USER COMPREHENSION OF PASSWORD REUSE RISKS AND MITIGATIONS IN PASSWORD MANAGERS

by

Robbie MacGregor

Submitted in partial fulfillment of the
requirements for the degree of
Master of Computer Science

at

Dalhousie University
Halifax, Nova Scotia
April 2020

Table of Contents

List of Figures

Abstract

Passwords are a familiar and cheap way to authenticate the users of online services. Most users have many online accounts, but just a few unique passwords, resorting to strategies for the creation and recall of passwords that leave them vulnerable to password reuse attacks. The dangers associated with password reuse are not well understood by most users, and even partial reuse results in increased vulnerability.

This thesis describes an experiment testing a model reuse notification, delivered via a prototype password manager, among a group of online survey participants. I present evidence that a meaningful improvement in users' comprehension of password reuse and associated risks can be achieved by presenting users with password reuse dialogues in an explicitly cross-site context. This work directly addresses a current issue in the field of usable privacy and security, providing concrete data, and offering direction to researchers and developers seeking to better secure vulnerable users.

List of Abbreviations Used

2FA	Two-factor authentication
FWER	Family-wise error rate
HIT	Human Intelligence Task
ITS	Information Technology Services
MTurk	Amazon's Mechanical Turk
PII	Personally identifying information
SeBIS	Security Behavior Intentions Scale
SQL	Structured Query Language
SSL	Secure Socket Layer
SSO	Single sign-on
TLS	Transport Layer Security

Acknowledgements

I was fortunate to enjoy the support of a group of extraordinarily capable, kind, and generousness people while pursuing this research. I need to acknowledge a few of them by name.

· First, my mentor and supervisor, **Dr. Srinivas Sampalli** (Dalhousie University). Srini has supported my efforts for years, setting me up to succeed again and again despite myself. He has always shared his time and attention, and has always had my best interests at heart.

· **Dr. Kirstie Hawkey** (Dalhousie University) has been a patient, and caring mentor to me. Kirstie has always offered just the right amount of help, pointed me in the right direction, and left me to find the answers myself. Dr. Hawkey helped nurture my interest in usable privacy and security, and guided my entry into a community of researchers that has helped support and encourage me since.

· **Dr. Vlado Keselj** (Dalhousie University) has supported me in my degree from beginning to end. Vlado was my first professor, and taught me how to program. He is among the readers who evaluated this thesis. Dr. Keselj has always humoured me, offered advice, provided assistance.

· A confidant and kind support, **Dr. Raghav Sampangi** (Dalhousie University) is a skilled researcher who has always treated me like a peer. Raghav made time to talk through some of the ideas that would inform my thesis early on, piloted my survey, helped identify likely venues for publication.

· I am significantly in the debt of past researchers, **Dr. Elissa Redmiles** (Princeton University), **Dr. Maximilian Golla** (University of Ruhr), **Miranda Wei** (University of Washington), and **Dr. Blase Ur** (University of Chicago). Each of these researchers is a leader in their field, and each made time to welcome me at conferences, connect for in-person meetings, patiently reply to emails, offer guidance related to experimental design, surveys, and statistics.

· **Dr. Danielle Navarro** (University of New South Wales) is known internationally as an expert in conducting ethical research with human subject recruited via Amazon's

Mechanical Turk service. I was fortunate to have Danielle correspond with me, and provide guidance, as I developed a strategy for employing these tools in my data collection.

· My closest colleague and stalwart friend, **(Mir) Masood Ali** (University of Illinois, Chicago). Masood was a constant companion and support when we shared a lab. Since leaving to pursue his PhD, he has to insisted on continuing to proof my work, and reminded me of conference deadlines, pushed me to be a better researcher. He is like extended family, an extraordinary and devoted friend and peer.

· **Mitchell Kane** (Dalhousie University) has been with me since Day One. We have nurtured a great friendship over the course of many years, challenging each other, forcing each other to grow and consider new points of view. Mitchell has offered me formal and informal support without ever having to be asked. He is not someone I will ever take for granted.

