9:00 am	
Failed sign-in attempt	Facebook	Home	Saturday	9:00 am	Event 55
Sign-in	LinkedIn	Café	Saturday	2:00 pm	
Sign-in	Gmail	Café	Saturday	2:00 pm	
Sign-in	Instagram	Mall	Saturday	4:00 pm	
Sign-in	Twitter	Mall	Saturday	4:00 pm	
Sign-in	Instagram	Home	Saturday	6:00 pm	
Sign-in	Twitter	Home	Saturday	8:00 pm	
Sign-in	LinkedIn	Home	Saturday	8:00 pm	
Sign-in	Netflix	Home	Saturday	9:00 pm	

Figure 4.4: The four unusual events in *Textlog*. They were not disclosed in the survey, but artificially labelled here for clarity using red boxes.

Visual Encodings

In *Diagram*, we implemented design changes based on our first study. We moved the legend vertically with a darker line to demarcate the difference between the activity log and the legend. In addition, we used a larger dataset of events to produce a denser visualization. Each account was represented by its commercial icon, and the location of access was represented by a line glyph which enclosed the account icon. We used a mug glyph to represent *Cafe*, a house glyph to represent *Home*, a briefcase glyph to represent *Office*, and a red triangle with an accompanying red exclamation mark to represent *Unknown*. Contrary to our first study, *Diagram* was a static visualization without pop-overs. We implemented each event as a hotspot to enable participants to click them.

Diagram uses the same calendar layout from our first study; the vertical axis is a continuous scale for the time of day and the horizontal axis is an ordinal scale for the day of the week. Within the day of week, events that happened within the same hour block were placed from left to right in chronological order. The fictional dataset consisted of six variables which were mapped to *Diagram* in the following manner:

1. the vertical placement signalled the hour of day
2. the horizontal placement indicated the day of week
3. an icon identified the account domain, e.g., Twitter
4. an enclosing line glyph identified the location, e.g., Office
5. a greyed-out appearance indicated a failed sign-in attempt, e.g., see Events 51 and 55 in Figure 4.3
6. a fully-coloured appearance indicated a successful sign-in, e.g., see Event 16 in Figure 4.3

As in the first iteration, we used saturated colours on a white background, and we used horizontal proximity to indicate grouping by hour of day so that participants could infer the hourly usage patterns across the week. Because form, colour, and spatial position are pre-attentively processed, we use the form of enclosures (ie,

glyphs) to signal location, the colour, red, to flag events from an unknown location, and spatial position to represent time. We appeal to the *law of similarity* [129] by denoting location with glyphs. This is because we wanted participants to be able to infer which access events occurred from the same location.

Textlog listed the same 63 events, but in a table format. We used black text on a white and light grey background for readability. In order to make both activity logs comparable, we designed *Textlog* to display the information as closely as possible to *Diagram*. We used shading to distinguish the days of the week from one another because they are distinguished in *Diagram*. Instead of using an exact time stamp, we listed the hour block during which the event had occurred as is depicted in *Diagram*.

4.2.5 Participants

Participants were recruited in two ways: through Qualtrics and through social media. We paid Qualtrics $5.67 CAD per participant (including 13% tax in Canada). From this, participants received reward points and monetary compensation for completing the survey. We initially collected 196 responses, 190 of which were recruited by Qualtrics. Six participants voluntarily took part in the survey as a result of our own recruitment efforts without receiving compensation. The survey consisted of two main parts: security questions (to address ***RQ1*** and ***RQ2***), and activity log questions (to address ***RQ3***). Of the complete responses we kept for analysis, we did not keep all of the responses to the activity log questions. Table 4.1 shows how the participants were distributed across groups and conditions.

We excluded 26 participants from the original 196 because they picked the same Likert response in the entire survey, answered open-ended questions with non-words or non-English, did not follow our instructions to write *"no comment"* when they did not have an answer, on rare occasion were assigned to the wrong group, or failed our quality-check question. Our quality check question was designed to test attention: *What account is this survey about?* Participants could pick from six choices, including *Facebook* and *Google/Gmail*. The order of the choices was randomized. This resulted in 170 valid responses used to answer ***RQ1*** and ***RQ2***.

We retained the original participant pseudonyms and did not re-number them after

Table 4.1: How participants are distributed by group, *Facebook* or *Google*, and condition, *Diagram* or *Textlog*.

	Facebook	**Google**	**Total**
Diagram	28	37	65
Textlog	46	23	69
Responses for RQ1 and RQ2 only	21	15	36
Total	95	75	$N = 170$

excluding responses. Of our valid 170 responses, we excluded the activity log responses of 36 participants because they did not understand what the activity log was (12 participants in the *Diagram* condition, and nine in the *Textlog* condition), mistakenly believed that the activity log was their own (two *Diagram* and five *Textlog*), did not understand the questions (one *Diagram* and three *Textlog*), or misinterpreted the events that are adjacent to one another for being simultaneous (one *Diagram* and three *Textlog*). We used the remaining 134 responses (65 *Diagram* and 69 *Textlog*) for analysis for ***RQ3***.

Half of the participants were female. Their ages ranged from 18 to 83 years old, with a mean age of 48, $SD = 15$ years. Thirty-nine of 170 were technical users based on their responses to the question, *Do you, or have you ever worked in a computer, computer security, or information technology (IT)-related field?*

4.3 Data Analysis Process

4.3.1 Qualitative Analysis

We analyzed participants' responses to four open-ended questions using a thematic analysis approach [20]. The researcher started by reading the provided answers several times to understand the data as a collective body. She then derived codes directly from the words that conveyed key concepts. Depending on the level of detail that participants provided, the researcher grouped codes into categories, and merged the ones with the same underlying concepts. This was done until a code applied to each provided answer. This yielded information about participants' reports of account

compromise, which other entities participants hold responsible for their account security, what would lead them to delete their Facebook or Google account, and how participants decided events in the activity logs were unusual.

4.3.2 Quantitative Analysis

We used SPSS to analyze the data to answer each of our research questions.

Our first research question is: ***RQ1.*** *Who do end users perceive is responsible for (a) preventing attacks, (b) alerting the user or reporting to the service provider of unusual activity, and (c) for recovering the account after an attack?* We measured this with 5-point Likert scale questions. We asked participants about a total of 10 *responsibility items*. To determine whether there are significant differences between the two entities (ie, user and service provider) in how participants allocated responsibility for account security, we ran 10 Wilcoxon signed rank tests. We used this test because the Likert scale responses are ordinal data from two dependent groups [40]. This is because we compare the responsibility of the *end user* (i.e., the participant), to the responsibility of the *service provider* (i.e., Facebook or Google) given by all the participants. In other words, each participant, regardless of group, provided two Likert scale responses for 10 responsibility items: one for themselves, and one for the service provider (see Section 4.4.2). To control for familywise error, we applied a Bonferroni correction [40].

As an extension of *RQ1*, we also analyzed the difference between the two groups (i.e., Facebook and Google) in how they allocated responsibility. To this end, we ran 10 Mann-Whitney tests. We chose this test because the Likert scale responses are ordinal data from two independent groups [40]. This is because we compare the Facebook participants' Likert scale responses to the Google participants, on the 10 responsibility items. To control for familywise error, we applied a Bonferroni correction [40].

To answer ***RQ2.*** *Why do users trust (or mistrust) Facebook and Google?*, we asked the following question in the survey: *What makes [service provider] trustworthy? Check all that apply.* We gave 10 choices, the 11th being the option to define their own reason. Participants could pick more than one item. We calculated the total

number of times that each close-ended response was selected. We also compared those frequencies in the Facebook group with the Google group. We also asked for Likert scale responses to the statement, *[Facebook/Google] is able to keep my data safe.* We used a Mann-Whitney test to determine whether there are differences between the two groups, Facebook and Google, in participants' perception of their SP's ability to keep their accounts safe. We chose this test because the Likert scale responses are ordinal data from two independent groups (Facebook and Google) [40].

Our third research question was: ***RQ3.*** *How does the visual activity log (Diagram) compare with the text activity log (Textlog) in (a) how effective it is in identifying unusual activity, and (b) how comprehensible it is?* We used Mann-Whitney tests to determine whether there are differences between the two activity logs on the participants' discovery and comprehension scores. Each participant received a score out of four, corresponding to how many of the four unusual events they discovered. They also received a score out of two, corresponding to how many of the two comprehension questions they answered correctly. We chose this test because the scores are from two independent groups (*Diagram* and *Textlog*) and not normally distributed [40].

4.4 Results

We report on participants' existing attitudes and beliefs with regards to account monitoring. We will then present the results organized by research question. In figures, we use grey and black for *Diagram* and *Textlog* and blue and yellow for Facebook and Google.

4.4.1 Existing Attitudes and Practices

We report on our participants' beliefs and attitudes in order to establish a context for our findings. We asked the participants, *How concerned are you about the security of your [Facebook/Google] account?* to which they could pick from a 5-point Likert scale. Figure 4.5 shows the percentage of participants who picked each Likert response. Participants from the Facebook group were significantly more concerned about the security of their accounts than the Google group, $U = 2,968$, $z = -1.98$, $p = .048$, $r = -0.15$. We speculate that the higher concern for security is a result of recent

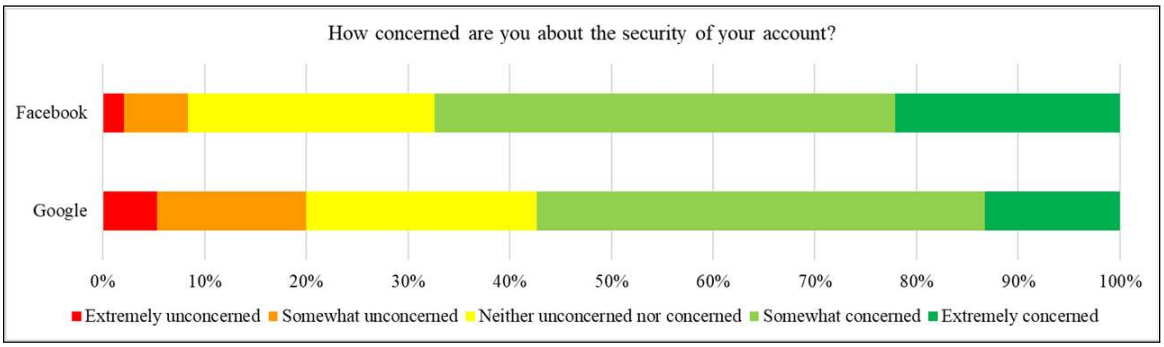


Figure 4.5: Participant responses to, *How concerned are you about the security of your account?*, by group.

media coverage on third-party access to Facebook users' data [38].

When asked if their Facebook/Google account has ever been compromised, a minority of the participants (11 Facebook and six Google) answered yes (Figure 4.6). Those participants were asked a follow-up question, *What happened and how did you find out?* Within the Facebook group, three participants indicated that Facebook notified them of the unusual activity. Two participants explained that they had been impersonated by someone online. One participant explained, *"It was hijacked by FaceBook staff in order to get me to change my name. I needed access for school project but they refused me access to my School pages"* (P157). Although we defined *compromise* in the survey as *someone has gained access to your account without your permission*, this response reveals that the term can be interpreted subjectively. It further raises the question of what users perceive as their rights in contrast with their responsibilities in conforming to the service provider's terms of use. One participant did not comment and two participants simply answered that their Facebook account had been compromised. Two Facebook participants explained that they were notified by their friends of posts or messages that did not seem to have been posted by the participant. This points to a voluntary "neighbourhood watch" mindset that people engage in as part of their online circles. When asked if any other entities were responsible for account security, one participant answered, *"friends noticing unusual activities"* (P024).

Within the Google group, three participants explained that they were notified by Google of unusual account activity. One participant witnessed third-party activity

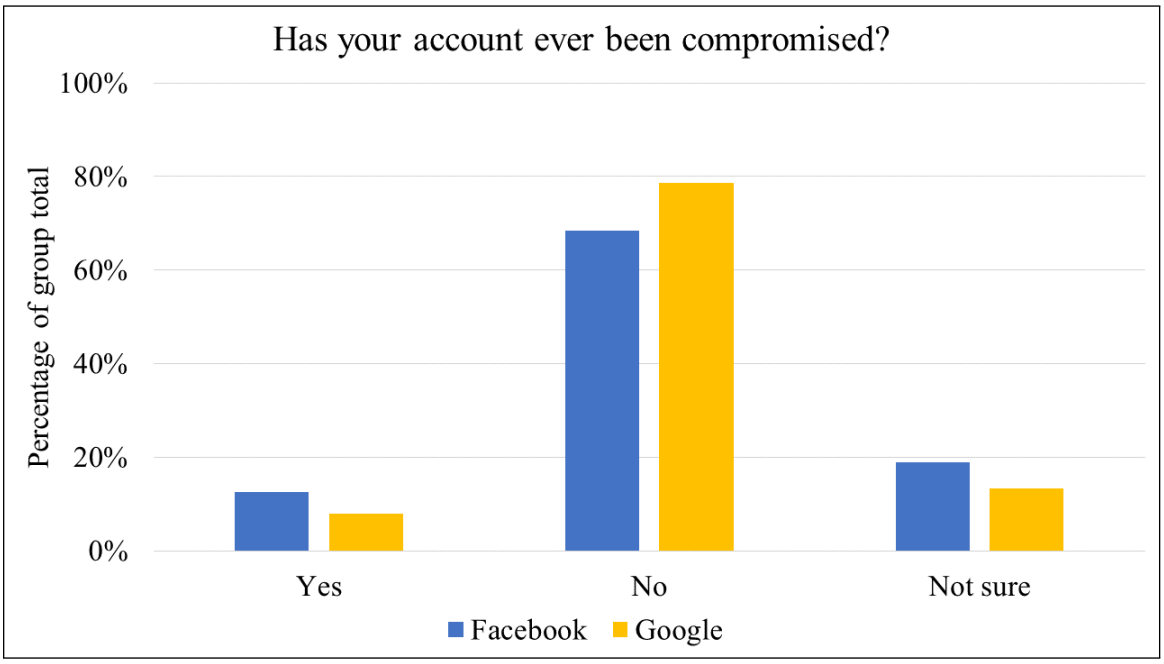


Figure 4.6: The percentage of participants whose accounts were compromised at some point, by group.

on her accounts; *"all my online mail accounts were hacked, items would be deleted as I was looking at them"* (P004). One participant indicated what we interpreted as having received unwanted emails (P049), and another participant *"was being sent email meant for another user"* (P027). P027 and P049 seemed to have interpreted account compromise differently than our definition.

We defined *security alert* as an alert about *a sign-in event or potential risk to your [Facebook/Google] account. (For example, an attempted sign-in from an unrecognized device.)* Overall, 35% of Facebook participants and 52% of Google participants reported receiving at least one security alert within the last year. Within this group, most reported that they check their activity logs. Figure 4.7 shows a side-by-side comparison of the two groups.

Figure 4.8 shows the breakdown of the responses to the close-ended question, *How often do you check the history of recent sign-ins for your account?* Overall, 39% of the Facebook participants and 51% of the Google participants report checking their activity logs at least once a year. In other words, 61% of Facebook participants and 49% of Google participants never check their activity logs. The most common reason

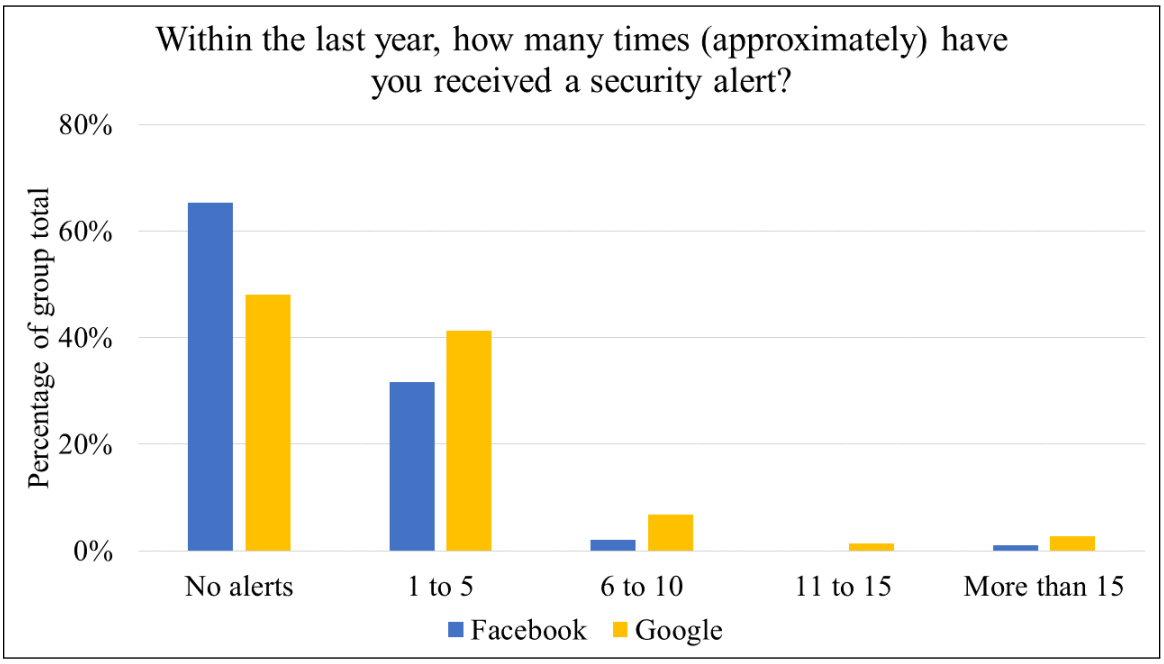


Figure 4.7: The percentage of participants who received security alerts regarding their account within the last year, by group.

for not checking is that they do not know that this information exists. It is surprising that so many Facebook participants reported checking their activity logs, because doing so is arguably an unintuitive process requiring download of a user's profile data (see Section 2.2.2).

These findings are in contrast with our first study: 80% of participants do not check their account activity logs. We postulate possible explanations for these reversed findings. Perhaps an increase in high-profile security incidents within the last

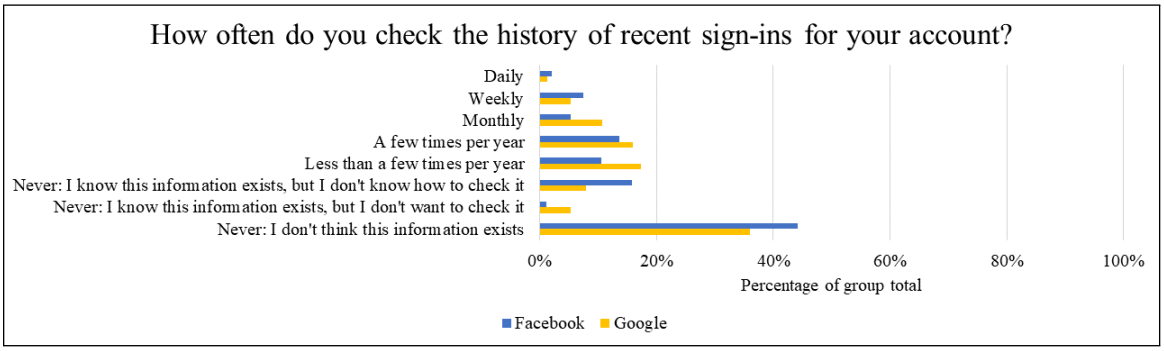


Figure 4.8: Percentage of participants who check their sign-in history, by group.

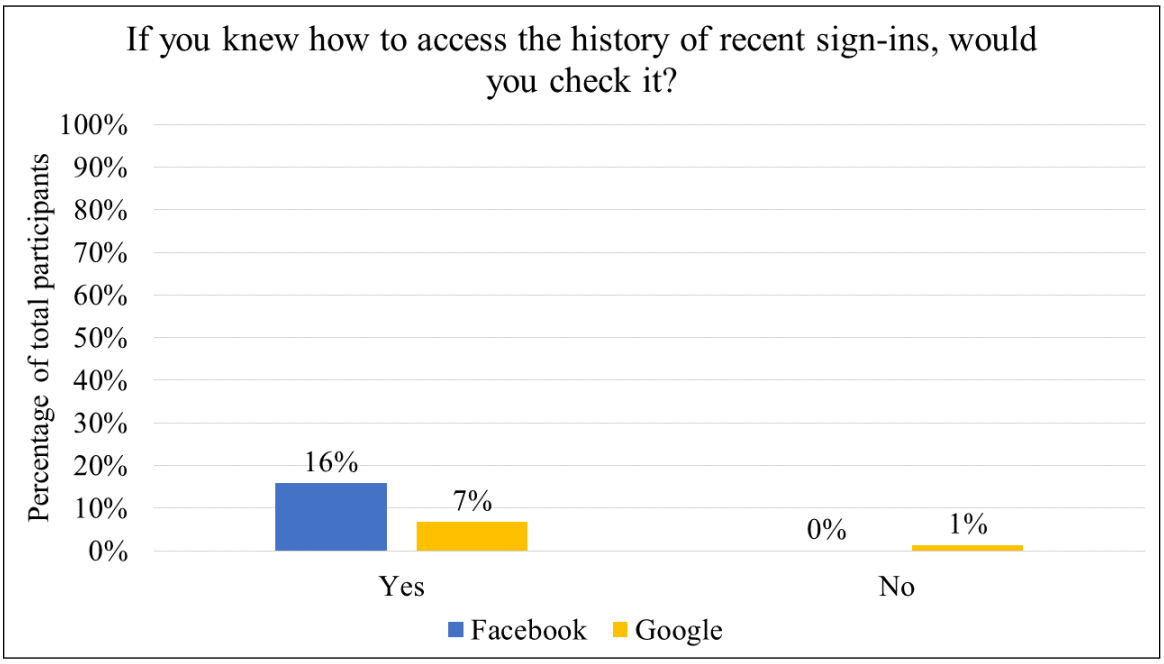


Figure 4.9: Percentage of participants who would check their sign-in history if they knew how to access it, by group.

year has increased end-user awareness of their account security. Alternatively, results from Study 1 could be due to the homogeneity and/or size of our sample.

For the participants who do not know, we asked the follow-up question, *If you knew how to find this information, would you check it?* For participants who do not want to check, we asked them a close-ended question, *Why would you not check the history of recent sign-ins for your account?* For participants who were unaware, we asked, *If this information is available, would you check it?* Figures 4.9, 4.10, and 4.11 show the breakdown of the answers. The majority of participants who do not currently check their activity logs indicate a desire to do so.

We asked the participants who check their activity logs, *How easy is it to understand the history of recent sign-ins for your [Facebook/Google] account?* Figure 4.12 shows their answers on a 5-point Likert scale. A Mann-Whitney test found no significant differences between the Google and Facebook participants in how easy it is to understand their activity logs, $U = 780.5$, $z = .87$, $p = .385$, $r = .10$.

We asked the close-ended question, *I would like [Facebook/Google] to determine whether the account activity is from me by comparing it to:* the five options depicted

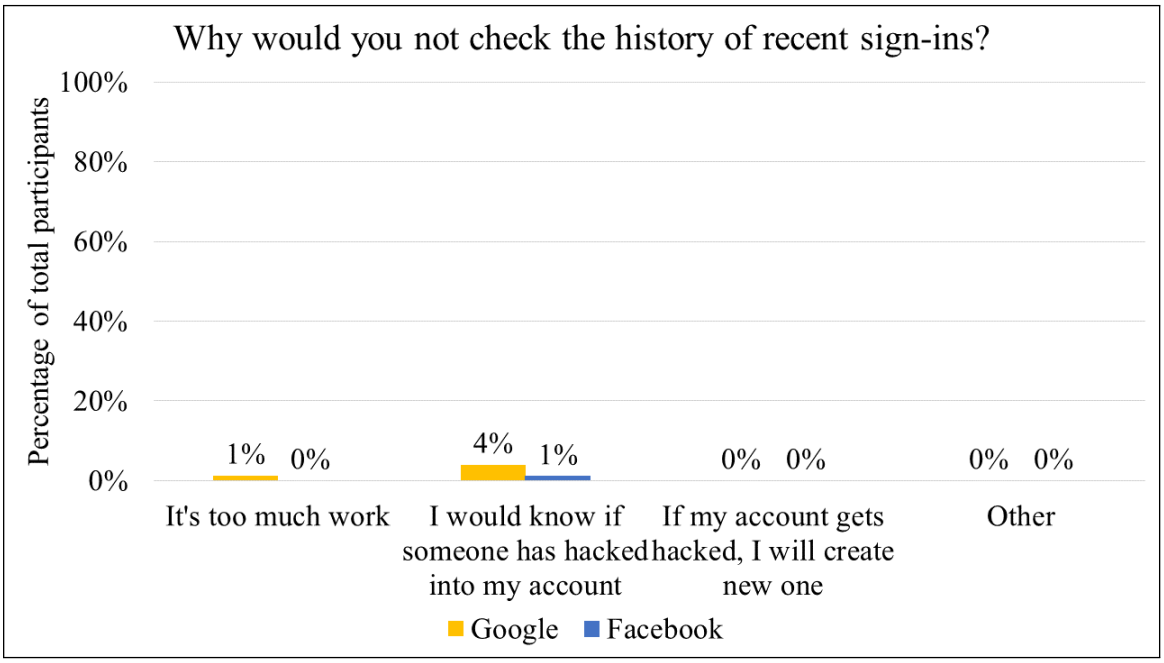


Figure 4.10: Percentage of participants who do not want to check their sign-in history, by group.

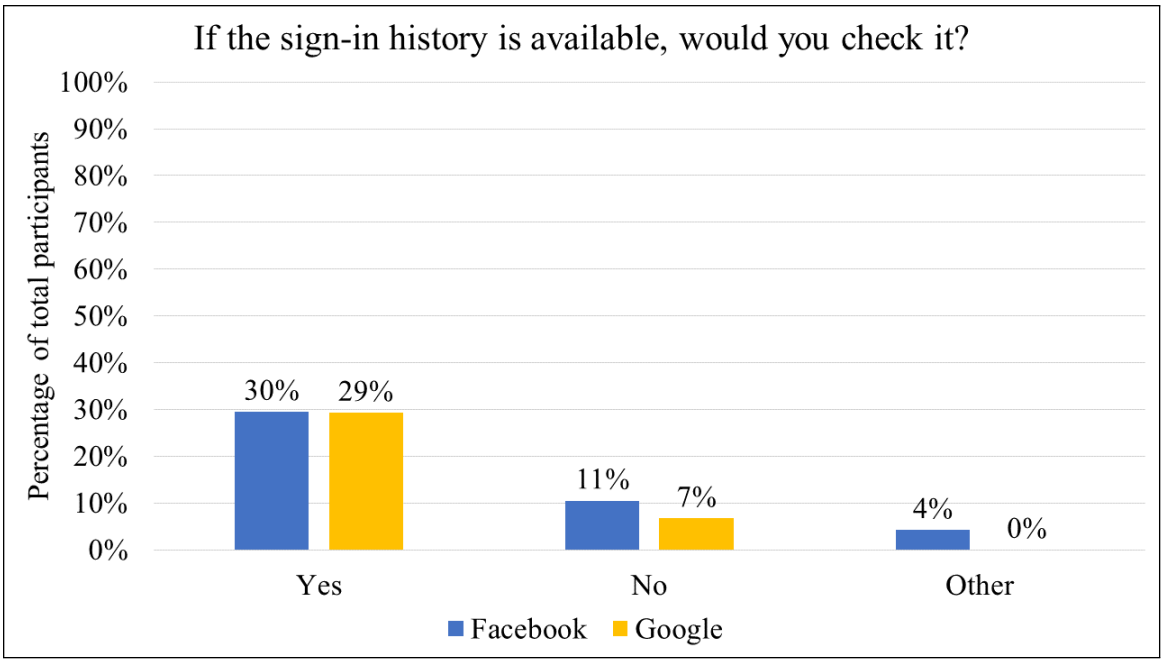


Figure 4.11: The overall percentage of participants who would check their sign-in history if it was available, by group.

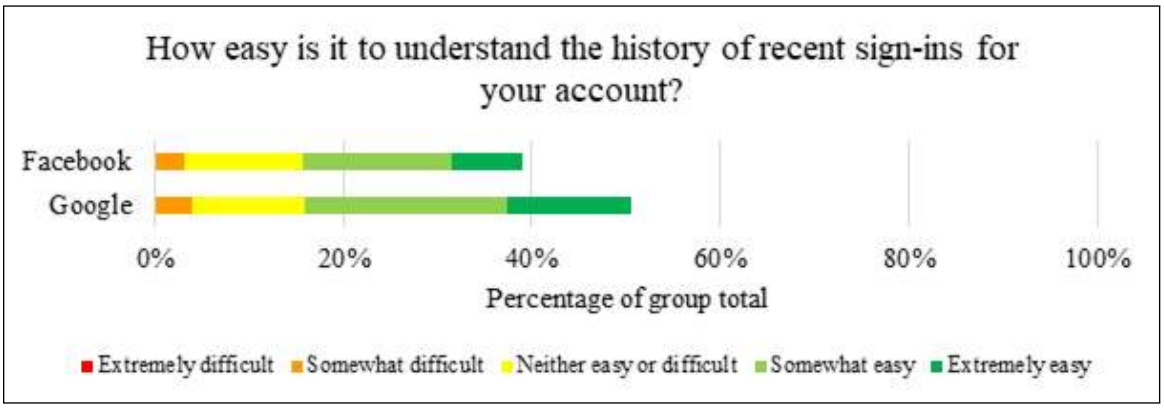


Figure 4.12: Percentage of total participants who rated how easy it is to understand their activity logs, by group. The remaining participants did not see this question because they do not check their activity logs.

in Figure 4.13. As shown, most responses are split between comparison with general patterns, and comparison with personal activity.

We asked the open-ended question, *What would lead you to delete your [Facebook/Google] account?* We organized their answers into 12 categories, depicted in Figure 4.14. Some participants provided more than one reason, and a few did not provide any reasons. The top reason in both groups was compromise or lack of protection for their accounts. Only Facebook participants were concerned with risks to reputation (2%), fraud (3%) and social dangers (10%), while only Google participants were concerned with spam (5%).

Summary

Participants from the Facebook group were significantly more concerned about the security of their accounts than the Google group. A minority of our participants reported having their account compromised in the past. 35% of Facebook participants and 52% of Google participants reported receiving at least one security alert within the last year. 61% of the Facebook participants and 49% of the Google participants never check their account activity logs. The most common reason for not checking them is that they do not know that this information exists. Most of the participants who do not check their activity logs indicate a desire to do so. To identify unusual account activity, most participants would like their service provider to either compare

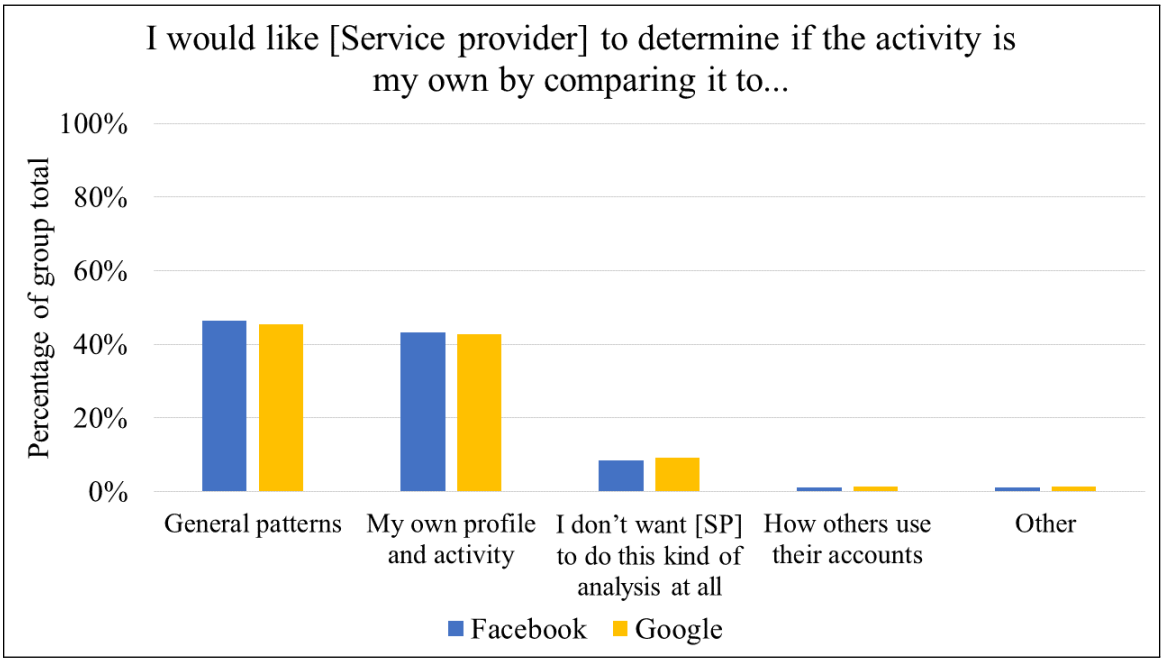


Figure 4.13: How participants would like their activity to be analyzed, by group.

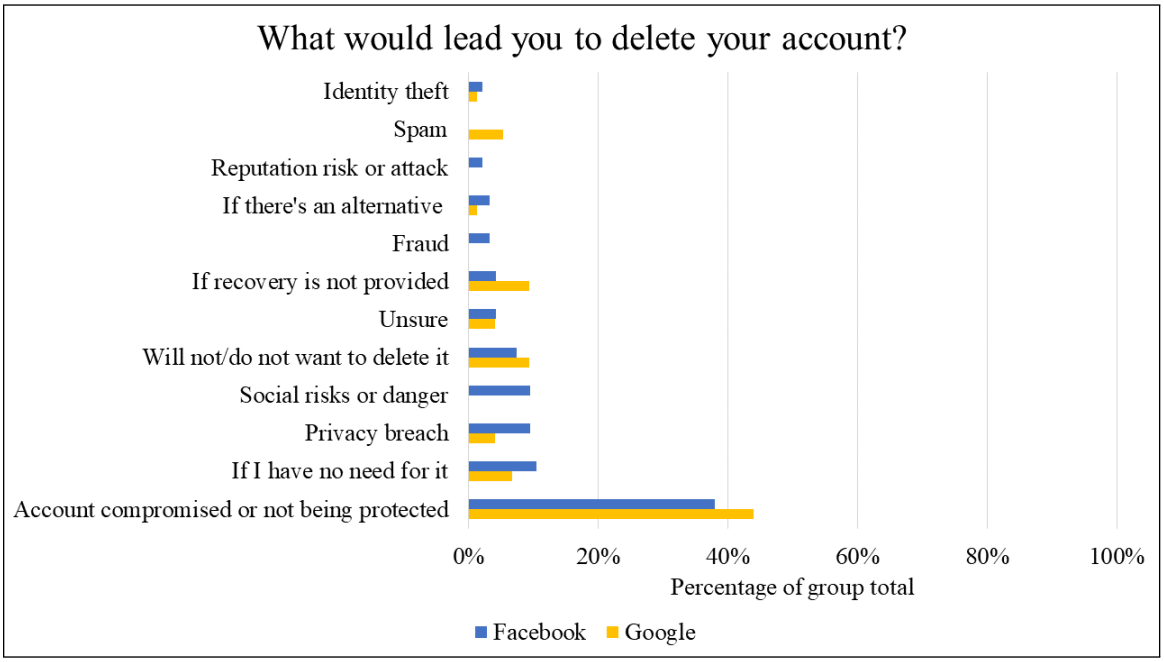


Figure 4.14: The reasons provided by participants for why they would delete their account, by group.

their account activity with general patterns or with their own profile and data. The top reason that would lead them to delete their accounts are either their accounts being compromised or lacking protection from their service provider.

4.4.2 Responsibility for Account Security

RQ1 asked, *Who do end users perceive is responsible for (a) preventing attacks, (b) alerting the user or reporting to the service provider of unusual activity, and (c) for recovering the account after an attack?* For the purpose of this thesis, we focus on responsibility, not accountability because the former implies a subjective, moral meaning, whereas the latter implies an objective, legal meaning. To answer this research question, we asked participants: *To what extent do you believe that [Facebook/Google] is responsible for...* We also asked them *To what extent do you believe that you are responsible for...* for each of 10 responsibility items. Using Likert scale questions, we assessed the distribution of responsibility across the 10 items for the two entities, the service provider and the user. Figure 4.15 depicts the mean Likert score responses for both Facebook and Google combined.

The definition of phishing. When we asked participants about phishing, we defined it as "the attempt to obtain sensitive information such as usernames, passwords, and credit card details (and money), often for malicious reasons, by disguising as a trustworthy entity." When we asked them about their perceived responsibility for preventing phishing attacks, we worded the question as follows, "To what extent do you believe [service provider/you] are responsible for [...] preventing you from falling victim to a phishing attack?" By falling victim, we intended the meaning to be that the user would have fallen for the "bait" of the attacker and as a result, experienced stolen credentials, stolen data, financial loss, etc. Therefore, there are two parts to phishing, the disguise used by the attacker, and the bait. The bait comes in different forms, such as a URL in an email that links to malicious code or a webpage that looks exactly like the user's online banking site. We did not include the bait in our definition of phishing. In retrospect, giving examples of bait to our participants would have been better to ensure that the meaning of phishing was clear to them.

Did participants attribute responsibility to themselves differently than

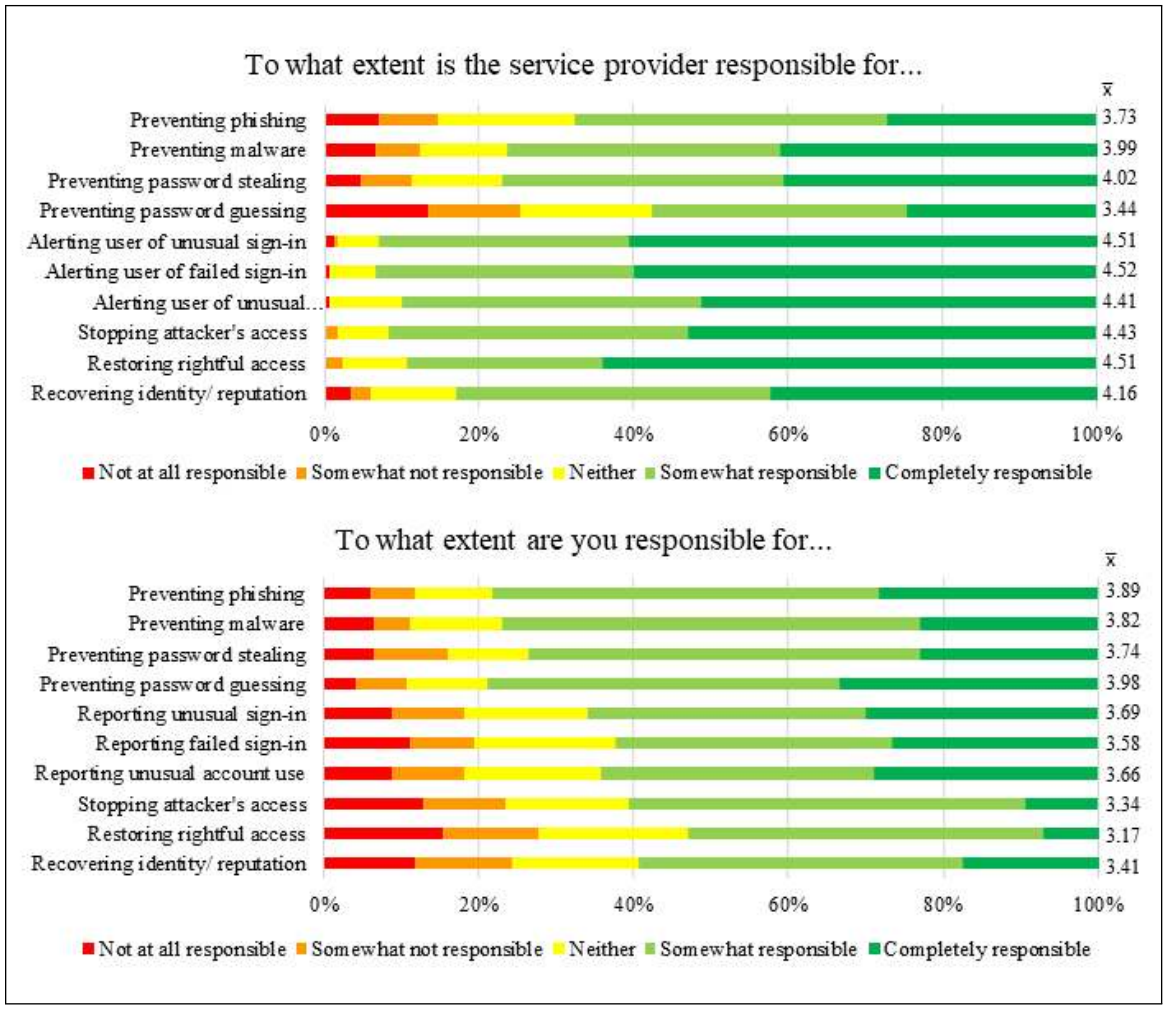


Figure 4.15: Likert scores depicting how participants allocated responsibility for account security to the service provider (top) and to themselves (bottom).