· The many excellent **labmates** with whom I have shared highs and lows deserve special, collective thanks. These researchers have been variously kind, compassionate, and patient as I struggled to develop this work in our shared space. This document contains the word "password" 401 times. My labmates have heard the word perhaps 100000 times more than that, yet somehow remained cheerful and even encouraging.

· Finally, as always, particular thanks are owed to my wife, **Tiina Johns**, and my small son, **Finlay MacGregor**, for their faith and flexibility. They have given me the time and opportunity to conduct research and share my findings; to become a computer scientist.

Chapter 1

Introduction

Legacy, textual passwords are a fact of life. They have been with us since the early days of UNIX time-share systems [1], and they appear likely to persist into the future as a tool for authenticating the users of online accounts. The reasons for this are quite straightforward: passwords are familiar, cheap, and convenient. They are easy to use, readily deployable, and their longstanding incumbency has placed them in an almost unassailable position [2–4].

The problems associated with passwords have been well documented and studied [5–7]. Many issues have been apparent from the very beginning [8]. Where passwords are concerned, researchers and service providers are trying to address problems related to usability and security that have persisted for decades, but experts agree that it seems unlikely that an alternative authentication scheme will provide all of the benefits that textual passwords do while simultaneously offering users meaningful gains [3].

Many users of online accounts and services are overwhelmed by the demands placed upon them by password systems [7, 9–12]. In response to this, there have been many attempts to replace textual passwords with more secure alternative authentication methods [13], but these tools are not universally available, and are not always usable or readily accepted [14, 15].

With the proliferation of online accounts, users regularly resort to strategies to support the creation and recall of passwords that leave them vulnerable to password reuse attacks. Password reuse (using the same or similar passwords for multiple accounts) happens frequently. It is common for users to have dozens of online accounts, yet they often have just a few unique passwords [5]. Any reuse will result in increased vulnerability and an increased chance of account compromise. It makes little difference if these users have a handful of truly unique passwords shared across a hundred accounts or reuse a single string with superficial modifications. Any password creationstrategy

broadly involving reuse will leave a user vulnerable to attacks leveraging targeted guessing [16–18].

The risks associated with password reuse are poorly understood by users. While password reuse in general has been studied [5, 19, 20], there has been little work examining how reuse is communicated to users, to try to understand why and where their mental models fail. Prior work by Golla et al. [21] represents some of the first research to focus on password reuse notifications. These researchers explored strategies to help motivate users to change their existing passwords to make them less vulnerable to password reuse attacks, and to enhance their understanding of how and why to make these changes. They were challenged to improve users' understanding of a complex, cross-site problem, despite hewing to best practices from the field of warnings and alerts. Golla et al. propose seeking "ecosystem-level" solutions to problems related to password reuse in their call for further work; engaging users via third-party applications like browsers and password managers, rather than stand-alone notifications generated by online service providers [21].

The research described in this thesis takes up the call of Golla et al. [21], shifting domains and investigating an alternative channel for communicating password reuse to the users of online accounts. Working with the framework developed and validated by these researchers, I executed an experiment ($n = 260$) testing a model reuse notification via an online scenario-based survey. Participants encountered a dialogue in the context of a prototype password manager (*NewPass*), and reported attitudes and opinions reflecting their understanding of both the model notification and password reuse more generally. The experiment, survey, and sampling frame all closely reflect past experiments, supporting my inference that delivery channel can meaningfully impact user comprehension of password reuse. Participants viewing reuse notifications originating from a single service provider were seldom able to correctly identify the cause of a warning in past studies [21]. The great majority of my participants (more than 92% of respondents, across experimental conditions) could identify password reuse as the cause of the notifications they saw in the context of a password manager (an application with explicit access to credentials for multiple sites). Through a detailed accounting of my experimental results, I will paint a clear picture of the impact of my specific observations, and their potential importance in guiding future research in the

field of usable privacy and security.

In the chapters that follow I will unpack both foundational and recent research informing this thesis (§2 Background). I will discuss the motivations for the project and outline its place in the field of usable privacy and security, as well as present its relative merits. I will describe in detail my experiment (§3 Study), enumerating the research questions it was designed to address, methodologies employed, and the statistics I used to test my hypotheses and analyse my data. I will present my observed results (§4 Results). I will first describe my sample and the overall shape of my experimental data, before making cautious inferences to draw meaning from participant responses, and explore more complex interactions between experimental and subject variables. Finally, I will outline the concrete contributions of this thesis (§5 Conclusion), elaborating upon the likely value of my findings, as well as exploring the limitations of this project and avenues for future research.

Chapter 2

Background

Related work focused on passwords, password reuse, and password managers is summarized in this chapter (2.1 Related Work). Past research from the field of usable privacy and security focussed on password reuse notifications, warnings and alerts, improving users' comprehension of risk, and the adoption of security-positive behaviours is thoroughly unpacked (2.2 Prior Research). The factors motivating this research project are identified and discussed (2.3 Motivation), and a brief summary of the research objectives informed by prior work outlined (2.4 Objectives).

2.1 Related Work

When it comes to passwords, researchers and service providers are still trying to address problems that have been with us from the very beginning. The majority of users continue to be authenticated to systems and services by means of a memorized secret. Submitting a unique set of credentials (i.e. the combination of a username and textual password) is a considered sufficient proof of identity to verify a user and grant access to resources. The system closely resembles that described by Morris and Thompson for handling access control on early UNIX time-share systems [1] in research dating back more than 40 years [8]. While computing has advanced significantly in the intervening decades, the issues faced by those administering remote-access systems in the 60s and 70s mirror those faced by online service providers in the present where passwords and password security are concerned. Morris and Thompson enumerated a set of concerns related to password security, highlighting the need for users to adopt strong and unique passwords (that were unlikely to be either brute-forced or easily guessed by an attacker) and challenges associated with the secure storage of credentials. In both cases the rationale related to the need to limit the availability of data that might aid attackers or reduce search space in the event of a breach. In their short paper, the authors repeatedly discuss issues familiar to researchers in the fields of

usable privacy and security, and authentication, paying particular attention to the need for unique passwords and dangers associated with password reuse [8].

Despite acknowledged shortcomings, researchers in the fields of usable privacy and security, and authentication largely agree that legacy, textual passwords are here to stay. While there will occasionally be stories in the popular press heralding the imminent "death of the password" [22], those working on authentication and online identity management largely believe passwords will persist well into the future. Herley and van Oorschot made the case for just why passwords were likely to remain with us in a foundational paper [2], and much of what they discussed remains relevant almost a decade later. While passwords are readily deployable, they offer poor usability and provide a middling list of security benefits. Of principal concern to those hoping to replace passwords is improving the security of authentication systems. The main barriers to replacing passwords discussed by Herley and van Oorschot are their incumbency, as well as factors related to cost and the deployability of alternatives [2]. They argue that no single approach, protocol, or scheme can replace passwords in all of the places where they are currently the *de facto* method of authentication. This remains as true now as it was at the authoring of their paper. The challenge then becomes striking the correct balance when selecting tools to enhance user security, while working to understand the likely costs and benefits of various authentication schemes. In their conclusion, the authors link the predominant place of textual passwords in part to the lack of a systematic approach toward research in the field of authentication. They would seek to address this with their own framework, reflecting approaches from the field of human computer interaction [23], later the same year [3].

Researchers continue to be swayed by the arguments of Herley and van Oorschot [2], as well as those fleshed out in cooperation with Bonneau and Stajano [3], with whom they established the criteria by which authentication schemes are still judged. Joseph Bonneau, Cormac Herley, Paul C. van Oorschot, and Frank Stajano (an incredible team of authentication researchers, who by collegial agreement have listed their names alphabetically when publishing collectively) created a comprehensive framework for the evaluation of web authentication schemes in 2012 [3], to support an exhaustive survey of related tools and technologies[1]. Through their investigations they cemented the

[1] An extended technical report from the University of Cambridge presents the researchers' findings in detail for all schemes evaluated [24].

primacy of textual passwords for the authentication of users of online accounts. These researches showed that despite a "...near universal desire to replace passwords" [2] and a long-acknowledged set of shortcomings [4], textual passwords were the scheme to beat. As a long-standing incumbent technology (offering a unique set of usability, deployability, and security benefits), they determined that any authentication scheme seeking to supplant textual passwords should provide at minimum the same set of benefits, while improving some aspects of either usability or security. Bonneau et al. could not identify such a scheme at the time of their study, concluding that no option could even rise to the standard of "legacy passwords" let alone supersede them [3]. Researchers assessing modern authentication schemes have similarly found them wanting in comparison to textual passwords [25, 26], thought they have identified schemes which may augment their use and provide additional security.