Table 4.2: Mean, standard deviation (SD), and median Likert scores of participants' allocation of responsibility between themselves and the service provider. Items that are significantly different between the entities are bolded.

Responsibility item	Mean (SD)		Median	
	User	SP	User	SP
Prevent phishing	3.89 (1.07)	3.73 (1.15)	4	4
Prevent malware	3.82 (1.05)	3.99 (1.16)	4	4
Prevent password stealing	3.74 (1.11)	4.02 (1.10)	4	4
Prevent password guessing	3.98 (1.04)	3.44 (1.34)	4	4
Alert/report unusual sign-in	3.69 (1.24)	4.51 (.73)	4	5
Alert/report failed sign-in	3.58 (1.27)	4.52 (.66)	4	5
Alert/report unusual account use	3.66 (1.24)	4.41 (.68)	4	5
Stop attacker's access	3.34 (1.19)	4.43 (.70)	4	5
Restore access	3.17 (1.21)	4.51 (.75)	4	5
Restore identity/reputation	3.41 (1.25)	4.16 (.96)	4	4

they did to their service providers? We compared how participants attributed responsibility between themselves and their service provider, regardless of group. Wilcoxon signed rank tests with Bonferroni correction[1] find that the distribution of responsibility was significantly different between the two entities for 7 out of the 10 responsibility items listed. Table 4.3 lists the results, with significant results bolded.

Participants viewed themselves as more responsible than their service provider to prevent the guessing of their passwords. The service provider received more responsibility for alerting and recovery. Participants shared the responsibility with their service provider to prevent phishing attacks, malware attacks, and password stealing.

Did participants generally attribute responsibility differently between the two groups? We conducted 20 Mann-Whitney tests (10 responsibility items $\times 2$ entities) with Bonferroni correction,[2] looking for differences in between the two groups (Facebook and Google) on each of the responsibility items. Tables 4.4 and 4.6 list the descriptive statistics and Tables 4.5 and 4.7 show the results. There was no effect of group on how participants allocated responsibility for the items.

[1]Because we ran 10 Wilcoxon signed rank tests, we set the alpha level to .005. The SPSS output shows the original p value, but indicates which p values are significant with the new alpha level. We list the p values as they appear in SPSS.

[2]Because we ran 20 Mann-Whitney tests (10 responsibility items x 2 entities), we set the alpha level to .0025. The SPSS output shows the original p value, but indicates which p values are significant with the new alpha level. We list the p values as they appear in SPSS.

Table 4.3: Wilcoxon signed rank test results with Bonferroni correction for the differences in the distribution of responsibility between the two entities. Significant results are bolded.

Responsibility item	Result		
Prevent phishing	$T = 3,338$	$p = .217$	$r = .09$
Prevent malware	$T = 2,073.50$	$p = .152$	$r = -0.11$
Prevent password stealing	$T = 1,743.50$	$p = .030$	$r = -0.17$
Prevent password guessing	$\boldsymbol{T = 3{,}618}$	$\boldsymbol{p < .001}$	$\boldsymbol{r = .27}$
Alert/report unusual sign-in	$\boldsymbol{T = 384}$	$\boldsymbol{p < .001}$	$\boldsymbol{r = -.53}$
Alert/report failed sign-in	$\boldsymbol{T = 336.5}$	$\boldsymbol{p < .001}$	$\boldsymbol{r = -.59}$
Alert/report unusual account use	$\boldsymbol{T = 828}$	$\boldsymbol{p < .001}$	$\boldsymbol{r = -.49}$
Stop attacker's access	$\boldsymbol{T = 574}$	$\boldsymbol{p < .001}$	$\boldsymbol{r = -.60}$
Restore access	$\boldsymbol{T = 480}$	$\boldsymbol{p < .001}$	$\boldsymbol{r = -.68}$
Restore identity/reputation	$\boldsymbol{T = 1{,}617}$	$\boldsymbol{p < .001}$	$\boldsymbol{r = -.37}$

Table 4.4: Mean, standard deviation (SD), and median Likert scores of participants' allocation of responsibility to themselves, by group.

Responsibility item	Mean (SD)		Median	
	Facebook	Google	Facebook	Google
User prevent phishing	3.84 (1.19)	3.95 (.90)	4	4
User prevent malware	3.76 (1.15)	3.91 (.90)	4	4
User prevent password stealing	3.75 (1.19)	3.73 (1.02)	4	4
User prevent password guessing	3.94 (1.09)	4.03 (.97)	4	4
User report unusual sign-in	3.63 (1.28)	3.76 (1.20)	4	4
User report failed sign-in	3.53 (1.29)	3.65 (1.26)	4	4
User report unusual account use	3.67 (1.25)	3.64 (1.23)	4	4
User stop attacker's access	3.14 (1.25)	3.59 (1.05)	4	4
User restore access	3.00 (1.31)	3.39 (1.04)	4	4
User restore ID / reputation	3.25 (1.31)	3.61 (1.14)	4	4

Table 4.5: Mann-Whitney test results with Bonferroni correction for differences in how participants allocate responsibility to themselves, by group. None of the results are significant.

Responsibility item	Result			
User prevent phishing	$U = 3,544$	$z = -.063$	$p = .95$	$r = -.005$
User prevent malware	$U = 3,714.5$	$z = .52$	$p = .6$	$r = .04$
User prevent password stealing	$U = 3,395$	$z = -.57$	$p = .570$	$r = -.04$
User prevent password guessing	$U = 3,693.5$	$z = .44$	$p = .659$	$r = .03$
User report unusual sign-in	$U = 3,717$	$z = .51$	$p = .613$	$r = .04$
User report failed sign-in	$U = 3,734.5$	$z = .56$	$p = .58$	$r = .04$
User report unusual account use	$U = 3,467.5$	$z = -.31$	$p = .757$	$r = -.02$
User stop attacker's access	$U = 4,260$	$z = 2.36$	$p = .018$	$r = .18$
User restore access	$U = 4,045$	$z = 1.61$	$p = .109$	$r = .39$
User restore ID / reputation	$U = 4,075.5$	$z = 1.68$	$p = .092$	$r = .13$

Table 4.6: Mean, standard deviation (SD), and median Likert scores of participants' allocation of responsibility for each service provider.

Responsibility item	Mean (SD)		Median	
	Facebook	Google	Facebook	Google
SP prevent phishing	3.83 (1.16)	3.60 (1.13)	4	4
SP prevent malware	4.09 (1.19)	3.85 (1.12)	5	4
SP prevent password stealing	4.06 (1.19)	3.96 (.99)	4	4
SP prevent password guessing	3.47 (1.34)	3.39 (1.36)	4	4
SP alert unusual sign-in	4.45 (.81)	4.57 (.62)	5	5
SP alert failed sign-in	4.56 (.63)	4.48 (.70)	5	5
SP alert unusual account use	4.42 (.71)	4.39 (.66)	5	4
SP stop attacker's access	4.47 (.71)	4.37 (.67)	5	4
SP restore access	4.60 (.69)	4.40 (.81)	5	5
SP restore ID / reputation	4.20 (1.01)	4.11 (.91)	4	4

Table 4.7: Mann-Whitney test results with Bonferroni correction for differences in how participants allocate responsibility to the service provider, by group. None of the results are significant.

Responsibility item	Result			
SP prevent phishing	$U = 3,085$	$z = -1.57$	$p = .116$	$r = -.12$
SP prevent malware	$U = 2,959$	$z = -2.01$	$p = .044$	$r = -0.15$
SP prevent password stealing	$U = 3,162.5$	$z = -1.34$	$p = .181$	$r = -.10$
SP prevent password guessing	$U = 3,434$	$z = -.42$	$p = .677$	$r = -.03$
SP alert unusual sign-in	$U = 3,782$	$z = .80$	$p = .424$	$r = .06$
SP alert failed sign-in	$U = 3,349.5$	$z = .77$	$p = .439$	$r = -.06$
SP alert unusual account use	$U = 3,404.5$	$z = -.55$	$p = .581$	$r = -.04$
SP stop attacker's access	$U = 3,204.5$	$z = -1.26$	$p = .207$	$r = -.10$
SP restore access	$U = 3,086.5$	$z = -1.76$	$p = .078$	$r = -.14$
SP restore ID / reputation	$U = 3,225.5$	$z = -1.14$	$p = .253$	$r = -.09$

Are any other entities responsible for account security?

We asked our participants to list any other entities they believe are responsible for prevention, alerting, and recovery. Only 13 of 170 provided a third entity. Except otherwise listed, each answer was mentioned by only one participant. Facebook participants identified their friends, mobile device, email provider, third-party accounts linked to Facebook, antivirus program ($n = 2$), and their browser. Google participants mentioned other software, antivirus program, third-party accounts linked to Google, and the government ($n = 3$).

It was not surprising to us that a few participants mentioned antivirus as a responsible entity. In fact, we had included antivirus as an entity in our first iteration of the survey, but removed it for brevity. The role of third-party services (e.g., linked accounts, devices, software, browsers, email provider) in the security of first-party domain accounts is unclear to us and needs further investigation.

Summary

We found no significant differences in how much responsibility participants attributed to Facebook versus Google as service providers. However, participants identified clear roles with respect to primary responsibility between themselves and their service provider. Figure 4.16 depicts a high-level summary of these findings.

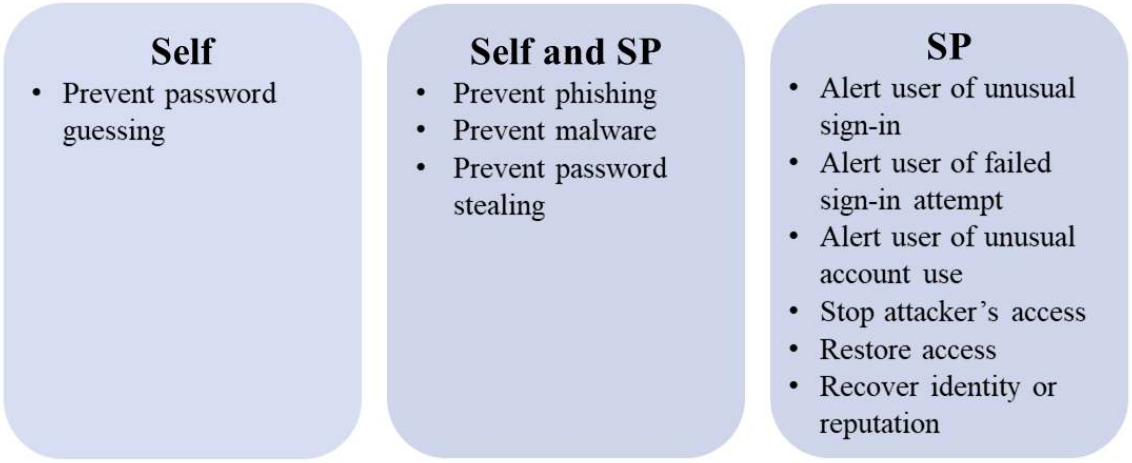


Figure 4.16: How participants attributed primary responsibility for their account security.

4.4.3 Trust

Our second research question was, ***RQ2*****:** *Why do users trust (or mistrust) Facebook and Google?* To address it, we asked the multiple-choice question, *What makes [Facebook or Google/Gmail] trustworthy?* We used most of the *trust cues* from participants in our first study (Section 3.4.3) regarding how our app can gain their trust and added more choices, for a total of 10 options.

Participants could select multiple responses. The eleventh choice (*Other*) allowed participants to write another reason or comment in response to the question. Figure 4.17 shows the percentage of participants who picked each *trust cue*. The top three trust cues for Facebook were *many people have accounts with them* (39%), *they are a well-known company* (36%), and *they use secure technology to protect my account* (33%). The least cited trust cue for Facebook was *they monitor my account* (12%). The top three trust cues for Google were *they have a good reputation* (72%), *they use secure technology to protect my account* (57%), and *they are a well-known company* (48%). The least cited trust cues for Google were *security experts have accounts with them* and *people I know have recommended them* (both 23%). Overall, Google received a higher percentage of selections in all trust cues except for one, *many people have accounts with them*, in which Facebook received 2% more of the group total than Google.

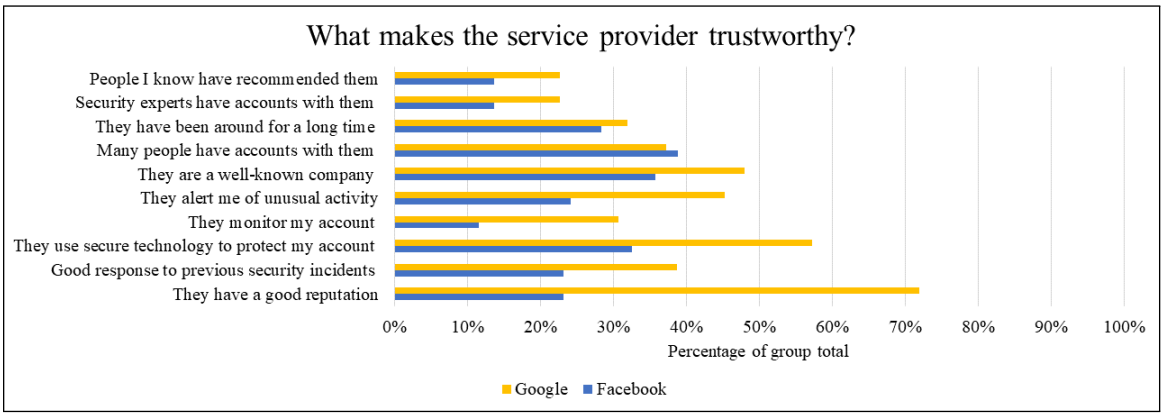


Figure 4.17: Responses to what makes each service provider trustworthy. Multiple responses allowed. Descriptions of responses in the *Other* category are available in *Written Responses*).

Written Responses: Eighteen participants in the Facebook group provided written responses to the *Other* option. They believed that Facebook is not trustworthy (9), were unsure of Facebook's trustworthiness (4), and indicated that their trust in Facebook is declining (1). One participant indicated that none of the choices apply to Facebook. The remaining three comments were, *"no," "no comment,"* and *"Need to protect their reputation."* One participant in the Google group indicated that they do not trust Google.

We also asked participants to pick a Likert response to the following statement, *[Facebook/Google] is able to keep my account safe*. Figure 4.18 depicts the Likert scale responses. 55% of Facebook ($\overline{X}$ Likert score = 3.38, $SD = 1.04$) participants and 89% of Google participants ($\overline{X}$ Likert score = 4.16, $SD = .74$) selected *somewhat* or *strongly agree*. A Mann-Whitney test found a significant difference between the two groups, $U = 5,121$, $z = 5.30$, $p < .001$, $r = .41$. Google participants rated their SP's competence significantly higher than Facebook.

Summary

The trust cues that participants selected differed between Facebook and Google. Overall, Google received a higher percentage of selections in all but one trust cue. 15 participants doubted Facebook's trustworthiness, while only one participant indicated they do not trust Google. Google participants perceived their SP's ability to keep

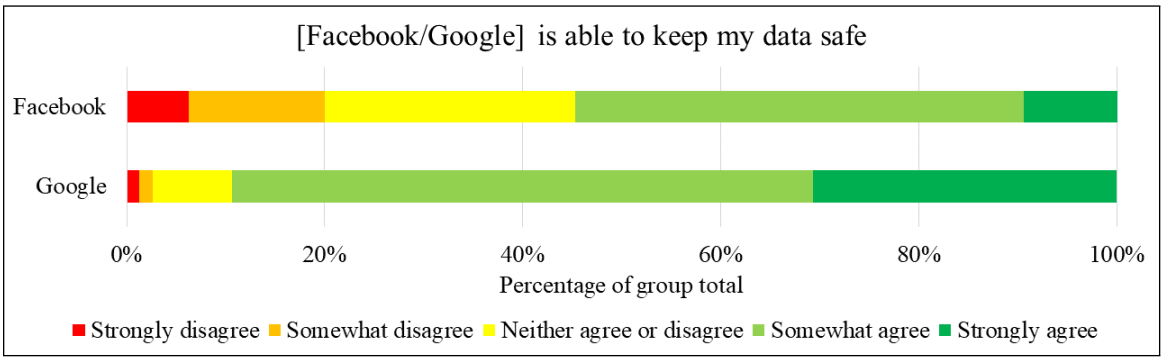


Figure 4.18: Perceived service provider's ability to keep user's account safe.

their accounts safe significantly more than Facebook participants.

4.4.4 Activity Log Effectiveness and Comprehensibility

In the activity log portion of the survey, participants were asked to identify the unusual account access events and explain their selections. This task was followed by three close-ended comprehension questions about the activity log.

Of our valid 170 responses, we excluded the activity log responses of 36 participants because they did not understand what the activity log was (12 participants in the *Diagram* condition, and nine in the *Textlog* condition), mistakenly believed that the activity log was their own (two *Diagram* and five *Textlog*), did not understand the questions (one *Diagram* and three *Textlog*), or misinterpreted the events that are adjacent to one another for being simultaneous (one *Diagram* and three *Textlog*). We used the remaining 134 responses (65 *Diagram* and 69 *Textlog*) for analysis for ***RQ3***.

Effectiveness in Discovering Unusual Events

To test whether participants could identify unusual activity, we asked participants: *Unusual access events are those that may not be from the person who owns these accounts. Based on the activity log below, which account access event(s) is/are the most unusual? Click on the events.* We presented the activity log with invisible hotspots over each of the 63 events. Clicks on these hotspots enabled us to determine which events were selected. Participants were prompted to pick at least one event, and they were allowed to pick as many as they wanted. There was no time limit

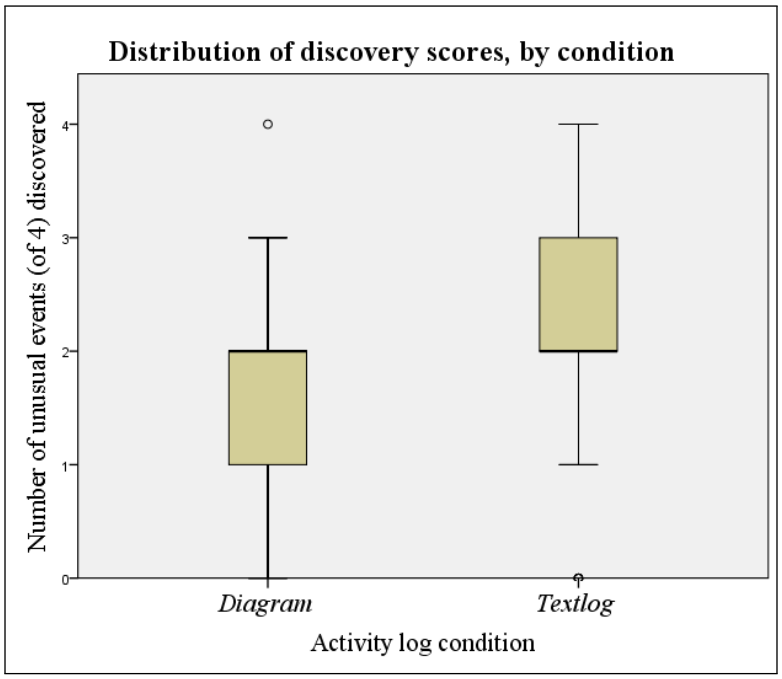


Figure 4.19: Distribution of discovery scores in *Diagram* compared with *Textlog*.

placed on this task. Figures 4.3 and 4.4 show the four events that we designed to be unusual. They were not disclosed in the survey, but are highlighted here for clarity.