With replacement unlikely, tools and security practices aimed at supporting the users of legacy, textual passwords are common. Herley and van Oorschot were early proponents of approaches designed to help users manage some of the burdens related to passwords, calling in particular for researchers to take a serious look at improving password managers [2]. It has since become common place to encourage users to adopt technologies to augment the use of passwords and improve security. Users are most frequently presented with advice to enable two-factor (2FA, or multi-factor) authentication schemes when these are available [15, 27], and to use password managers to support the creation and use of strong, unique passwords for all online accounts [28, 29]. This advice is not always well received. Users are often unclear why they should enrol in the multi-factor programs of service providers. Redmiles, in conjunction with the research group of Mazurek at the University of Maryland, produced a pair of related studies exploring why and how users adopted 2FA, as well as how to improve invitations to enrol in these programs [15, 30]. There is also recent evidence to suggest that users are doubly disadvantaged. Recent work by Ion et al. suggesting that many novice users may support false assumptions or have poor mental models when it comes to security-enhancing technologies [31]. Work by Fagan and Khan (picking up early threads from Herley [11]) suggests that more knowledgeable users may be rationally unwilling to accept trade-offs between security and convenience [14]. These factors are troubling, and need to be actively considered and addressed, given that the usability

of passwords is increasingly suffering.

2.1.1 Password Reuse

Textual passwords persist as a method of authentication, and problems related to passwords are becoming more common. This is largely due to two related issues: the proliferation of online accounts and password reuse. Faced with an exploding number of online accounts, the majority of users resort to password reuse (using the same or similar passwords for multiple accounts) in an effort to cope with the cognitive demands placed upon them [5–7, 9, 10]. Incidences of password reuse have been catalogued and observed extensively in recent years, and users' rationale of reuse collected and considered.

Gaw and Felten made an early set of observations focussed on reuse explicitly [7]. Working with a convenience sample of undergraduate students they recorded evidence suggesting both that password reuse was common and, importantly, increasing as users accumulated a greater number of online accounts while refusing to create new, unique credentials. These researchers report participants sharing rationale for their behaviours which closely reflected the findings of Herley [11], and the foundational observations of Adams and Sasse [12]. Adams and Sasse observed that users frequently understood that there were trade-offs between usability and security when it came to creating passwords (e.g. that a simple, memorable password was likely to compromise security) [12], while Herley found that the refusal to create strong, unique passwords for new accounts or adopt other security-enhancing habits, was at least in part an active decision they could justify to themselves [11]. Shay et al. further qualified these kinds of observations [10]. With the average user having literally dozens of accounts by the time of their writing, each requiring "strong" textual passwords that adhere to the password policies of account providers, users were reusing passwords as a way of coping with unrealistic demands [10]. The net result, as reported by the teams of Das et al. [20] and Wash et al. [32], is that the majority of passwords are now reused.

This is doubly true if serious consideration is given to the practice of partial password reuse, where users make trivial modifications or additions to an existing base password (adding or removing special characters or digits, "leetifying" common characters by exchanging them for confusable digits, etc.) [5,6]. Pearman et al. presented

evidence to suggest that those working in usable privacy and security needed to be doing just this. These authors greatly expanded our understanding of how users actually engage with passwords [5], presenting the results of a large-scale, longitudinal, *in situ* study designed to engage with past password research (notably that of Wash et al. [32], who had conducted the most thorough investigations into passwords and reuse prior to this). Pearman et al. were able to access much more granular data, and instrument behaviour outside of the browser, by leveraging Carnegie Mellon University's *Security Behavior Observatory* [33] for their data collection. The data reflected an extraordinary level of ecological validity, and through the creation of special-purpose software Pearman's team were able to measure and observe the partial reuse of passwords clearly for the first time [5]. A central take-away from the findings of Pearman et al. is that password reuse in all its forms is more prevalent that previously thought, and that partial reuse is rampant.

Partial reuse matters. Measuring the strength of passwords can be contentious, but most agree that naive calculations relying upon Shannon Entropy [34] are insufficient. The majority of password policies will result in the composition of passwords that should be immune to brute force attacks [35]. What is of much greater concern is the "guessability" of passwords [36–39]. In the last decade, researchers in the field of authentication have adopted an approach to estimating the strength of passwords that more closely reflects their ability to support users' security goals. Kelley et al. [39] were early proponents of measuring the strength of passwords based upon how likely they were to fail when common password cracking techniques and algorithms were applied. More recently this thread has been picked up by Wheeler [38], whose strength estimation algorithm reflected four of the most common guessing attacks. The teams of Melicher et al. [37], Guo and Zhang [36], have all attempted to reflect real-word attacks in modelling password strength, addressing the efficacy of textual passwords for securing online accounts by considering how credentials are actually compromised, and the risks reuse (partial or exact) can represent; combinations and permutations on previously captured credentials featuring prominently in the cracking strategies most frequently employed.

Users have many accounts, frequently reuse passwords, and often have just a handful of usernames associated with these [5, 20, 32]. This scenario results not just in

users becoming more vulnerable, but also in an increased motivation for attackers to mine available data from breaches to attempt account compromises. Where password reuse is concerned, users are most vulnerable to targeted guessing based attacks, where permutations on a known username and password pair are systematically tested in an effort to compromise related online accounts [16–18].

Data breaches and account compromise significantly increase the vulnerability of affected users, increasing their risk of future compromise and exposing them to targeted attacks. With reuse commonplace, the availability of a username and textual password associated with one account or service will frequently provide an opportunity to attackers to limit their search space when attempting to hack other online accounts [16]. The risks associated with password reuse become clear when a larger online ecosystem is considered. The compromise of an account with one service provider, can result in a cascade of failures and compromises. Han et al. validated this empirically [17]. In a large-scale study sourcing data from a series of real data breaches, these authors showed that similar usernames and passwords were frequently used across multiple accounts. They further presented empirical, quantitative evidence supporting the claim that leaked password data could be used to improve the efficiency of known attacks, algorithms and popular cracking tools [40]. Han et al. also attended to partial reuse in their experiment, and were able to demonstrate in a manner reflecting a high level of ecological validity how weak passwords could reveal information about related stronger passwords, simplifying cracking and account compromise [17]. Wang et al. [18] further made the risks associated with reuse plain, publishing a study reflecting real-world attack conditions and presenting empirical measures to show how targeted guessing could be effective in online contexts where lock-out and throttling might reasonably frustrate attackers. These researchers were able to demonstrate conclusively how the availability of related password and account data (e.g. username) could reliably enable the cracking of secondary passwords and compromise of associated accounts [18].

2.2 Prior Research

Password reuse attacks represent a real danger, but are not widely or well understood by the majority of users, nor are the practices and behaviours that result in increased vulnerability [6, 31]. Golla et al. [21] were the first researchers to seriously investigate

how password reuse was communicated to users, and how users responded to the strategies employed by online account providers to alert them to password reuse and encourage corrective action. The authors completed a pair of linked studies. The first study of Golla et al. investigated existing strategies to help motivate users to change duplicate passwords, to make them less vulnerable to password reuse attacks. After completing an exhaustive survey of password reuse messages generated by online account providers in response to data breaches, they assessed the efficacy of state-of-the-art approaches with a large-scale user study. They captured quantitative as well as qualitative data related to user attitudes and comprehension, and their findings were not encouraging. Testing six, representative notifications (those generated by Netflix, LinkedIn, Instagram, Facebook, for Google accounts and Gmail), Golla et al. found that none were reliably understood by participants to be related to password reuse, nor were they likely to result in appropriate action on the part of users [21].