We counted how many of the four events each participant discovered. The distribution of discovery scores is depicted in Figure 4.19. Using these discovery scores, a Mann-Whitney test found a significant difference between discovery scores in *Diagram* ($\overline{X} = 1.62$, $SD = .86$) and *Textlog* ($\overline{X} = 2.06$, $SD = 1.01$), $U(134) = 2,838.5$, $z = 2.82$, $p = .005$, $r = .24$. Participants in the *Textlog* condition discovered significantly more unusual events than participants in the *Diagram* condition.

Upon further examination, we find that although *Textlog* performed better overall, one event had a slightly higher discovery rate in *Diagram*: Event 52. Figure 4.20 shows a side-by-side comparison of scores for the four events. It is worth noting that Event 52 is the only unusual event enclosed in a red triangle accompanied by a red exclamation mark in *Diagram*, which is the glyph for an unknown location.

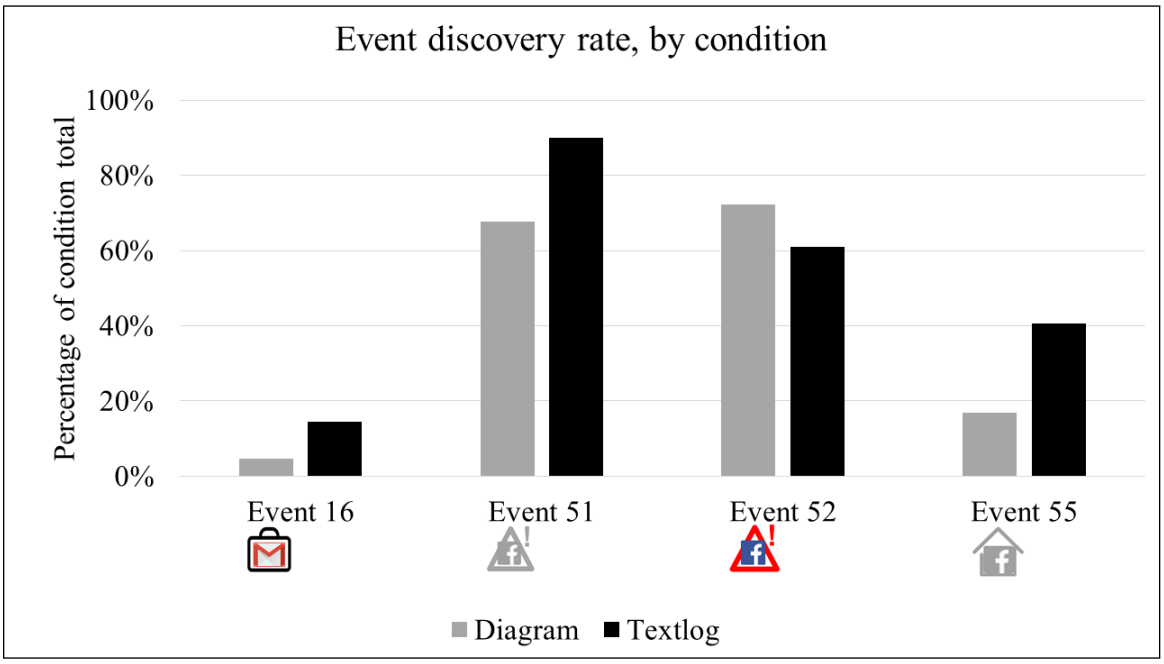


Figure 4.20: Discovery rate of the four unusual events in *Diagram* compared with *Textlog*. The glyphs under each event label appeared only in *Diagram*.

A re-calculation of discovery scores to control for false positives

Although we had designed four events to appear unusual, some participants interpreted other events as unusual. This would mean that a participant who picked all four correct events, in addition to another 10 events, received a perfect discovery score (4 out of 4) as a participant who picked only the four correct events. This is problematic when drawing conclusions about *Textlog* because the higher discovery rate could simply be a result of participants picking more events in *Textlog* than in *Diagram*. To control for false positives, we re-calculated the discovery scores for each activity log using two methods:

1. Using only the scores of participants who picked true positives, ie, they picked at least one of the four unusual events, but no other events. For example, if a participant picked Event 51 and Event 52, she received a discovery score of 2. If another participant picked Event 51 in addition to two other incorrect events, her discovery score was excluded from analysis. The resulting number of participants who picked only true positives were 46 in each activity log condition

Table 4.8: Mann-Whitney test comparing discovery between Diagram and Textlog using the discovery scores of participants who only picked true positives.

Diagram		Textlog	
Mean (SD)	Median	Mean (SD)	Median
1.87 (.72)	2	2.09 (.76)	2
Result:	$U(92) = 1{,}221$, $z = 1.42$, $p = .157$, $r = .15$		

Table 4.9: Mann-Whitney test comparing discovery between Diagram and Textlog using weighted discovery scores.

Diagram		Textlog	
Mean (SD)	Median	Mean (SD)	Median
1.44 (.94)	1.33	1.75 (.95)	2
Result:	$U(134) = 2, 645$, $z = 1.85$, $p = .064$, $r = .16$		

($N = 92$). Although Textlog still had a higher discovery rate, there was no significant difference between the activity logs. Descriptives and results of the Mann-Whitney test are listed in Table 4.8.

2. Using a *weighted discovery score* for every participant. We calculated the weighted discovery score by first deriving a ratio of true positives picked to all events picked. For example, P021 picked a total of 13 events, three of which were true positives. Her ratio would be $3/13 = 0.23$. This means that 23% of the events she picked were true positives. To convert this to a discovery score out of four, we multiplied her number of true positives (3) by her percentage (23%), resulting in a weighted discovery score of 0.7 out of 4. The total number of participants used in this analysis remained the same ($N = 134$). Although Textlog still had a higher discovery rate, there was no significant difference between the activity logs. Descriptives and results of the Mann-Whitney test are listed in Table 4.9.

These results indicate that *Diagram* and *Textlog* performed equally on discovery when we control for false positives.

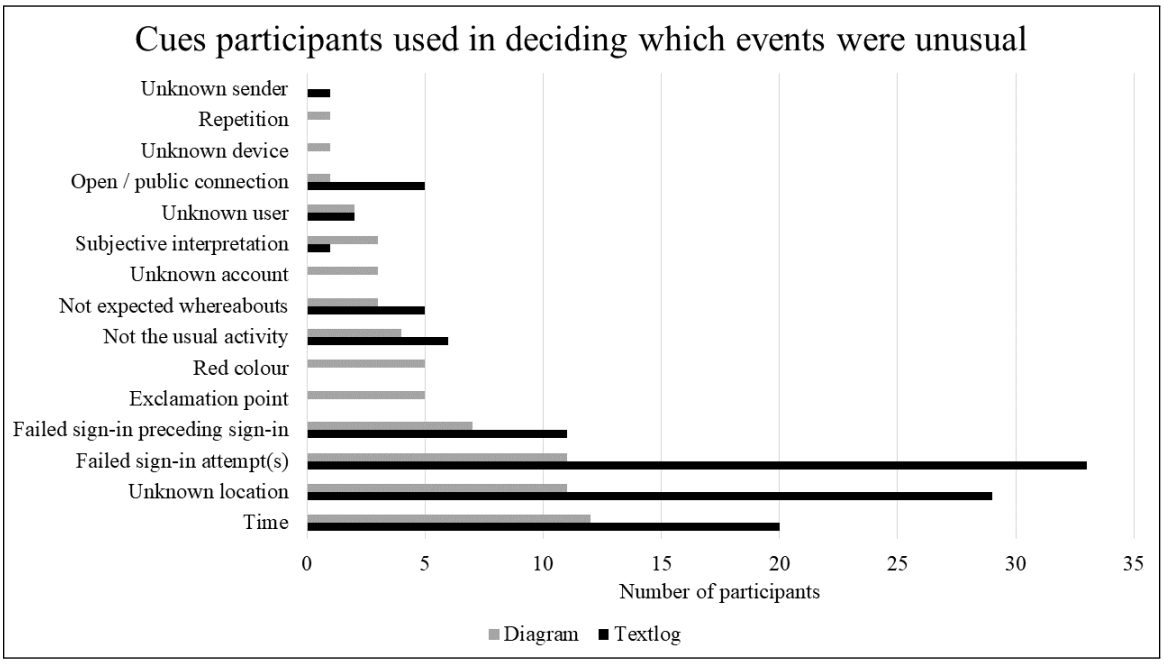


Figure 4.21: Number of participants who used particular cues in deciding which events were unusual in *Diagram* compared with *Textlog*. Within each cue, each participant is represented once. Some participants are represented in multiple cues.

Which cues did participants use in deciding that events are unusual?

After they clicked the unusual events, participants answered the open-ended question, *Why is this event/these events unusual?* Most participants identified more than one cue, and often indicated that it is the combination that made a particular event unusual. The tallied cues are depicted in Figure 4.21.

We found that the top four cues that participants used were reasonable: the timing of the unusual events (12 in *Diagram*, 20 in *Textlog*), the unknown location (11 in *Diagram*, 29 in *Textlog*), the failed sign-in attempt(s) (11 in *Diagram*, 33 in *Textlog*), and the failed sign-in attempt followed by sign-in (7 in *Diagram*, 11 in *Textlog*). These findings are similar to Study 1 (see Table 3.5), in which participants used timing, IP address, and location as the top three cues. More Textlog participants used these cues than Diagram. As expected, some *Diagram* participants used the colour red ($n = 5$) and the exclamation mark ($n = 5$) as an indication of unusual activity. These cues were used in Event 51 and Event 52 (Figure 4.3).

Some participants mentioned that the events simply did not follow the usual

pattern (*Diagram* $n = 4$, *Textlog* $n = 6$). Some participants specified that the unusual activity does not align with the user's expected whereabouts (3 in *Diagram*, 5 in *Textlog*). P194 said, *"He is at home at 8 pm."* P098 explained, *"[the account accessed] late at night at the office [is] usually used during the day."* P036 said, *"at home early morning and home during evening."*

One participant in *Diagram*, and five in *Textlog* deemed the events from the Cafe or Mall to be unusual because they are public places and the internet connection is not secure. Interestingly, three *Diagram* and one *Textlog* participant offered an alternative, subjective interpretation of the unusual events: *"office time, no personal information"* (P116). *"The person has gone into the account too many times at various times during the day. You usually don't go into Gmail that many times"* (P079). *"Facebook at an early time around 6am on a Saturday because you probably are off from working and would sleep in"* (P089). *"Bank [accessed at] home [on] Saturday 9 a.m. it is just unusual"* (P100). Some participants misinterpreted *unknown location* as *"unknown account"* (3 in *Diagram*), *"unknown user"* (2 in *Diagram*, 2 in *Textlog*), *"unknown device"* (1 in *Diagram*), and *"unknown sender"* (1 in *Textlog*).

These results indicate that although many users generally understood the content of both activity logs, there was room for misinterpretation. Overall, *Textlog* participants identified more cues than *Diagram* participants.

Comprehension

We posed three close-ended comprehension questions to the participants in both conditions:

1. **CQ1 (time):** We asked, *Based on the activity log above, what is the time of day when this person has most accessed their Outlook account this week?* to test how well the diagram enables participants to scan for horizontal patterns. Participants could choose from four options: (1) *12am to 6am*, (2) *6am to 12pm noon* (correct answer), (3) *12 pm noon to 6pm*, (4) *6pm to 11:59pm*.

2. **CQ2 (day):** We asked, *Based on the activity log above, which account was the least accessed on Sunday?* to test vertical pattern scanning. Participants could

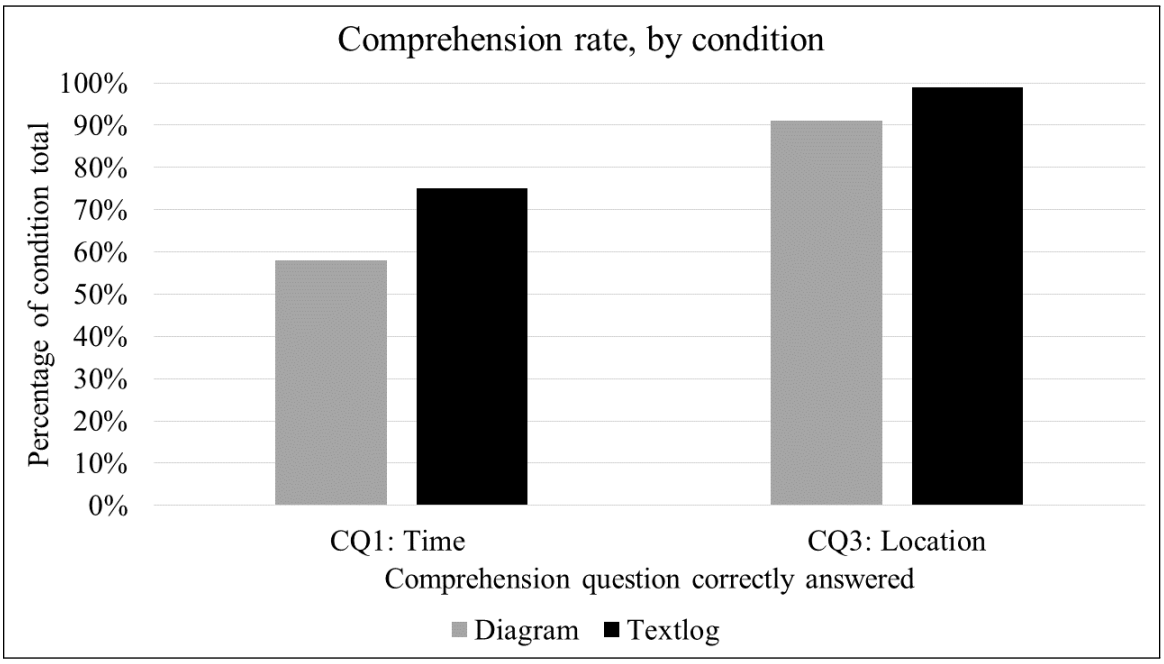


Figure 4.22: Percentage of correct answers to the two comprehension questions in *Diagram* compared with *Textlog*.

choose from five options: (1) *Facebook*, (2) *Outlook*, (3) *Instagram*, (4) *Twitter* (correct answer), (5) *Google/Gmail*.

3. **CQ3 (location):** We asked, *Where does this person usually access their Facebook account?* to test the comprehensibility of the glyphs used in the diagram to represent location. Participants could choose from five options: (1) *Cafe*, (2) *Home* (correct answer), (3) *Office*, (4) *Mall*, (5) *Unknown*.

The wording of the CQ2 is problematic because some people interpreted "least accessed" as having been accessed zero times and picked an account that was not accessed at all on Sunday. As a result, we excluded CQ2 from analysis.

Figure 4.22 shows the percentage of correct answers to the two comprehension questions for *Diagram* ($\overline{X} = 1.49$, $SD = .62$) and *Textlog* ($\overline{X} = 1.74$, $SD = .47$). More *Textlog* participants correctly answered the two comprehension questions than the Diagram participants. CQ3, about location, was correctly answered by more participants. It was surprising to us that many participants incorrectly answered CQ1 about time. It is possible that the use of *am/pm* may have been confusing for

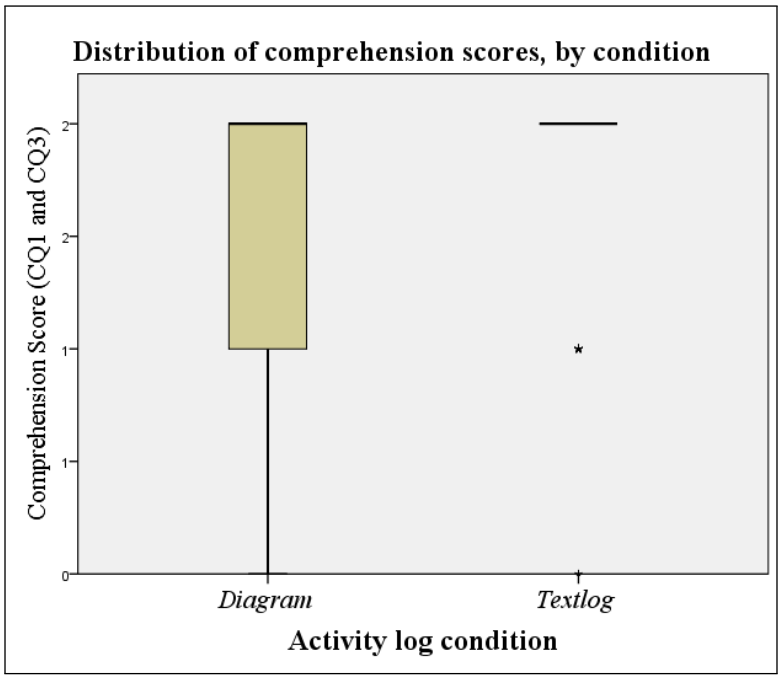


Figure 4.23: Distribution of comprehension scores in *Diagram* compared with *Textlog*.

some participants. For example, in the open-ended question, P002 referred to the early morning failed sign-in attempt as happening between *"5-6 pm"* although he answered CQ1 correctly.

We tabulated how many of the two comprehension questions each participant answered correctly. A distribution of the comprehension scores is illustrated in Figure 4.23. Using these comprehension scores, a Mann-Whitney test found a significant difference between comprehension scores in *Diagram* and *Textlog*, $U = 2,710$, $z = 2.51$, $p = .012$, $r = .22$. Participants in the *Textlog* condition correctly answered more comprehension questions than their *Diagram* counterparts.

Preferences

At the end of the survey, we asked the multiple-choice question, *Would you prefer to check your account activity for each of your online accounts separately, or through a combined activity log like [the one] above?* We included a picture of the activity log from their condition with the question. The answers were roughly equal across both conditions. Figure 4.24 shows the percentage of participants who picked each

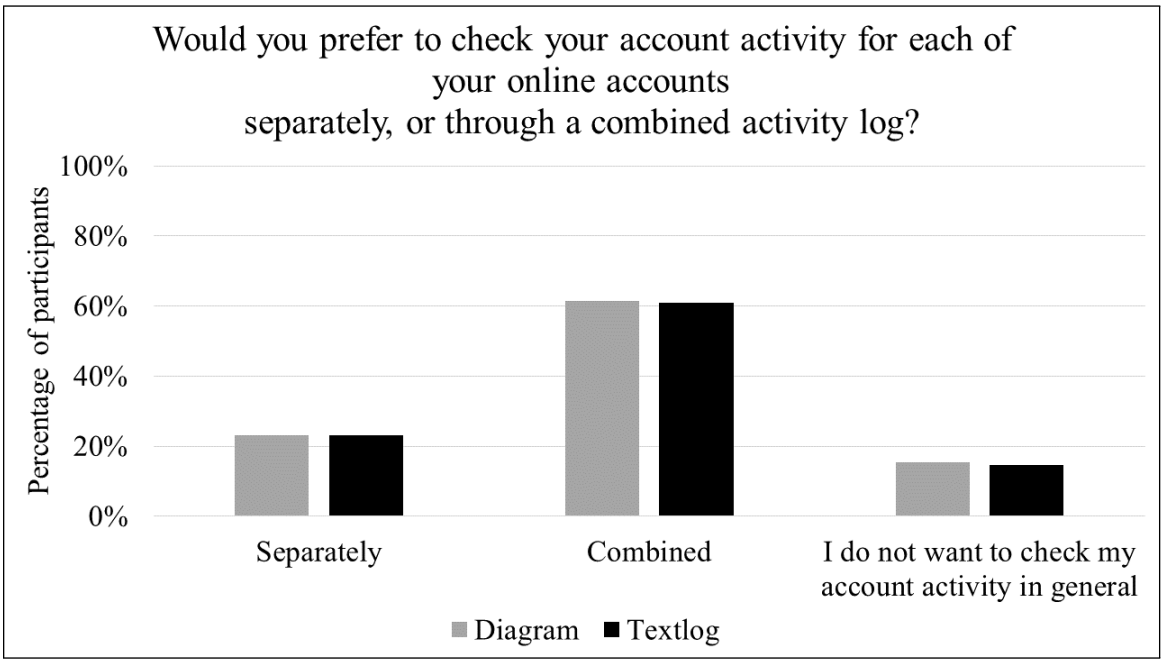


Figure 4.24: Percentage of participants who prefer checking their activity logs separately, in a combined format, or neither in *Diagram* compared with *Textlog*

choice. Most participants prefer checking their activity logs in a combined format, rather than separately.