Golla et al. were able to identify statistically significant differences in the ways participants responded to the individual notifications, even if none adequately accomplished the goals of informing and motivating users. The differences observed permitted the researchers to isolate factors which tended to improve the effectiveness of password reuse notifications. the researchers identified five goals that any password reuse notification needed to achieve. Treating these as a framework, they devised a model reuse notification, designed to improve upon existing reuse messaging. In their second study, the group sought to validate this model notification. They conducted a large-scale, online case study, exposing participants to 15 variations of a model reuse notification. The researchers captured both quantitative and qualitative data in an attempt to measure a meaningful difference in users' responses to their conditions, as well as asses the validity of their model in light of previous results from the first study. While the overall picture was still not good, the model notification of Golla et al. outperformed the state-of-the-art, and informed a set of best practices they propose others generating password reuse notifications ought to follow [21].

The results of Golla et al. suggest that password reuse can not be addressed satisfactorily by online account providers alone. By these researchers' assessment, this is a reflection of the complex, cross-site nature of password reuse attacks, and the inaccurate and inadequate mental models of the users of textual passwords. They

conclude their paper by recommending that other researchers hoping to improve reuse notifications and address problems associated with users' understanding of password reuse focus on "ecosystem-level" strategies. They suggest others explore leveraging tools and channels users associate with more than one online account: browsers and password managers [21].

2.3 Motivation

This thesis responds directly to the call of Golla et al. [21], shifting domains and investigating an alternative channel for communicating password reuse to the users of online accounts. Where past research has assessed the impact of password reuse notifications related to data breaches and threat analysis conducted by the providers of online accounts [21], this project is the first to take a closer look at how users understand and respond to notifications generated by third-party applications explicitly connected to passwords and security. Password managers, broadly speaking, are an important tool for addressing password reuse. In this research, I do not differentiate between stand-alone password managers [41–45] and those included in major modern browsers [46, 47]. The latter have come in recent years to offer all of the features commonly associated with other third-party offerings, distinguishing themselves principally by leveraging convenience in an attempt to encourage users to enrol. Conversely, the most popular stand-alone password managers function largely as web apps, and deliver notifications in the browser and across sites by means of plugins and extensions. In effect, seriously considering password managers as a means to communicate password reuse results in addressing both password managers and browsers as communication channels simultaneously.

Whether or not password managers can reliably improve the security of users is a somewhat contentious issue. While this thesis reflects an intuition expressed by Golla et al. [21], I have not blindly assumed that the cross-site cues provided by password managers will fix users' mental models or solve the reuse problem. I have instead picked up another recent thread from the field of authentication. A 2018 study by Lyastani et al. [29] focused on the impact of password managers on the creation of strong, unique passwords. The work of these researchers addresses earlier claims that password managers provide no net security benefit. The use of password managers

alone has not been shown to significantly improve the security of users nor to uniformly reduce incidences of password reuse in past studies. Neither the team of Wash et al. [32] nor that of Pearman et al. [5] found evidence of a significant security benefit associated with password manager use. However, these well known and frequently cited studies relied significantly upon inference to determine how and when participants were interacting with password managers. Researchers could with a high degree of certainty determine if participants had a password manager installed, and observe them logging into some online services absent any keyboard input: that is the extent of what could be measured [5, 32].

Lyastani et al. conducted a large-scale study, instrumenting carefully many aspects of their participants' interactions with password managers. The goal of these researchers was to do away with inference, and measure the impact of password managers on password strength and password reuse [29]. Their findings offer some hope. While they confirmed that merely having a password manager installed was not enough to improve security outcomes, the researchers observed results suggesting that interaction with password managers did indeed impact password strength and uniqueness. Interestingly, they found that security benefits were dependent largely upon how password managers were used. Participants who engaged with password managers during the creation of new passwords (i.e. to save a new credential pair, rather than store an existing one) tended to generate passwords that were stronger and more likely to be unique than those that did not. The use of password managers to support tasks not related to the creation of new passwords (i.e. to autofill fields) was observed to either have no meaningful impact on user security (password strength and reuse) or to aggravate existing problems related to weak passwords and reuse. Participants using password managers as simple aids to memory or as tool to automate the input of existing passwords, were observed to be less secure than those who did not use password managers at all. In their call for further work, Lyastani et al. enjoin researchers to examine how password managers can "better support users' password strategies in order to improve password security as well as stop aggravating existing problems" [29]. This call coupled with their observation that some password manager-based interventions can significantly impact password reuse [29] informs my thesis and the decision to conduct experiments in line with the intuitions of Golla et al. [21].