Participants also rated how confident they are in using the activity logs to identify unusual activity, and how secure the activity log would make them feel. As we can see in Figures 4.25 and 4.26, their answers are roughly equal. There was no significant difference in confidence between *Diagram* ($\overline{X}Likertscore = 3.68$, $SD = 1.03$, $Median = 4$) and *Textlog* ($\overline{X} = 3.77$, $SD = .89$, $Median = 4$), $U = 2,304$, $z = .31$, $p = .756$, $r = .03$. There was no significant difference in the feeling of security between *Diagram* ($\overline{X} = 3.75$, $SD = .81$, $Median = 4$) and *Textlog* ($\overline{X} = 3.59$, $SD = .63$, $Median = 4$), $U = 1,887$, $z = -1.78$, $p = .075$, $r = -0.15$. Most participants in both conditions would feel somewhat or very confident in their ability to identify unusual activity using the activity log. Similarly, most of them would also feel somewhat or very secure.

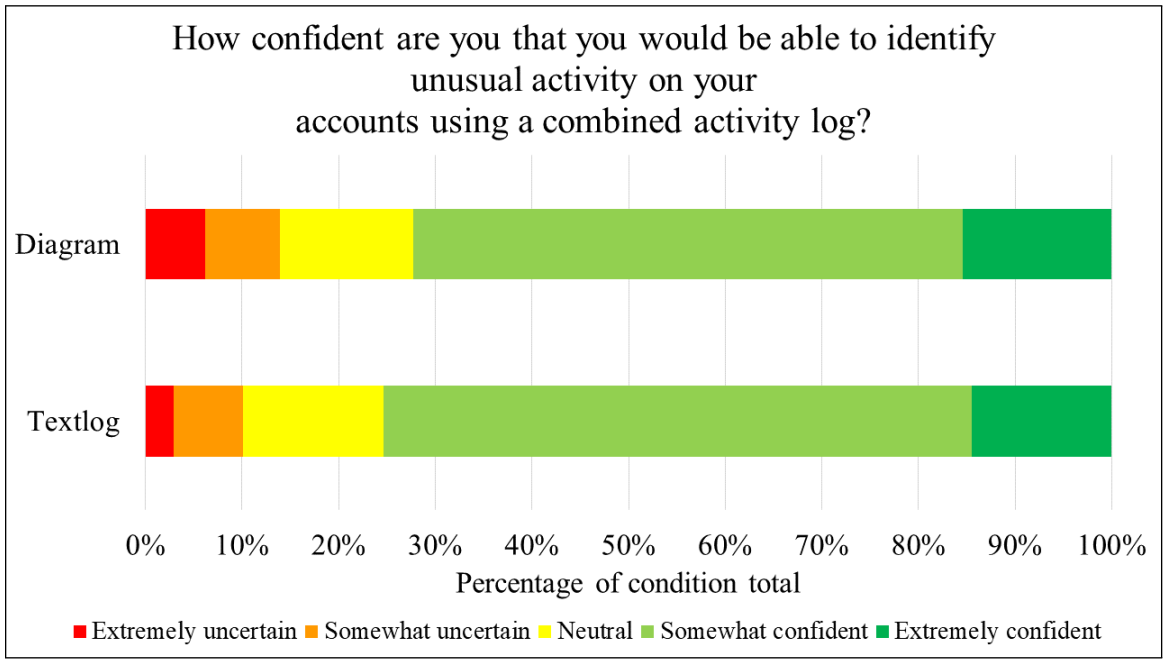


Figure 4.25: Confidence in ability to identify unusual events using the activity log.

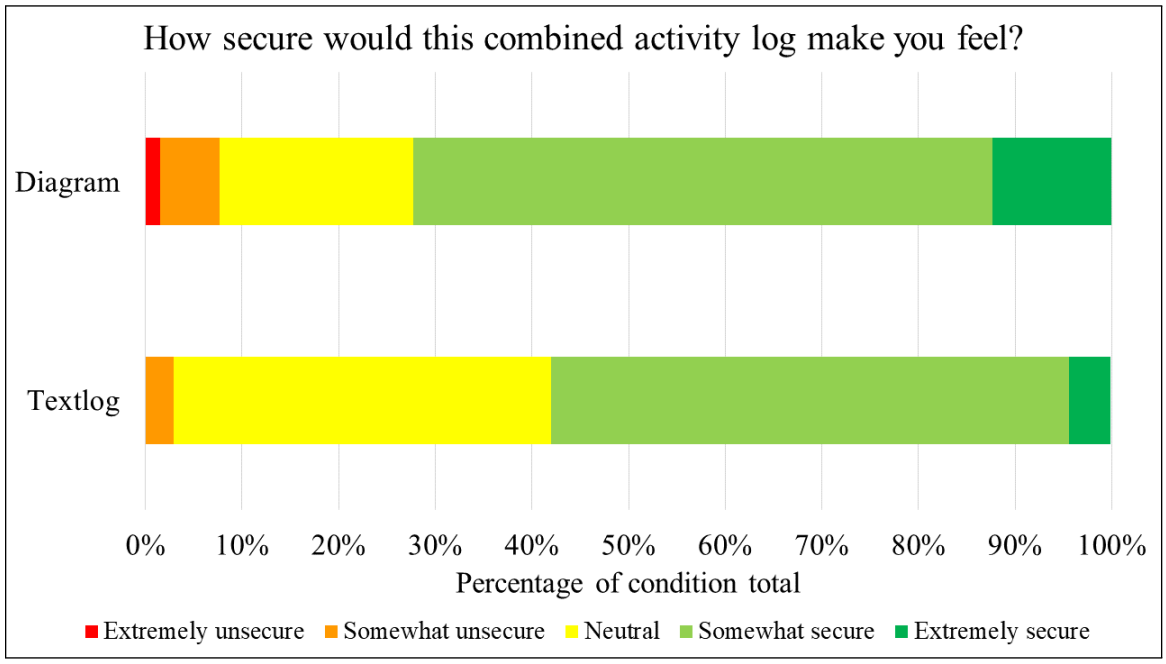


Figure 4.26: Perceived security of the activity log.

Summary

Textlog had a slightly higher discovery rate than *Diagram*. When we control for the selection of false positives, we find that they both performed equally.

Textlog participants correctly answered the comprehension questions significantly more than participants in the *Diagram* condition.

Most participants prefer checking their activity logs in a combined format, rather than separately and felt confident in their ability to identify unusual activity. Similarly, the majority would also feel secure with its use.

4.5 Discussion

RQ1. asked, *Who do end users perceive is responsible for (a) preventing attacks, (b) alerting the user or reporting to the service provider of unusual activity, and (c) for recovering the account after an attack?* We hypothesized that **H1.** *Users will (a) assign responsibility of prevention to Google and Facebook, (b) share the responsibility of alerting/reporting, and (c) assign most of the responsibility to Google and Facebook for recovery.* Our findings partially supported our hypothesis: participants attributed (a) most of the responsibility to themselves for preventing password guessing, and shared responsibility with the SP for preventing phishing, malware, and password stealing, (b) most of the responsibility to the SP for alerting, and (c) most of the responsibility to the SP for recovery. These findings confirm attitudes in our first study. Some participants believed that they should not have to bear the responsibility of monitoring their accounts because it is the service provider's responsibility to alert them of unusual activity.

RQ2. asked, *Why do users trust (or mistrust) Facebook and Google?* We hypothesized that **H2.** *Users trust both Google and Facebook mainly because they are well-known service providers and because they use secure technology.* Contrary to our hypothesis, participants trust Facebook and Google for different reasons. Facebook participants were significantly more concerned about the security of their account than Google participants. We speculate that the recent privacy breach by a third-party app on Facebook [38] played a role in the lower trust ratings. One participant

wrote when asked *What would lead you to delete your Facebook account?*, *"I should actually delete it with all the information that's come out in recent months"* (P035). Another participant responded, *"Unsure, however I have debated deleting it after recent security issues that have been exposed"* (P060). Despite these attitudes, the top two trust cues that Facebook participants selected were, *many people have accounts with them* and *they are a well-known company.* This could indicate that a platform being in existence and widely used for a long time is a powerful reason for people to trust it enough to use it. Alternatively, people could be using Facebook because they do not have alternatives [52], [114], [13].

RQ3. asked, *How does the visual activity log (Diagram) compare with the text activity log (Textlog) in (a) how effective it is in identifying unusual activity, and (b) how comprehensible it is?* We hypothesized that **H3.** *Diagram will be more (a) comprehensible and (b) more effective in identifying unusual activity than Textlog. Textlog* significantly outperformed *Diagram* in comprehension. *Textlog* had a slightly higher discovery rate than Diagram, but the difference between the two was not significant when we controlled for the selection of false positives.

These results are contrary to our hypotheses. Upon reflection, we speculate that *Diagram* did not perform as well as *Textlog* because it violated some Gestalt laws. It may have been difficult for participants to group events by only their location or only their account type because the glyphs appear too detailed. In addition, the colours of the commercial icons were too many, perhaps all competing for the participant's attention. *Diagram* appeared too cluttered overall. Another factor that may have played a role was that participants were not trained on how to use *Diagram* before they were asked the corresponding questions. This could indicate that *Textlog* is a better solution due to a lower learning curve. Alternatively, a small amount of training time may be a good trade-off for potentially better performance using *Diagram.* Such speculations will require further research.

A lesson we learned from the performance of Diagram is that what users say they want may not necessarily be a usable solution. We based our design of *Diagram* on what participants recommended in Study 1, yet found that Textlog performed better.

Due to its neatly formatted text and distinguishable rows, *Textlog* itself is arguably

a visualization because it enables pattern scanning. This is different from existing activity logs which do not display the information so neatly. For Event 52, the discovery rate in *Diagram* was higher than *Textlog* by 12%. Event 52 is the only one that used a red bordering triangle accompanied by a red exclamation mark. It would not be surprising if the colour red contributed to this success rate [89]. Another speculation is that a hybrid activity log that combines effective features from both designs, such as colour and neatly formatted text, could enjoy a higher success rate.

A possible confounding variable in the performance of *Diagram* could be the way that the legend was presented within the visualization. Some participants may have seen the legend as part of the account access events because it was very closely placed next to the events on Sunday. Conversely, the use of line glyphs in *Diagram* had a high success rate in comprehension. Similarly, the use of a red triangle and exclamation mark were accurately understood by most *Diagram* participants as being a potentially dangerous event. These design components can be leveraged further to aid a faster comprehension time. Although the time that was spent on the survey was roughly equal between both activity log conditions, it would be worthwhile to explore efficiency in a future study by accurately measuring the time taken to complete each task.

An important aspect of *Diagram* is scalability. The dataset we visualized contained 63 events on 10 accounts, however, many users have more than 10 online accounts. If *Diagram* were to be implemented in the wild, it would likely become too cluttered. In addition, users who do not have regular access times and whereabouts would not be able to discern a pattern by which to judge unusual events. With bigger datasets, we deem interactivity to be necessary. Users will need zoom, filter, and search functionalities to be able to explore the data. We also expect that users would prioritize their most important accounts for monitoring, instead of all their accounts.

4.6 Limitations

Our survey had methodological limitations. Online surveys are leveraged for the access they provide to a wider population, yet they are known to yield less accurate results than in-lab studies or field studies [95]. In addition, our online survey may

have been better administered with some training time. This is because 21% of the participants did not understand the activity log (mostly *Diagram*) or the questions associated with it.

The design of the visual activity log, *Diagram*, was an improvement of our initial visual log from Study 1, however it needed more robust design iterations before being released in the survey. Two iterations of pilot testing did not reveal that the legend caused confusion, however, this consequence seems likely from the results we observed from the comprehension questions and the open-ended answers. Many participants interpreted *Unknown location* as an unknown user or unknown account. Some participants seemed to be linking the events with the legend. P082 explained her choices of unusual events: *"it appears there was an unusual log in attempt at a cafe, followed by an unknown log in attempt."* If we look at *Diagram*, the word *Cafe* in the legend vertically aligns with the failed sign-in attempt, Event 51.

Chapter 5

Discussion and Conclusion

In this section, we discuss our contributions, how our findings fit into the literature on account security, and the future research directions that have emerged.

5.1 Main Contributions

To re-iterate, our main contributions from this thesis are as follows:

1. We designed, prototyped, and tested a combined activity log tool.
2. We identified external factors contributing to user trust of security tools, and online service providers more generally.
3. We identified a mismatch in perceived responsibility between users and service providers.
4. We provide a look into end users' activity log practices.

5.2 Trust

We find from our studies and others [11,134] that trust is necessary for adoption. The cues that our participants use to derive trust confirm existing literature [8,61]. Some trust cues have to do with *competence* [70]. For example, Aurigemma et al. [11] found that one reason people were unwilling to adopt password managers is the concern for the security of their passwords. As in our first study, their participants were also uncomfortable having all their passwords in one place [11].

Other trust cues are social in nature. For example, we found that the top two reasons people trust Facebook is because it is used by many others and it is a well-known company. These trust cues are arguably part of *transitive trust* [8], but other social factors could also be at play. For example, the wide usage of a platform could

motivate people to join it to avoid being isolated, or simply because there is no alternative [94]. Another trust cue is *reputation* [8], which we found to be the top reason our participants trust Google.

Similar to what we learned in Study 1, Xiao *et al.* [134] found that security tool adoption by developers is highly influenced by "word of mouth" recommendations, which they categorize as a trusted communication channel. When recommendations for a new technology arrive through competent or knowledgeable people, end users are more likely to adopt it. Such socially-derived trust cues are in keeping with the *diffusion of innovations* theory [97]. Reputation takes a long time for a service provider to develop, but early adopters of a technology have an influential role in spreading it via recommendations [97]. In the case of our combined activity log tool, or other emerging security tools, this would imply that a good strategy would be to identify early adopters who can endorse it to potential end users.

Taking what we learned, it is important for researchers and designers to consider the methods that people use to derive trust when evaluating their products or services. For usable security specifically, if we produce a tool that is designed to make people safe, it is imperative that we meet the trust cues [8, 61, 70, 107, 113, 126] necessary for user adoption.

5.3 Mismatch in Perceived Responsibility

Our findings on responsibility indicate a mismatch between user and service provider expectations. Although users believed that they shared responsibility equally with their service provider to prevent some attacks, they held the service provider more responsible for alerting them of unusual account activity and recovering their accounts and identity/reputation after an attack. In contrast, Facebook's terms of service state, *"We make no guarantees that [our products] always will be safe, secure, or error-free [...] Under no circumstance will we be liable to you for any lost profits, revenues, information, or data, or consequential, special, indirect, exemplary, punitive, or incidental damages [...] even if we have been advised of the possibility of such damages"* [37]. Google's terms of service state, *"Google [...] will not be responsible for lost profits, revenues, or data, financial losses or indirect, special, consequential,*

exemplary, or punitive damages" [49]. We chose Facebook and Google due to their popularity, but this mismatch in expectations can extend to other SPs as well [94].

On a practical level, service providers such as Facebook and Google protect end users' accounts [120]. Arguably, it is not in their business interests not to do so. However, when users' expectations are violated, this could lead to a decline in their trust for the service provider. For example, in 2013, end users filed a lawsuit against Google because Gmail scans their emails for advertising purposes [91]. Although Google was not violating the terms of service, it was the mismatch in expectations that lead to those users mistrusting their service provider. In the context of account security, it is better for users to have full awareness of what their service providers are legally liable for rather than rely on protections (e.g., recovery or reimbursement in the aftermath of identity theft) that they may never receive.

The mismatch in perceived responsibility for privacy and security [67] is common. In fact, Rao *et al.* [94] identify different types of mismatches with respect to data practices. As a solution, they suggest a redesign of EULAs by service providers to highlight the commonly misunderstood data policies. Another solution is third-party "privacy decision support tools" which identify the consequences of the EULA that are most relevant to a person, based on their privacy preferences or characteristics. For example, Rao *et al.* find that younger people generally expect that they have the option of deleting their data from a website. A privacy decision support tool would highlight policies that do not allow data deletion [94]. A similar approach for security can also be applied. For example, a security/privacy decision support tool can highlight the consequences of sharing personally identifiable information (i.e., identity theft, for example) and the user's corresponding responsibility (i.e., to recover one's identity, for example).

An implication of this mismatch in perceived responsibility is that it necessitates that users take on a bigger role in reactive security measures, such as account monitoring. This is where our proposed combined activity log tool can be a potential solution. This is because aggregating users' account activity logs in one place is a more usable way of monitoring accounts than having to check 90 [17] different sources. We argue that existing aggregation mechanisms [1, 96] can be leveraged for this purpose.

5.4 Combining Insight from the Two Studies

In Study 1, most participants indicated that they do not check their account activity logs, whereas in Study 2, a higher percentage of participants do. Participants were generally able to understand our first design of the visual activity activity log in Study 1 and our second iteration of it in Study 2. In both studies, participants felt confident in their ability to identify unusual events using our proposed activity logs. In Study 2, they indicated that both activity logs (*Diagram* and *Textlog*) would help them feel secure. Most participants in Study 1 would not feel secure using our *calendar visualization* because they did not trust the app that it was a part of, whereas in Study 2, we presented the activity logs as stand-alone tools, without the context of an app.

In both studies, roughly 60% of our participants indicated their preference to check the account activity logs of their accounts in a combined form. The *trust cues* our participants mentioned as necessary for them to adopt our app extended to the cues participants use to trust Facebook and Google in Study 2. Facebook users trust their SP using a combination of trust cues, whereas Google participants were more likely to trust their SP for its reputation. It is also worth noting that Facebook users are significantly more concerned about account security than Google users. They also have less confidence in Facebook's ability to keep their accounts safe. Perhaps the lower perceived competence and security of Facebook is made up for with a more diverse set of trust cues. In other words, perhaps Facebook users look to trust cues other than competence and security to decide to continue using the platform. This implies that service providers who score lower on certain trust cues can still gain traction if they make up for it with other trust cues.

In Study 1, a portion of our participants allocated responsibility to their service provider for the monitoring of their accounts. These findings held true in Study 2, in which we found that participants allocated responsibility of monitoring (i.e., alerting the user of unusual activity) as well as recovery to the service provider. The only item they assigned primary responsibility to themselves for was to prevent password guessing. This could be because creating strong passwords is a proactive security measure that end users have control over. They assigned primary responsibility to

the SP for alerting and recovery. This is arguably because users are unable to identify failed access attempts, nor do they generally have the technical capability [67] to recover their accounts. In addition, the service provider is the entity who has access to the account activity logs before the end user can even access them.

Most of our participants either check or want to check their activity logs, and 60% prefer having a combined format for doing so. This is a favourable pretext for more research on combined activity log tools. It is also an indication of a proactive attitude towards reactive security measures. Additionally, if account monitoring is something end users have control over like creating strong passwords, the perceived responsibility for it may shift over to the shared category (see the three categories depicted in Figure 4.16). Although users cannot possibly identify failed access attempts before their service provider does, they may be able to identify other types of unusual activity from the logs made available to them. For example, users may be able to identify an access event that occurs from a usual device and location, but is not theirs. Such an event requires human logic that web mining techniques may not be able to recreate.

5.5 Limitations

A methodological limitation of our first study is that we conducted it with simulated data in a lab setting. Schechter *et al.* [100] found that role-playing had a significant negative effect on the security behaviour of the participants using fake login credentials in their study, compared with participants who used real login credentials. This finding relates to our study in which we asked participants to play the role of a user whose account activity they analyzed for unusual events. The *role-playing effect* may have resulted in less rigorous analysis by our participants. This is because motivation to find unusual activity on a fictional person's accounts would not be the same as finding unusual activity on their own accounts. In addition, the data did not have a personal meaning to the participants so it was hard to identify anomalies.

Surveys are known to yield less accurate results than in-lab studies or field studies [95]. In addition, our online survey in Study 2 may have been better administered with some training time for the activity log portion. This is because 21% of the participants did not understand the activity log (mostly *Diagram*) or the questions

associated with it. We also found that the wording of one of our comprehension questions was problematic.

The comprehension and event discovery rates of *Diagram* were lower than expected. In retrospective analysis, we observed potentially confounding variables in the visual design of *Diagram* that may have played a role. The placement of the legend was too close to the visualization, perhaps causing some participants to interpret it as part of the visualization. In addition, the glyphs representing location (*Office* and *Mall*, for example) may have appeared too similar for participants. We also observed that *Diagram* violated some Gestalt laws. We learned that there was a need for additional iterations and pilot testing of *Diagram*.

A limitation of activity logs in general is that IP addresses are problematic to use as a parameter for identifying unusual events. This is because IP addresses change, and the geolocation embedded within them is not always specific enough [12, 115]. Future work is needed to explore how static identifiers can be used for identification of unusual events.

5.6 Future Work

Based on our experience, we suggest future work in the following directions.

First, it would be useful to investigate account activity logs in the wild to assess if they increase reactive security behaviour. Real life behaviours and attitudes over a longer time period may differ from those exhibited in the lab, in particular if the data represents users' real activities.

Secondly, we would like to explore whether consolidation of activity logs across real users' multiple accounts will make usage patterns and anomalies more evident. With fictitious data, it was sometimes difficult for users to assess what was considered 'unusual' because they lacked context surrounding the actions.

We are not advocating better account protection from monitoring that is exclusively user-sided, nor do we expect that users would want to take on this task alone. For these reasons, we would also like to explore the potential of a hybrid solution that leverages both the artificial intelligence of web usage mining as well as human judgment. This would be particularly useful in situation when systems flag false positives

or not flag true positives.

5.7 Conclusion

Although we found that our participants perceived our account monitoring prototype favourably in how easy and convenient it would be to use, we also identified trust cues that impacted adoption. In Study 2, we found that those trust cues extended to the online service providers Facebook and Google. We find that users generally trust Facebook or Google to keep their account safe, yet there is a mismatch in expectations between the two entities. We found that participants identified clear roles with respect to primary responsibility between themselves and their service provider for the security of their accounts.

We provide a look into users' account activity log practices and find that most participants either check or want to check their activity logs. We learned that the main reason why participants do not check their activity logs is that they are unaware that this information exists. We tested our combined activity log designs in Study 2 and find that our participants were generally able to comprehend them, and to a lesser extent, successfully identify some types of unusual events. We found that 60% of our participants prefer using an aggregated method for checking their account activity logs. We highlight the reasons why enabling reactive security behaviours is important for end users, and argue that our proposed concept of combined activity logs has the potential to serve as a tool for that purpose.

Bibliography

[1] Joao Aires and Daniel Gonçalves. Personal information dashboard-me, at a glance. In *PIM 2012 Workshop*, pages 1–8, 2012.

[2] Alexa Internet, Inc. Top sites in Canada. https://www.alexa.com/topsites/countries/CA, May 2017.

[3] Ahmed AlKalbani, Hepu Deng, and Booi Kam. Investigating the role of socio-organizational factors in the information security compliance in organizations. In *Australasian Conference on Information Systems*, 2015.

[4] Mansour Alsaleh, Abdullah Alqahtani, Abdulrahman Alarifi, and AbdulMalik Al-Salman. Visualizing phpids log files for better understanding of web server attacks. In *Proceedings of the Tenth Workshop on Visualization for Cyber Security*, pages 1–8. ACM, 2013.

[5] Mohamed Alsharnouby, Furkan Alaca, and Sonia Chiasson. Why phishing still works: User strategies for combating phishing attacks. *International Journal of Human-Computer Studies (Elsevier)*, 82:69 – 82, 2015.

[6] V Anitha and P Isakki. A survey on predicting user behavior based on web server log files in a web usage mining. In *International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE),*, pages 1–4. IEEE, 2016.

[7] Mohd Anwar and Ashiq Imran. A comparative study of graphical and alphanumeric passwords for mobile device authentication. In *MAICS*, pages 13–18, 2015.

[8] Donovan Artz and Yolanda Gil. A survey of trust in computer science and the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 5(2):58–71, 2007.

[9] Hala Assal, Sonia Chiasson, and Robert Biddle. Cesar: Visual representation of source code vulnerabilities. In *Symposium on Visualization for Cyber Security (VizSec)*. IEEE, 2016.

[10] ATLAS.ti. What is ATLAS.ti? http://atlasti.com/product/what-is-atlas-ti/, May 2017.

[11] Salvatore Aurigemma, Thomas Mattson, and Lori Leonard. So much promise, so little use: What is stopping home end-users from using password manager applications? In *Proceedings of the 50th Hawaii International Conference on System Sciences*, 2017.

[12] Mahesh Balakrishnan, Iqbal Mohomed, and Venugopalan Ramasubramanian. Where's that phone?: Geolocating IP addresses on 3G networks. In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement*, pages 294–300. ACM, 2009.

[13] Eric PS Baumer, Phil Adams, Vera D Khovanskaya, Tony C Liao, Madeline E Smith, Victoria Schwanda Sosik, and Kaiton Williams. Limiting, leaving, and (re) lapsing: an exploration of facebook non-use practices and experiences. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 3257–3266. ACM, 2013.

[14] Anshul Bhargav and Munish Bhargav. Pattern discovery and users classification through web usage mining. In *International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*, pages 632–636. IEEE, 2014.

[15] Rainer Böhme and Stefan Köpsell. Trained to accept?: A field experiment on consent dialogs. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2403–2406. ACM, 2010.

[16] Rita Borgo, Johannes Kehrer, David HS Chung, Eamonn Maguire, Robert S Laramee, Helwig Hauser, Matthew Ward, and Min Chen. Glyph-based visualization: Foundations, design guidelines, techniques and applications. In *Eurographics (STARs)*, pages 39–63, 2013.

[17] Tom Le Bras. Online overload its worse than you thought. *Dashlane blog (online)*, 2015.

[18] Cristian Bravo-Lillo, Saranga Komanduri, Lorrie Faith Cranor, Robert W Reeder, Manya Sleeper, Julie Downs, and Stuart Schechter. Your attention please: designing security decision UIs to make genuine risks harder to ignore. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, page 6. ACM, 2013.

[19] Michael Buhrmester, Tracy Kwang, and Samuel D Gosling. Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data? *Perspectives on psychological science*, 6(1):3–5, 2011.

[20] Paul Cairns and Anna L Cox. *Research methods for human-computer interaction*, volume 12. Cambridge University Press Cambridge, 2008.

[21] Stuart K Card, Jock D Mackinlay, and Ben Shneiderman. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.

[22] John M Carroll. *Making use: scenario-based design of human-computer interactions*. MIT press, 2000.

[23] Sonia Chiasson, Chris Deschamps, Elizabeth Stobert, Max Hlywa, Bruna Freitas Machado, Alain Forget, Nicholas Wright, Gerry Chan, and Robert Biddle. The MVP web-based authentication framework. In *International Conference on Financial Cryptography and Data Security*, pages 16–24. Springer, 2012.

[24] Sonia Chiasson, Paul C van Oorschot, and Robert Biddle. A usability study and critique of two password managers. In *Usenix Security*, volume 6, 2006.

[25] Tom Clarke and Alan Costall. The emotional connotations of color: A qualitative investigation. *Color Research and Application*, 33(5):406–410, 2008.

[26] Mark Coeckelbergh. Virtual moral agency, virtual moral responsibility: on the moral significance of the appearance, perception, and performance of artificial agents. *AI & society*, 24(2):181–189, 2009.

[27] Kari Gwen Coleman. Computing and moral responsibility. *Stanford Encyclodpedia of Philosophy (online)*, 2012.

[28] Jessica Colnago, Summer Devlin, Maggie Oates, Chelse Swoopes, Lujo Bauer, Lorrie Cranor, and Nicolas Christin. "It's not actually that horrible": Exploring adoption of two-factor authentication at a university. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 456. ACM, 2018.

[29] Robert Cooley, Bamshad Mobasher, and Jaideep Srivastava. Web mining: Information and pattern discovery on the world wide web. In *Ninth IEEE International Conference on Tools with Artificial Intelligence*, pages 558–567. IEEE, 1997.

[30] Kenneth C Cox, Stephen G Eick, Graham J Wills, and Ronald J Brachman. Brief application description; visual data mining: Recognizing telephone calling fraud. *Data Mining and Knowledge Discovery*, 1(2):225–231, 1997.

[31] Alexander De Luca, Alina Hang, Emanuel Von Zezschwitz, and Heinrich Hussmann. I feel like I'm taking selfies all day!: Towards understanding biometric authentication on smartphones. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1411–1414. ACM, 2015.

[32] Pierre Dragicevic and Stéphane Huot. Spiraclock: a continuous and non-intrusive display for upcoming events. In *CHI'02 extended abstracts on Human factors in computing systems*, pages 604–605. ACM, 2002.

[33] Lisa Eadicicco. Why Google+ failed, according to Google insiders. *Business Insider (online)*, 2015.

[34] Edison Research. Infinite dial 2018. https://www.slideshare.net/webby2001/infinite-dial-2018, 2018.

[35] Niklas Elmqvist and Philippas Tsigas. Trustneighborhoods in a nutshell. In *Proceedings of the 2006 ACM Symposium on Software Visualization*, SoftVis '06, pages 189–190, New York, NY, USA, 2006. ACM.

[36] Federico Michele Facca and Pier Luca Lanzi. Recent developments in web usage mining research. In *International Conference on Data Warehousing and Knowledge Discovery*, pages 140–150. Springer, 2003.

[37] Facebook Inc. Terms of service. https://www.facebook.com/legal/terms, 2018.

[38] Facebook Inc. An update on our plans to restrict data access on Facebook. https://newsroom.fb.com/news/2018/04/restricting-data-access/, 2018.

[39] David Feng, Yueh Lee, Lester Kwock, and Russell M Taylor II. Evaluation of glyph-based multivariate scalar volume visualization techniques. In *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization*, pages 61–68. ACM, 2009.

[40] Andy Field. *Discovering statistics using IBM SPSS statistics*. sage, 2013.

[41] John M Findlay and Iain D Gilchrist. Eye guidance and visual search. In *Eye guidance in reading and scene perception*, pages 295–312. Elsevier, 1998.

[42] N FitzGerald. New facebook worm — don't click da'button baby. https://www.geek.com/news/facebook-worm-wants-you-to-click-da-button-baby-992052/, 2009.

[43] Ivan Flechais, Jens Riegelsberger, and M Angela Sasse. Divide and conquer: the role of trust and assurance in the design of secure socio-technical systems. In *Proceedings of the 2005 workshop on New security paradigms*, pages 33–41. ACM, 2005.

[44] Alain Forget, Sonia Chiasson, Paul C van Oorschot, and Robert Biddle. Improving text passwords through persuasion. In *Proceedings of the 4th symposium on Usable privacy and security*, pages 1–12. ACM, 2008.

[45] Siwei Fu, Jian Zhao, Hao Fei Cheng, Haiyi Zhu, and Jennifer Marlow. T-cal: Understanding team conversational data with calendar-based visualization. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 500. ACM, 2018.

[46] Johannes Fuchs, Fabian Fischer, Florian Mansmann, Enrico Bertini, and Petra Isenberg. Evaluation of alternative glyph designs for time series data in a small multiple setting. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 3237–3246. ACM, 2013.

[47] Hongyu Gao, Jun Hu, Tuo Huang, Jingnan Wang, and Yan Chen. Security issues in online social networks. *IEEE Internet Computing*, 15(4):56–63, 2011.

[48] David Gefen. Reflections on the dimensions of trust and trustworthiness among online consumers. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, 33(3):38–53, 2002.

[49] Google. Terms of service. https://policies.google.com/terms, 2018.

[50] Ashika Gupta, Rakhi Arora, Ranjana Sikarwar, and Neha Saxena. Web usage mining using improved frequent pattern tree algorithms. In *Issues and Challenges in Intelligent Computing Techniques (ICICT), 2014 International Conference on*, pages 573–578. IEEE, 2014.

[51] Penny Hagen, Toni Robertson, Melanie Kan, and Kirsten Sadler. Emerging research methods for understanding mobile technology use. In *Proceedings of the 17th Australia conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future*, pages 1–10. Computer-Human Interaction Special Interest Group (CHISIG) of Australia, 2005.

[52] Antti Hakkala, Olli I Heimo, Sami Hyrynsalmi, and Kai K Kimppa. Security, privacy'); drop table users;-and forced trust in the information age?: when trusting an information system is not optional and why it matters. *ACM SIGCAS Computers and Society*, 47(4):68–80, 2018.

[53] Sun Hao, Shen Zhaoxiang, and Zhang Bingbing. A user clustering algorithm on web usage mining. In *Electronics Instrumentation & Information Systems (EIIS), 2017 First International Conference on*, pages 1–4. IEEE, 2017.

[54] Marian Harbach, Emanuel Von Zezschwitz, Andreas Fichtner, Alexander De Luca, and Matthew Smith. It's a hard lock life: A field study of smartphone (un) locking behavior and risk perception. In *Symposium on Usable Privacy and Security (SOUPS)*, pages 213–230, 2014.

[55] Eiji Hayashi, Oriana Riva, Karin Strauss, AJ Brush, and Stuart Schechter. Goldilocks and the two mobile devices: going beyond all-or-nothing access to a device's applications. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, page 2. ACM, 2012.

[56] Larue Tone Hosmer. Trust: The connecting link between organizational theory and philosophical ethics. *Academy of management Review*, 20(2):379–403, 1995.

[57] Facebook Inc.

[58] Renáta Iváncsy and István Vajk. Frequent pattern mining in web log data. *Acta Polytechnica Hungarica*, 3(1):77–90, 2006.

[59] Audun Josang. Trust-based decision making for electronic transactions. In *Proceedings of the Fourth Nordic Workshop on Secure Computer Systems (NORDSEC99)*, pages 496–502, 1999.

[60] Audun Josang and Roslan Ismail. The beta reputation system. In *Proceedings of the 15th Bled electronic commerce conference*, volume 5, pages 2502–2511, 2002.

[61] Audun Jøsang, Roslan Ismail, and Colin Boyd. A survey of trust and reputation systems for online service provision. *Decision support systems*, 43(2):618–644, 2007.

[62] Priit Kallas. Top 15 most popular social networking sites and apps. *Dreamgrow (online)*, 2018.

[63] Patrick Gage Kelley. Conducting usable privacy and security studies with Amazons Mechanical Turk. In *Symposium on Usable Privacy and Security (SOUPS)*, 2010.

[64] Raymond Kosala and Hendrik Blockeel. Web mining research: A survey. *ACM Sigkdd Explorations Newsletter*, 2(1):1–15, 2000.

[65] Kat Krol, Eleni Philippou, Emiliano De Cristofaro, and M Angela Sasse. "They brought in the horrible key ring thing!" Analysing the usability of two-factor authentication in uk online banking. *arXiv preprint arXiv:1501.04434*, 2015.

[66] Michelle Kwasny, Kelly Caine, Wendy A Rogers, and Arthur D Fisk. Privacy and technology: folk definitions and perspectives. In *CHI'08 Extended Abstracts on Human Factors in Computing Systems*, pages 3291–3296. ACM, 2008.

[67] Mohammad Mannan and Paul C van Oorschot. Security and usability: the gap in real-world online banking. In *Proceedings of the 2007 Workshop on New Security Paradigms*, pages 1–14. ACM, 2008.

[68] Amy Martin and Wendy Ju. Bloom: an interactive, organic visualization of starred emails. In *ACM SIGGRAPH 2010 Posters*, page 33. ACM, 2010.

[69] Lucas Matney. Google has 2 billion users on Android, 500M on Google photos. https://techcrunch.com/2017/05/17/google-has-2-billion-users-on-android-500m-on-google-photos/, 2017.

[70] Roger C Mayer, James H Davis, and F David Schoorman. An integrative model of organizational trust. *Academy of management review*, 20(3):709–734, 1995.

[71] Daniel McCarney, David Barrera, Jeremy Clark, Sonia Chiasson, and P. C. van Oorschot. Tapas: design, implementation, and usability evaluation of a password manager. In *Annual Computer Security Applications Conference (ACSAC)*, pages 89–98, 2012.

[72] Aleecia M McDonald and Lorrie Faith Cranor. The cost of reading privacy policies. *ISJLP*, 4:543, 2008.

[73] Christine Mekhail, Leah Zhang-Kennedy, and Sonia Chiasson. Visualizations to teach about mobile online privacy. In *International Conference on Persuasive Technology*, pages 43–47, 2014.

[74] Todd Miller and John Stasko. Infocanvas: A highly personalized, elegant awareness display. In *Supporting Elegant Peripheral Awareness, workshop at CHI03*, 2003.

[75] Miniwatts Marketing Group. World internet users and 2018 population stats. https://www.internetworldstats.com/stats.htm, 2018.

[76] Chris Muelder, Kwan-Liu Ma, and Tony Bartoletti. Interactive visualization for network and port scan detection. In *International Workshop on Recent Advances in Intrusion Detection*, pages 265–283. Springer, 2005.

[77] Lik Mui, Mojdeh Mohtashemi, and Cheewee Ang. A probabilistic rating framework for pervasive computing environments. In *Proceedings of the MIT Student Oxygen Workshop (SOW2001)*, 2001.

[78] Lik Mui, Mojdeh Mohtashemi, Cheewee Ang, Peter Szolovits, and Ari Halberstadt. Ratings in distributed systems: A bayesian approach. In *Proceedings of the Workshop on Information Technologies and Systems (WITS)*, pages 1–7, 2001.

[79] Lik Mui, Mojdeh Mohtashemi, and Ari Halberstadt. A computational model of trust and reputation. In *System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on*, pages 2431–2439. IEEE, 2002.

[80] Lik Mui, Mojdeh Mohtashemi, and Ari Halberstadt. Notions of reputation in multi-agents systems: A review. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1*, pages 280–287. ACM, 2002.

[81] Tsuyoshi Murata and Kota Saito. Extracting users' interests from web log data. In *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*, pages 343–346. IEEE Computer Society, 2006.

[82] Vinayak Musale and Devendra Chaudhari. Web usage mining tool by integrating sequential pattern mining with graph theory. In *Intelligent Systems and Information Management (ICISIM), 2017 1st International Conference on*, pages 160–163. IEEE, 2017.

[83] Guillaume Nadon, Marcus Feilberg, Mathias Johansen, and Irina Shklovski. In the user we trust: Unrealistic expectations of Facebook's privacy mechanisms. In *Proceedings of the 9th International Conference on Social Media and Society*, pages 138–149. ACM, 2018.

[84] Shahnaz Parvin Nina, Mahmudur Rahman, Khairul Islam Bhuiyan, and Khandakar Entenam Unayes Ahmed. Pattern discovery of web usage mining. In *International Conference on Computer Technology and Development*, volume 1, pages 499–503. IEEE, 2009.

[85] Helen Nissenbaum. Accountability in a computerized society. *Human values and the design of computer technology*, pages 41–64, 1997.

[86] Jonathan A Obar. Big data and the phantom public: Walter Lippmann and the fallacy of data privacy self-management. *Big Data & Society*, 2(2):2053951715608876, 2015.

[87] Kenneth Olmstead and Aaron Smith. Americans and cybersecurity. *Pew Research Center*, 26, 2017.

[88] Laura Portwood-Stacer. Media refusal and conspicuous non-consumption: The performative and political dimensions of Facebook abstention. *New Media & Society*, 15(7):1041–1057, 2013.

[89] Karyn Pravossoudovitch, Francois Cury, Steve G Young, and Andrew J Elliot. Is red the colour of danger? testing an implicit red–danger association. *Ergonomics*, 57(4):503–510, 2014.

[90] Sören Preibusch. Guide to measuring privacy concern: Review of survey and observational instruments. *International Journal of Human-Computer Studies*, 71(12):1133–1143, 2013.

[91] Associated Press. Google defends its right to scan Gmail accounts as outraged privacy advocates seek to end practice: Google's attorneys say the practice that helps the company sell ads is legal, and are asking a federal judge to dismiss a lawsuit that seeks to stop the practice. *Financial Post (online)*, 2013.