This thesis is positioned to make a meaningful contribution to the fields of usable privacy and security, and authentication. Where prior work explored the notifications of service providers [21], I explore the promise of using password managers to better communicate reuse and support users. Where past work has examined correlations between the use of password managers and the strength and uniqueness of user passwords [29], this project enhances the understanding of how users interpret the feedback they get from these applications. But the reasons to explore this topic go beyond the opportunity to modestly advance an existing body of research.

The users of online accounts are likely to continue using textual passwords for the foreseeable future [2–4], and the most common strategies employed by users to manage an increasing number of these passwords result in significant password reuse [5, 20, 32], leaving them vulnerable to account compromise [16–18]. In response to this situation, users are frequently encouraged to adopt new security practices; adopting more secure alternative authentication methods and single sign-on (SSO) options [13], enabling multi-factor authentication (also known as two-factor authentication or 2FA) [27], and using password managers [28]. These tools, unfortunately, are not widely adopted. The barriers to adoption can take many forms, but researchers have reliably observed that users are not well supported in the development of mental models that will leave them receptive to new tools and security behaviours [14, 15, 31].

2.3.1 Learning, Accessibility and Comprehension

In a 2017 study, Redmiles and Mazurek [15] examined the role of messaging and dialogues in motivating users to adopt new security behaviours. They worked with a small group of participants ($n = 12$) to generate qualitative data reflecting the efficacy of different strategies for encouraging users to enrol in the 2FA schemes of online service providers. The case study built on a prior investigation with Koss that featured a large-scale, census-representative survey [30]. This work yielded strong evidence that many users could and did learn new security behaviours from warnings, alerts, and related dialogues. More than 80% of participants reported that they learned new security behaviours from messages generated by applications and online tools [30]. This was evidence enough for Redmiles and Mazurek to choose to focus on messaging and its potential to motivate users in their 2FA invitation case study, which identified

a set of factors that could reliably improve the invitations online account providers generate to encourage users to enable multi-factor authentication. Their work resulted in a testable set of best practices, a prototype message developed via co-design, but their discussion of the study and its motivations is perhaps most valuable.

In their paper, Redmiles and Mazurek point to an opportunity to improve the mental models of users, to adapt to their needs, and support them through the design of applications and their dialogues [15]. Their small, case study reflects a first attempt to engage with the results of a much larger survey in a constructive way. Redmiles and Mazurek built upon past work to find a way to capitalize on the fact that many users appear ready to adopt new security practices if presented with the right kind of dialogue. In making the case for other researchers to further explore options related to messaging and motivation (distinct from reactive alerts designed to correct user behaviour) the researchers introduce an important, related topic; accessibility [15].

Accessibility is frequently a focus for researchers in the field of usable privacy and security. Models and research goals have to account for users who are not always able-bodied, highly-educated, well-resourced, or in possession of technical expertise. The research of Redmiles and Mazurek [15] is positioned in response to prior work by Rader et al. [48], as well as previously published work by Redmiles and the team at the University of Maryland [30, 49], which highlights the importance of messaging and dialogues in helping users who (largely due to economic status) might need to rely more on applications as a source of security knowledge. Users learn new security behaviours from many sources (at work, school, from friends and family) [48], but for many, warnings, alerts, and other application dialogues are a crucial source of information [30]. Poorer users are less likely to access new security knowledge, less likely to be exposed to new security behaviours at work or school, and less likely to have close social relations with others who do. In the case of these users, the messages and notifications generated by applications can be one of the best sources of information and security advice [15, 49].