[92] Rinto Priambodo and Riri Satria. User behavior pattern of mobile online social network service. In *International Conference on Cloud Computing and Social Networking (ICCCSN)*, pages 1–4. IEEE, 2012.

[93] Ariel Rabkin. Personal knowledge questions for fallback authentication: Security questions in the era of Facebook. In *Proceedings of the 4th symposium on Usable privacy and security*, pages 13–23. ACM, 2008.

[94] Ashwini Rao, Florian Schaub, Norman Sadeh, Alessandro Acquisti, and Ruogu Kang. Expecting the unexpected: Understanding mismatched privacy expectations online. In *Symposium on Usable Privacy and Security (SOUPS)*, volume 4, page 2, 2016.

[95] Elissa M Redmiles, Yasemin Acar, Sascha Fahl, and Michelle L Mazurek. A summary of survey methodology best practices for security and privacy researchers. *UMD*, 2017.

[96] Tiago Rodrigues, Prateek Dewan, Ponnurangam Kumaraguru, Raquel Melo Minardi, and Virgílio Almeida. uTrack: track yourself! monitoring information on online social media. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 273–276. ACM, 2013.

[97] Everett M Rogers. Diffusion of innovations: modifications of a model for telecommunications. In *Die diffusion von innovationen in der telekommunikation*, pages 25–38. Springer, 1995.

[98] Yvonne Rogers, Helen Sharp, and Jenny Preece. *Interaction design: beyond human-computer interaction*. John Wiley & Sons, 2011.

[99] Heather Rosoff, Jinshu Cui, and Richard S John. Behavioral experiments exploring victims' response to cyber-based financial fraud and identity theft scenario simulations. In *Symposium on Usable Privacy and Security (SOUPS)*, pages 175–186, 2014.

[100] Stuart E Schechter, Rachna Dhamija, Andy Ozment, and Ian Fischer. The emperor's new security indicators. In *Security and Privacy, 2007. SP'07. IEEE Symposium on*, pages 51–65. IEEE, 2007.

[101] Panda Security. Facebook survey: More than 50% of users don't trust the social network to control the news. *Panda Security Mediacenter (online)*, 2008.

[102] Panda Security. Data leak might have exposed the details of more than 120 million Facebook users. *Panda Security Mediacenter (online)*, 2018.

[103] Murli Manohar Sharma and Anju Bala. An approach for frequent access pattern identification in web usage mining. In *Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference on*, pages 730–735. IEEE, 2014.

[104] Richard Shay, Iulia Ion, Robert W Reeder, and Sunny Consolvo. My religious aunt asked why I was trying to sell her viagra: experiences with account hijacking. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 2657–2666. ACM, 2014.

[105] Steve Sheng, Mandy Holbrook, Ponnurangam Kumaraguru, Lorrie Faith Cranor, and Julie Downs. Who falls for phish?: A demographic analysis of phishing susceptibility and effectiveness of interventions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 373–382. ACM, 2010.

[106] Hossein Siadati, Bahador Saket, and Nasir Memon. Detecting malicious logins in enterprise networks using visualization. In *Visualization for Cyber Security (VizSec), 2016 IEEE Symposium on*, pages 1–8. IEEE, 2016.

[107] Elizabeth Sillence, Pam Briggs, Lesley Fishwick, and Peter Harris. Trust and mistrust of online health sites. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 663–670. ACM, 2004.

[108] Judith Simon. Distributed epistemic responsibility in a hyperconnected era. In *The Onlife Manifesto*, pages 145–159. Springer, 2015.

[109] Judith Simon and Irina Shklovski. Lessening the burden of individualized responsibility in the socio-technical world. *Proceedings of ISIS Summit-The Information Society at the Crossroads*, pages 1–5, 2015.

[110] P. C. van Oorschot Sonia Chiasson, Robert Biddle. Materials for a usability study of password managers. *SOUPS*, 2006.

[111] Andreas Sotirakopoulos, Kirstie Hawkey, and Konstantin Beznosov. On the challenges in usable security lab studies: Lessons learned from replicating a study on SSL warnings. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*, page 3. ACM, 2011.

[112] Thomas Steiner. Bots vs. Wikipedians, anons vs. logged-ins. *CoRR*, abs/1402.0412, 2014.

[113] R Todd Stephens. A framework for the identification of electronic commerce design elements that enable trust within the small hotel industry. In *Proceedings of the 42nd annual Southeast regional conference*, pages 309–314. ACM, 2004.

[114] Stefan Stieger, Christoph Burger, Manuel Bohn, and Martin Voracek. Who commits virtual identity suicide? Differences in privacy concerns, internet addiction, and personality between facebook users and quitters. *Cyberpsychology, Behavior, and Social Networking*, 16(9):629–634, 2013.

[115] Mingxuan Sun, Guangyue Xu, Junjie Zhang, and Dae Wook Kim. Tracking you through DNS traffic: Linking user sessions by clustering with dirichlet mixture model. In *Proceedings of the 20th ACM International Conference on Modelling, Analysis and Simulation of Wireless and Mobile Systems*, pages 303–310. ACM, 2017.

[116] KR Suneetha and Raghuraman Krishnamoorthi. Identifying user behavior by analyzing web server access log file. *IJCSNS International Journal of Computer Science and Network Security*, 9(4):327–332, 2009.

[117] Techopedia. Tooltip. www.techopedia.com/definition/5482/tooltip, 2018.

[118] Soon Tee Teoh, Kwan-Liu Ma, S Felix Wu, and Xiaoliang Zhao. A visual technique for internet anomaly detection. In *IASTED International Conference on Computer Graphics and Imaging*. Citeseer, 2002.

[119] The Harris Poll. Annual reputation rankings for the 100 most visible companies in the U.S. https://theharrispoll.com/the-harris-pollr-today-released-its-17th-annual-reputation-quotientr-rqr-summary-report-revealing-corporate-reputation-ratings-for-the-100-most-visible-companies-in-the-u-s-as-perceived-by/, 2018.

[120] Kurt Thomas, Frank Li, Ali Zand, Jacob Barrett, Juri Ranieri, Luca Invernizzi, Yarik Markov, Oxana Comanescu, Vijay Eranti, Angelika Moscicki, et al. Data breaches, phishing, or malware?: Understanding the risks of stolen credentials. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 1421–1434. ACM, 2017.

[121] Maarten Van Dantzich, Daniel Robbins, Eric Horvitz, and Mary Czerwinski. Scope: Providing awareness of multiple notifications at a glance. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, pages 267–281. ACM, 2002.

[122] Fernanda B Viégas, Danah Boyd, David H Nguyen, Jeffrey Potter, and Judith Donath. Digital artifacts for remembering and storytelling: Posthistory and social network fragments. In *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*, pages 10–pp. IEEE, 2004.

[123] Fernanda B Viégas, Scott Golder, and Judith Donath. Visualizing email content: portraying relationships from conversational histories. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 979–988. ACM, 2006.

[124] Fernanda B Viégas and Marc Smith. Newsgroup crowds and authorlines: Visualizing the activity of individuals in conversational cyberspaces. In *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*, pages 10–pp. IEEE, 2004.

[125] T Franklin Waddell, Joshua R Auriemma, and S Shyam Sundar. Make it simple, or force users to read?: Paraphrased design improves comprehension of end user license agreements. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 5252–5256. ACM, 2016.

[126] Omar Abdel Wahab, Jamal Bentahar, Hadi Otrok, and Azzam Mourad. A survey on trust and reputation models for web services: Single, composite, and communities. *Decision Support Systems*, 74:121–134, 2015.

[127] Ding Wang and Ping Wang. On the usability of two-factor authentication. In *International Conference on Security and Privacy in Communication Systems*, pages 141–150. Springer, 2014.

[128] Rui Wang, Shuo Chen, and XiaoFeng Wang. Signing me onto your accounts through Facebook and Google: A traffic-guided security study of commercially deployed single-sign-on web services. In *Security and Privacy (SP), 2012 IEEE Symposium on*, pages 365–379. IEEE, 2012.

[129] Colin Ware. *Information visualization: perception for design*. Elsevier, 2012.

[130] Rick Wash and Emilee J Rader. Too much knowledge? security beliefs and protective behaviors among United States Internet users. In *Symposium on Usable Privacy and Security (SOUPS)*, pages 309–325, 2015.

[131] Alan F Westin. Social and political dimensions of privacy. *Journal of social issues*, 59(2):431–453, 2003.

[132] Andrew Whitby, Audun Jøsang, and Jadwiga Indulska. Filtering out unfair ratings in Bayesian reputation systems. In *Proc. 7th Int. Workshop on Trust in Agent Societies*, volume 6, pages 106–117, 2004.

[133] Allison Woodruff, Vasyl Pihur, Sunny Consolvo, Lauren Schmidt, Laura Brandimarte, and Alessandro Acquisti. Would a privacy fundamentalist sell their dna for $1000... if nothing bad happened as a result? The Westin categories, behavioral intentions, and consequences. In *Symposium on Usable Privacy and Security (SOUPS)*, volume 5, page 1, 2014.

[134] Shundan Xiao, Jim Witschey, and Emerson Murphy-Hill. Social influences on secure development tool adoption: why security tools spread. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 1095–1106. ACM, 2014.

[135] Wei Xu, Fangfang Zhang, and Sencun Zhu. Toward worm detection in online social networks. In *Proceedings of the 26th Annual Computer Security Applications Conference*, pages 11–20. ACM, 2010.

[136] Zhi Yang, Christo Wilson, Xiao Wang, Tingting Gao, Ben Y Zhao, and Yafei Dai. Uncovering social network sybils in the wild. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 8(1):2, 2014.

[137] Leah Zhang-Kennedy, Sonia Chiasson, and Robert Biddle. Password advice shouldn't be boring: Visualizing password guessing attacks. In *Anti-Phishing Working Group (APWG) eCrime Researchers Summit*, pages 1–11. IEEE, 2013.

[138] Leah Zhang-Kennedy, Sonia Chiasson, and Robert Biddle. Stop clicking on 'update later': Persuading users they need up-to-date antivirus protection. In *International Conference on Persuasive Technology*, pages 302–322. Springer LNCS, 2014.

[139] Leah Zhang-Kennedy, Sonia Chiasson, and Robert Biddle. The role of instructional design in persuasion: A comics approach for improving cyber security. *International Journal of Human-Computer Interaction (IJHCI)*, 32:215–257, 2016.

[140] Leah Zhang-Kennedy, Sonia Chiasson, and P. C. van Oorschot. Revisiting password rules: Facilitating human management of passwords. In *APWG eCrime*. IEEE, 2016.

Appendix A

Study 1: Recruitment Poster, Consent Form, Pre-Test Questionnaire, Semi-Structured Interview Script and Usability Test, Post-Test Questionnaire

Recruitment Poster

Carleton UNIVERSITY
Canada's Capital University

Try out a new security app!

Study Title: Account Security App
Supervising Professor: Dr. Sonia Chiasson

To be eligible, you must be:

- at least 18 years old
- able to read and speak English fluently
- a regular Internet user

What will participants do?

- complete online questionnaires
- use a mobile app
- participate in an interview which may be audio-recorded.

How long is the session?

- One 60-minute session
- $10 compensation

Any risks involved?
There are no risks associated with this study.

What is the research about?
This study aims to explore whether visual representations of online account activity logs within a mobile app can help end users better identify threats to the accounts and security breaches. The account activity logs used in this study are fictional, and no personal information about account activity will be collected.

For more information or to book a session:
Please contact the researcher at **yomna.abdelaziz@carleton.ca**
The ethics protocol for this project has been reviewed and cleared by CUREB-B, Clearance # 106745: Account Security App. If you have any ethical concerns with the study, please contact Dr. Andy Adler, Chair, Carleton University Research Ethics Board-B (by phone at 613-520-2600 ext. 4085 or via email at ethics@carleton.ca).

Consent Form

Title: Account Security App

Funding Source: NSERC Discovery Grant

Date of ethics clearance: May 11, 2017; CUREB-B Clearance # 106745

Ethics Clearance for the Collection of Data Expires: May 31, 2018

I _____, choose to participate in a study on online account security. This study aims to explore whether visual representations of online account activity logs within a mobile app can help end users better identify threats to the accounts and security breaches. The account activity logs used in this study are fictional, and no personal information about account activity will be collected. **The researcher for this study is Yomna Abdelaziz in the School of Computer Science, Carleton University.** She is working under the supervision of Dr. Sonia Chiasson in the School of Computer Science.

I will be asked to complete online questionnaires, use a mobile app, and participate in an interview which may be audio-recorded. There are no risks associated with this study. This study involves one 60 minute session and I will receive $10 for my time.

To be eligible, **I** must be:

- at least 18 years old
- able to read and speak English fluently
- a regular Internet user

The company running the online questionnaires is Qualtrics, which stores collected data on secured servers in a secured data centre based in the USA. No names or IP addresses will be linked to the data provided. Once copied from the survey-hosting server, the data will be deleted from the server. It will then be stored on the researcher's password-protected computer. The prototype app will record your interactions, such as where I clicked.

I will then be interviewed about my experience with the app and my own opinions regarding account security. If I consent, the interview will be audio-recorded for transcription. The audio-recordings will not be used for any other purposes and once the interview is transcribed the audio-recording will be deleted. Transcripts of the interview will be analyzed by the researchers. If **I** do not consent to be audio-recorded, the researcher will record **my** answers by hand writing notes on paper and typing notes into a computer. All research data will be anonymized and kept on a password-protected computer. Any paper copies of data (including consent forms) will be kept in a locked cabinet at Carleton University. Research data will only be accessible by the researcher and the research supervisor.

I may skip any questions in the questionnaires or interview that I do not feel comfortable answering.

I have the right to end my participation in the study at any time during the session, for any

reason; I will just tell the researcher that I want to end the study. If I withdraw from the study, all information I have provided will be immediately destroyed. However, I understand that withdrawal is not possible after the completion of the study. I will be compensated $10 for my time, even if I withdraw from the study.

I understand that once the project is completed, electronic research data will be kept and potentially used for other research projects on this same topic. Paper materials will be shredded after one year. Anonymized results may be used in publications or presentations.

The ethics protocol for this project was reviewed by the Carleton University Research Ethics Board, which provided clearance to carry out the research, CUREB-B Clearance # 106745. If I have any ethical concerns with the study, I may contact Dr. Andy Adler, Chair, Carleton University Research Ethics Board-B (by phone at 613-520-2600 ext. 4085 or via email at ethics@carleton.ca).

Researcher contact information:
Yomna Abdelaziz
School of Computer Science
Carleton University
Email: yomna.abdelaziz@carleton.ca

Supervisor contact information:
Sonia Chiasson
School of Computer Science
Carleton University
Email: chiasson@scs.carleton.ca
Tel: (613) 520-2600 ext. 1656

Do you agree to be audio recorded? ___ Yes ___No

_____ _____

Signature of participant Date

_____ _____

Signature of researcher Date

Questionnaire

Demographics

Thank you for participating in our Account Security App study. Please answer the following questions to the best of your ability. You may skip any questions you are not comfortable answering.

1. Please specify your age.

2. Gender:
   - Male
   - Female
   - Other
   - Prefer not to answer

3. Completed level of education:

4. Please specify your degree if you are currently in a post-secondary program:

5. Current occupation or job:

6. Do you, or have you ever worked in a computer, computer security, or information technology (IT)-related field?
   - Yes
   - No

7. [if *yes* is selected:] Please specify what type of computer, computer security, or information technology (IT)-related job(s) you have held in the past or currently hold.

Pre-Test

1. Name the top 5 online accounts that are the most sensitive to you or that you want to protect the most.

	Account Name	Why do you want to protect it?
Account #1	-	-
Account #2	-	-
Account #3	-	-
Account #4	-	-
Account #5	-	-

2. Which of the following online account provider(s) do you currently have an account with, and how often do you access them?

Online Account	No account	Access less than a few times per year	Access a few times per year	Access Monthly	Access Weekly	Access daily
Carleton account/ email	o	o	o	o	o	o
Google/Gmail	o	o	o	o	o	o
Youtube.com	o	o	o	o	o	o
Facebook.com	o	o	o	o	o	o
Reddit.com	o	o	o	o	o	o
Wikipedia.org	o	o	o	o	o	o
Amazon.ca	o	o	o	o	o	o
Twitter.com	o	o	o	o	o	o
Yahoo.com	o	o	o	o	o	o
Netflix.com	o	o	o	o	o	o
Imgur.com	o	o	o	o	o	o
Kijiji.ca	o	o	o	o	o	o
Instagram.com	o	o	o	o	o	o
Diply.com	o	o	o	o	o	o
Linkedin.com	o	o	o	o	o	o
Twitch.tv	o	o	o	o	o	o
Online banking (TD, Scotiabank, RBC, etc)	o	o	o	o	o	o
Cbc.ca	o	o	o	o	o	o
Tumblr.com	o	o	o	o	o	o
outlook.live.com / outlook.office.com	o	o	o	o	o	o
Wikia.com	o	o	o	o	o	o
Ebay.ca	o	o	o	o	o	o
Pinterest.com	o	o	o	o	o	o
Craigslist.ca	o	o	o	o	o	o
Wordpress.com	o	o	o	o	o	o
Ebay.com	o	o	o	o	o	o
Stackoverflow.com	o	o	o	o	o	o
Cra-arc.gc.ca	o	o	o	o	o	o
Apple.com / iTunes / AppStore	o	o	o	o	o	o
Github.com	o	o	o	o	o	o
Indeed.com	o	o	o	o	o	o
WhatsApp	o	o	o	o	o	o
Travel reservation website (e.g. Hotwire, AirCanada, etc)	o	o	o	o	o	o

3. How concerned are you about the security of each previously specified account?

	Extremely unconcerned	Somewhat unconcerned	Neither unconcerned nor concerned	Somewhat concerned	Extremely concerned
Account #1	o	o	o	o	o
Account #2	o	o	o	o	o
Account #3	o	o	o	o	o
Account #4	o	o	o	o	o
Account #5	o	o	o	o	o

4. Within the last year, how many times have you received a security alert from your online account provider about a login event or potential risk to your online account? (For example, an attempted login from an unrecognized device.)

	Number of alerts within the past year (approximately)
Account #1	-
Account #2	-
Account #3	-
Account #4	-
Account #5	-

5. Tell me about a time when one of your accounts was compromised. What happened and how did you deal with it?

6. How often do you check the activity log for your online account(s)? The activity log is a list of events related to your account such as usage history, log-ins, and attempted logins.

	Never: I don't think it exists	Never: I know it exists, but I don't want to check it	Never: I know it exists, but I don't know how to access it	Less than a few times per year	A few times per year	Monthly	Weekly	Daily
Account #1	o	o	o	o	o	o	o	o
Account #2	o	o	o	o	o	o	o	o
Account #3	o	o	o	o	o	o	o	o
Account #4	o	o	o	o	o	o	o	o
Account #5	o	o	o	o	o	o	o	o

7. How quick is it to access the activity log of your online account?

	Extremely slow	Somewhat slow	Neither slow nor quick	Somewhat quick	Extremely quick
Account #1	o	o	o	o	o
Account #2	o	o	o	o	o
Account #3	o	o	o	o	o
Account #4	o	o	o	o	o
Account #5	o	o	o	o	o

8. How easy is it to understand the activity log of your online account?

	Extremely difficult	Somewhat difficult	Neither easy nor difficult	Somewhat easy	Extremely easy
Account #1	o	o	o	o	o
Account #2	o	o	o	o	o
Account #3	o	o	o	o	o
Account #4	o	o	o	o	o
Account #5	o	o	o	o	o

Semi-Structured Interview Script and Usability Test

Thank you for agreeing to participate in my study. The session consists of three parts: The first part is a questionnaire to help me get your opinion about account security and to give me ideas about how you keep your own accounts safe. I will not be asking any private information about your accounts themselves. The second part is when I'll show you the mobile app and ask you to use it from the perspective of a fictional user. I will also interview you to ask you questions about the app. The third and final part consists of a short questionnaire. Please note that we are testing the app, not your performance. There are no right or wrong answers to any of the questions that you'll be asked, and your feedback is very valuable to us.

Before we begin, please read and sign this consent form first.

[*After signing consent form*:] We will start with a questionnaire. If any question is unclear to you, let me know. When part 1 is complete, you will see a message on the screen, so let me know when that happens.

[*After questionnaire is complete*:] I will now show you the mobile app. This is a prototype of a new mobile app designed to help people monitor their accounts. In the app, you will see the account activity of a fictional user named Jane Doe. Here is Jane Doe's profile. I want you to read it because it is relevant to her account activity that you will see in the app.

[*After they read profile*:] I want you to do two things: (1) try the app, and (2) look at Jane Doe's account activity in the app. So, I want you to pretend that Jane Doe is signing up for this app and that it's her online account activity that you will be looking at. Please think aloud as you go through the app. I will now start the voice recorder.

[*At the account activity visualization:*]
This is a visualization of Jane's activity during the week. Each letter represents an account, and you can tap each letter to see more information about the particular event. As you can see the account listed along the bottom, Jane has registered 14 accounts in this app. Each colour represents an IP address that Jane has saved as a particular location. For example, blue represents Jane's home, and green represents Carleton University.

Keeping Jane Doe's profile in mind, have you spotted any unusual or suspicious events? [*The researcher verbally asked these questions while the participant can see the visualization on the app, and instructed them to think aloud while she noted their answers*.]

And why does that event look suspicious to you?

[*At the password change detection screen*:] What would you do if you saw this? What does it mean?

[*After using the app*:]

- What are the pros and cons of using an account security app like this one?
- How likely are you to use an app like this one?
- How easy was this app to use? Did you find any parts confusing?
- Which parts of the app are most useful?

- Which information is more important to you in deciding whether or not a particular event is suspicious; the device information, the IP address, the location, or something else?
- Do you have any comments about the way it looks?
- Are there any other things you want this app to do?
- Would you trust this app to access all your accounts? If not, what would it take for this app to win your trust?
- [*For parents*] Would this app be useful as a parental monitoring tool?

Thank you for taking the time to try the app. For the last part of our session, I will ask you to go through a short questionnaire about your experience using the app.

[*After post-test questionnaire is complete*:] Thank you once again for your time, I really appreciate it. Please accept this payment as compensation, and sign this receipt.

Post-Test

1. In the future, how often do you intend on checking the activity log for each online account?

	Never	A few times per year	Monthly	Weekly	Daily
Account #1	o	o	o	o	o
Account #2	o	o	o	o	o
Account #3	o	o	o	o	o
Account #4	o	o	o	o	o
Account #5	o	o	o	o	o

2. Why would you check or not check your account activity log(s)?

3. Would you prefer to check the account activity logs for each account separately, or would you prefer to check the activity logs of your accounts through a centralized app like the one you tested?
   - Separately
   - Combined
   - I do not want to check my account activity in general
   - Other: [*open text field*]

4. Why? Please explain.

5. How easy was it to understand the account activity using the mobile app?
   - Extremely difficult
   - Somewhat difficult
   - Neither easy nor difficult
   - Somewhat easy
   - Extremely easy

6. How fast was it to understand the account activity using the mobile app?
   - Extremely slow
   - Somewhat slow
   - Average
   - Somewhat fast
   - Extremely fast

7. How confident are you that you would be able to identify suspicious behaviour on your accounts using the mobile app?
   - Extremely uncertain
   - Somewhat uncertain
   - Neutral
   - Somewhat confident
   - Extremely confident

8. How secure would the mobile app make you feel?
    - Extremely unsecure
    - Somewhat unsecure
    - Neutral
    - Somewhat secure
    - Extremely secure

9. Please provide any additional comments you have about our study or the topic of online account security.

Appendix B

Study 1: Qualitative Content Analysis Codes

Conventional Content Analysis of Interview Transcripts: List of High-Level Codes

1. Attitudes and practices (56 sub-codes): The attitudes, mental models, and practices that participants reported around online account security.
2. Drawbacks of the app ("cons") (14 sub-codes): The functions and features, or lack thereof, that participants believed made the app a poor candidate for adoption.
3. Desired changes to the app (33 sub-codes): What participants indicated would make the app a better candidate for use, or increase their likelihood of adoption.
4. Ease of use (2 sub-codes): Whether or not participants felt the app was easy to use.
5. Events missed (4 sub-codes): The unusual events that participants misattributed as benign and any reasons they mentioned for their misattributions.
6. Events spotted (18 sub-codes): The unusual events that participants attributed as such and their reasons for doing so.
7. Future design (38 sub-codes): Any improvements of the app alluded to by participants that could inspire future work or a later design of the app.
8. Future study (4 sub-codes): The limitations of the study that became apparent throughout testing. These were noted for the purpose of improving potential future studies.
9. General observations (contains 5 sub-codes): Includes notes on the physical interaction with the tablet computer, and relevant behaviour of participants during the session.
10. Information used for deciding suspiciousness of events (9 sub-codes): The cues and indications that participants used to decide whether or not an event was unusual or suspicious to them.
11. Likelihood of using the app (4 sub-codes): How likely participants were to use Account Sentinel if it were a real app.
12. Misconceptions (6 sub-codes): Misconceptions that participants had of how the app works.
13. Parental monitoring (7 sub-codes): The attitudes around potentially using the app as a parental monitoring tool.
14. Strengths of the app ("pros") (8 sub-codes): What the participants believed was good about Account Sentinel.
15. Password change detection screen (9 sub-codes): How participants perceived the password change detection screen and how they would respond to it.
16. Trust (35 sub-codes): The reasons participants gave for whether or not they would trust Account Sentinel with monitoring their accounts, and the attitudes around trusting mobile apps in general.
17. Useful features (16 sub-codes): The features that participants found useful to them, and the reasons for those beliefs.

Appendix C

Study 2: Recruitment Notice, Consent Form, and Online Survey

Recruitment Notice

The following notice was used by Qualtrics to recruit participants, except where otherwise indicated.

CUREB clearance #: 108976

Do you have opinions on your account security and what companies are doing to keep your data safe? We want to hear from you!

Our research group at the School of Computer Science at Carleton University is conducting a research study of online account monitoring. During a 15-minute online confidential survey, participants will be asked questions relating to account monitoring (e.g., identifying unusual sign-ins) and how they make the decision to trust their online account providers. Participants will also be asked for feedback on a report related to account monitoring. Your responses will be anonymous. Data collected during your session will be associated with an anonymous pseudonym that has no connection with any personally identifiable data.

[*For participants we recruited from social media*:] If you are interested in participating, please e-mail Yomna Abdelaziz at: yomna.abdelaziz@carleton.ca

To be eligible, participants should:

- Be over 18 years old
- Have accounts with Google/Gmail and Facebook
- Regularly use their Google/Gmail and Facebook accounts
- Currently live in Canada or the United States.

[*For participants we recruited from social media*:] Participation is entirely voluntary and there is no compensation.

Participation is entirely voluntary. You will be informed prior to the start of the survey how you will be compensated. You will be compensated the amount you agreed upon before you entered into the survey. You will only receive compensation if you finish the survey.

Ethical Review:
This research has been reviewed and cleared by Carleton University Research Ethics Board (CUREB-B).
Date of Clearance: June 6, 2018
Ethics Clearance for the Collection of Data Expires: June 30, 2019
CUREB contact information: ethics@carleton.ca

Consent Form

The following consent form was used by Qualtrics in the survey, except where otherwise indicated. Depending on which group participants were assigned to, they saw either Facebook or Google/Gmail under "participation criteria."

Title: A survey investigating online account monitoring
Funding Source: NSERC Discovery Grant
Carleton University Research Ethics Board (CUREB) clearance #: 108976
Date of ethics clearance: June 6, 2018
Ethics Clearance for the Collection of Data Expires: June 30, 2019

During a 15-minute online confidential survey, you will be asked questions relating to account monitoring (e.g., identifying unusual sign-ins) and how you make the decision to trust online account providers. You will also be asked for feedback on reports(s) related to account monitoring. Your responses will be anonymous.

The researchers for this study are Prof. Sonia Chiasson and Yomna Abdelaziz (Master's Student) in the School of Computer Science, Carleton University.

Task: The research study involves filling out a survey to provide:
-some demographic information;
-your opinions on account monitoring;
-your opinions on how you trust your account service providers, and;
-your feedback on report(s) related to account monitoring

[*For participants we recruited from social media*:]
Compensation: The survey is voluntary and there is no compensation.

Compensation: You will be compensated the amount you agreed upon before you entered into the survey. You will only receive compensation if you finish the survey.

Participation criteria: Participants must be over 18 years old, must have accounts with [Facebook *or* Google/Gmail] and must regularly use these accounts. Participants must currently reside in Canada or the United States.

Risks: There are no known risks associated with this study. Data collected during your session will be associated with an anonymous pseudonym that has no connection with any personally identifiable data. The survey is being run by Qualtrics. All responses will be confidential. Qualtrics will not collect participants' IP addresses. After data collection is complete, data will be stored on a password-protected computer, associated with the participant's anonymous pseudonym. Only researchers directly involved in the research will have access to the study data.

[*For participants we recruited from social media*:]
Withdrawal: Your participation is voluntary and you have the right to withdraw from the survey at any time, for any reason, up until you hit the "submit" button. You can withdraw by closing the webpage

containing the survey. If you withdraw from the study, your data will be deleted and not used for analysis. As the survey responses are anonymous, it is not possible to withdraw after the survey is submitted.

Withdrawal: You have the right to withdraw from the survey at any time, for any reason, up until you hit the "submit" button. You can withdraw by closing the webpage containing the survey. If you withdraw from the study, your data will be deleted and not used for analysis. As the survey responses are anonymous, it is not possible to withdraw after the survey is submitted. Payment will not be issued if you withdraw from the study.

Data use: The researchers may access the data collected through the survey. Upon project completion, all research data will be kept in password protected format so that it may be compared to the results of other research related to this same topic. Results of the research may be used in research publications or for teaching purposes.

Clearance: The ethics protocol for this project was reviewed by the Carleton University Research Ethics Board (CUREB-B), which provided clearance to carry out the research.

If you have any ethical concerns with the study, please contact:
CUREB contact information:
Professor Andy Adler, Chair (CUREB-B)
Carleton University Research Ethics Board
Carleton University
511 Tory
1125 Colonel By Drive
Ottawa, ON K1S 5B6
Tel: 613-520-2600 ext 4085
ethics@carleton.ca

Researchers' contact information:
Yomna Abdelaziz
School of Computer Science
Carleton University
1125 Colonel By Drive
Ottawa, Ontario K1S 5B6
Email: yomna.abdelaziz@carleton.ca

Prof. Sonia Chiasson
School of Computer Science
Carleton University
1125 Colonel By Drive
Ottawa, Ontario K1S 5B6
Email: chiasson@scs.carleton.ca

Survey Questions

Screening Question

1. Which account do you use regularly?
    - Google/Gmail
    - Facebook
    - Both Google/Gmail and Facebook

Demographic Questions

1. Gender:
    - Male
    - Female
    - Other
    - Prefer not to answer

2. Please specify your age in years.

3. What is your current occupation or job?

4. Do you, or have you ever worked in a computer, computer security, or information technology (IT)-related field?
    - Yes
    - No

5. [if *yes* is selected:] Please specify what type of computer, computer security, or information technology (IT)-related job(s) you have held in the past or currently hold.

Part 1: Security Questions

The following questions appeared to participants in both groups, Facebook and Google/Gmail. Depending on which group participants were assigned to, they were asked about their account domain, as indicated by the square brackets, [Facebook *or* Google/Gmail].

1. How concerned are you about the security of your [Facebook *or* Google/Gmail] account?
   - Extremely unconcerned
   - Somewhat unconcerned
   - Neither unconcerned nor concerned
   - Somewhat concerned
   - Extremely concerned

2. Has your Facebook account ever been compromised? Compromised means that someone has gained access to your account without your permission.
   - Yes
   - No
   - Not sure

3. [*If yes is selected*:] What happened and how did you find out? (If you do not have a comment, write "no comment".)

4. Within the last year, how many times (approximately) have you received a security alert from [Facebook *or* Google/Gmail] about a sign-in event or potential risk to your [Facebook *or* Google/Gmail] account? (For example, an attempted sign-in from an unrecognized device.)
   - 0
   - 1 to 5 alerts
   - 6 to 10 alerts
   - 11 to 15 alerts
   - More than 15 alerts

5. How often do you check the history of recent sign-ins for your [Facebook *or* Google/Gmail] account?
   - Never: I don't think this information exists
   - Never: I know this information exists, but I don't want to check it
   - Never: I know this information exists, but I don't know how to check it
   - Less than a few times per year
   - A few times per year
   - Monthly
   - Weekly
   - Daily

Depending on what participants selected in Q5:

6. Assuming that the history of your recent sign-ins is available, would you check this information?
   - Yes
   - No

- Other: [*open text field*]

7. Why would you not check the history of recent sign-ins of your Facebook account? Check all that apply.
   - Checking my sign-in history will not keep my account safe.
   - Checking my sign-in history is too much work.
   - I don't need to check my sign-in history because I would know if someone has hacked into my account.
   - If my account gets hacked, I will just create a new one.
   - Other: [*open text field*]

8. If you knew how to access the history of your recent sign-ins, would you check this information?
   - Yes
   - No
   - Other: [*open text field*]

9. How easy is it to understand the history of recent sign-ins of your Facebook account?
   - Extremely difficult
   - Somewhat difficult
   - Neither easy or difficult
   - Somewhat easy
   - Extremely easy

10. To what extent do you believe [Facebook *or* Google/Gmail] is responsible for preventing each of the following attacks to your [Facebook *or* Google/Gmail] account?

*Phishing is the attempt to obtain sensitive information such as usernames, passwords, and credit card details (and money), often for malicious reasons, by disguising as a trustworthy entity.

	[Facebook *or* Google/Gmail] is not at all responsible	Somewhat not responsible	Neither	Somewhat responsible	[Facebook *or* Google/Gmail] is completely responsible
Preventing you from falling victim to a phishing* attack	O	O	O	O	O
Preventing your data or device(s) from getting infected with malware through your account	O	O	O	O	O
Preventing your password from being stolen (being hacked)	O	O	O	O	O
Preventing your password from being guessed by an attacker	O	O	O	O	O

11. To what extent do you believe that you are responsible for preventing each of the following attacks to your [Facebook *or* Google/Gmail] account?

*Phishing is the attempt to obtain sensitive information such as usernames, passwords, and credit card details (and money), often for malicious reasons, by disguising as a trustworthy entity.

	I am not at all responsible	Somewhat not responsible	Neither	Somewhat responsible	I am completely responsible
Preventing you from falling victim to a phishing* attack	O	O	O	O	O
Preventing your data or device(s) from getting infected with malware through your account	O	O	O	O	O
Preventing your password from being stolen (being hacked)	O	O	O	O	O
Preventing your password from being guessed by an attacker	O	O	O	O	O

12. To what extent do you believe that [Facebook *or* Google/Gmail] is responsible for alerting you of the following unusual activities on your [Facebook *or* Google/Gmail] account?

	[Facebook *or* Google/Gmail] is not at all responsible	Somewhat not responsible	Neither	Somewhat responsible	[Facebook *or* Google/Gmail] is completely responsible
A sign-in that is not from you	O	O	O	O	O
A failed sign-in attempt not from you	O	O	O	O	O
Account use not from you, [for example, a post on your Facebook account that you did not post *or* an email sent from your Google/Gmail account that you did not send.]	O	O	O	O	O

13. I would like [Facebook *or* Google/Gmail] to determine whether the account activity is from me by comparing it to:
    - my own profile and activity.
    - how other people use their accounts.
    - general patterns (for example, [Facebook *or* Google/Gmail] may consider the activity as unusual if there are three failed sign-ins within a 1-hour time period).
    - I don't want [Facebook *or* Google/Gmail] to do this kind of analysis at all.
    - Other: [*open text field*]

14. To what extent do you believe you are responsible for reporting to [Facebook *or* Google/Gmail] the following unusual activities that you notice on your [Facebook *or* Google/Gmail] account?

	I am not at all responsible	Somewhat not responsible	Neither	Somewhat responsible	I am completely responsible
A sign-in that is not from you	O	O	O	O	O
A failed sign-in attempt not from you	O	O	O	O	O
Account use not from you, [for example, a post on your Facebook account that you did not post *or* an email sent from your Google/Gmail account that you did not send.]	O	O	O	O	O

15. To what extent do you believe that [Facebook *or* Google/Gmail] is responsible for the following recovery efforts in the event of an attack to your [Facebook *or* Google/Gmail] account?

	[Facebook *or* Google/Gmail] is not at all responsible	Somewhat not responsible	Neither	Somewhat responsible	[Facebook *or* Google/Gmail] is completely responsible
Stopping the attacker's access to your account	O	O	O	O	O
Restoring rightful access to your account	O	O	O	O	O
Recovering your identity or reputation after an attack	O	O	O	O	O

16. To what extent do you believe that you are responsible for the following recovery efforts in the event of an attack to your [Facebook *or* Google/Gmail] account?

	I am not at all responsible	Somewhat not responsible	Neither	Somewhat responsible	I am completely responsible
Stopping the attacker's access to your account	O	O	O	O	O
Restoring rightful access to your account	O	O	O	O	O
Recovering your identity or reputation after an attack	O	O	O	O	O

17. [***Quality check question****; choices were randomized. The correct answer depended on which group participants were assigned to, either Facebook or Google/Gmail.*] What account is this survey about?
    - Instagram
    - Google/Gmail
    - Dropbox
    - Wordpress
    - Facebook
    - PayPal

18. List any other entities that you believe are responsible for preventing attacks to your [Facebook *or* Google/Gmail] account, alerting you of unusual activity on your [Facebook *or* Google/Gmail] account, or recovering your [Facebook *or* Google/Gmail] account after an attack. (If you have no comment, write "no comment.")

19. What makes [Facebook *or* Google/Gmail] trustworthy? Check all that apply.
    - They have a good reputation
    - They have had a good response to previous security incidents
    - They use secure technology to protect my account
    - They monitor my account
    - They alert me when there is unusual activity on my account
    - They are a well-known company
    - Many people have accounts with them
    - They have been around for a long time
    - Security experts have accounts with them
    - People I know have recommended them
    - Other: [*open text field*]

20. What would lead you to delete your [Facebook *or* Google/Gmail] account? Please explain.

21. [Facebook *or* Google/Gmail] is able to keep my data safe. (Your data is information about you, your account, and how you use your account.)
    - Strongly disagree
    - Somewhat disagree
    - Neither agree or disagree
    - Somewhat agree
    - Strongly agree

Part 2: Activity Log Questions

The following questions appeared to participants in both conditions, Diagram and Textlog. Depending on which condition participants were assigned to, they were asked about the specific activity log, as indicated within the square brackets.

1. The following [diagram] is a combined activity log that illustrates a fictional person's account access events. Please answer the following questions about the combined activity log [diagram]. Unusual account access events are those that may not be from the person who owns these accounts. Based on the activity log below, which account access event(s) is/are the most unusual? Click on the event(s) in the [diagram *or* combined activity log].

[*Activity log appears here.*]

2. Why is this event/these events unusual? Please explain.

3. [*Activity log inserted before question*.] Based on the activity log above, what is the time of day when this person has most accessed their Outlook account this week?
    - 12 am to 6 am
    - 6 am to 12 pm noon
    - 12 pm noon to 6 pm
    - 6 pm to 11:59 pm

4. [*Activity log inserted before question*.] Based on the activity log above, which account was the least accessed on Sunday?
    - Facebook
    - Outlook
    - Instagram
    - Twitter
    - Google/Gmail

5. [*Activity log inserted before question*.] Based on the activity log above, where does this person usually access their Facebook account?
    - Café
    - Home
    - Office
    - Mall
    - Unknown

6. [*Activity log inserted before question*.] Would you prefer to check your account activity for each of your online accounts separately, or through a combined activity log like the [diagram *or* one] above?
   - Separately
   - Combined
   - I do not want to check my account activity in general
   - Other: [*open text field*]

7. [*Activity log inserted before question*.] How confident are you that you would be able to identify unusual activity on your accounts using a combined activity log like the [diagram *or* one] above?
   - Extremely uncertain
   - Somewhat uncertain
   - Neutral
   - Somewhat confident
   - Extremely confident

8. How secure would this combined activity log make you feel?
   - Extremely unsecure
   - Somewhat unsecure
   - Neutral
   - Somewhat secure
   - Extremely secure

9. Thank you for your time. Do you any comments on the topic of account monitoring or our study?