This thesis, in addressing users' understanding of the complex problem of password reuse, also seeks in part to address the "digital divide" described by Redmiles et al. [49]. The project is about better understanding how users respond to password reuse messages presented in the context of the password manager, and exploring an

opportunity to better support and secure users interacting with these applications through the notifications and dialogues they encounter. Porter Felt et al. [50] argued persuasively that alerts and other dialogues could and should focus on improving users' understanding of the threats they face. They presented the case of balancing the goal of encouraging immediate adherence (compliance with directions or advice) against the more long-term goal of improving user comprehension regarding complex threats. They demonstrated that applications and their dialogues could contribute to users making better security decisions in future [50]. The finding of these researchers, and the framework they published run parallel to the best practices of Golla et al. [21] in important ways, and suggest that it is possible to extend the work of those researchers while focussing explicitly on improving user comprehension of a complex threat that could result in an overall improvement when it comes to online security [50]. While Porter Felt et al. conducted research related to SSL/TLS warnings and alerts, their goal of addressing the underlying complexity of a problem poorly understood by users is reflected in the work of researchers such as Golla et al. [21] concerning password reuse, and validates a set of strategies worth exploring.

It is possible to reach users through alerts, notifications, and other dialogues, and to engage with them in a way that will improve their understanding of the security risks they face [50]. Users can and do adopt new behaviours, and change old practices, based on the messages applications generate [30]. Researchers in the field of usable privacy and security are interested in users, and care about people. My thesis research is therefore not just motivated by a desire to improve password security, but to support and secure users. This work capitalizes on an exciting opportunity at the intersection of a number of concerns, and has the potential to inform future work to support the users of online accounts and related security focused applications like password managers through providing guidance related to the development of password reuse notifications. These kinds of notifications have been shown to contribute in a meaningful way to better securing vulnerable users with fewer resources [30, 48, 49], and are worthy of serious consideration, study, and further development. I have designed and carried out an experiment with these issues in mind.

2.4 Objectives

A modest set of research objectives are informed by the prior work discussed in this chapter. My thesis responds directly to calls from researchers in the fields of usable privacy and security, and authentication, presenting a preliminary investigation into the efficacy of communicating password reuse via password managers. This is an early, exploratory project to provide concrete data regarding the attitudes and intentions of users encountering password reuse messages originating with applications that have explicit cross-site access. I reflect past methodologies and frameworks for the development of notifications, in an attempt to facilitate a comparison of my own results with past measures. A series of regression analyses provide data helping to isolate factors improving the effectiveness of dialogues. The overall aim of this project is to expand upon past knowledge and offer some direction to researchers seeking to better secure vulnerable users and develop more effective and usable security-enhancing tools.

Chapter 3

Study

This chapter describes in detail the study at the heart of my thesis. I outline all aspects of my experiment. The research questions addressed by this thesis are first enumerated (3.1 Research Questions). The rationale for my study are revisited and further unpacked in context (3.2 Rationale). My experimental methodology (3.3 Methodology), survey tools (3.3.1 Survey), experimental conditions (3.3.2 Conditions), and study participants are all discussed independently (3.3.3 Participants). Finally, the statistics used to test my hypotheses and analyse my data are outlined (3.4 Statistics).

3.1 Research Questions

Past researchers have reported that even when password reuse notifications hew to established frameworks and reflect best practices from the field of usable privacy and security regarding warnings and alerts, they fail to effectively improve the understanding of users or improve the likelihood of users taking appropriate action to address risks associated with password reuse [21]. It has been suggested that this failure is in someway linked to the fact that most of the password reuse notifications users encounter come from a single online service provider. The problems associated with password reuse are complex, and cross-site in nature. Notifications originating with a single online service provider do little to improve, change or challenge mental models, and appear to have little effect [21].

The most important research question addressed by my thesis is that of whether password reuse can be effectively communicated when notifications originate with an application understood by users to have explicit access to multiple credentials, for multiple online accounts. An important related question is whether notifications in this context can be effective enough to reasonably be expected to improve security outcomes related to password reuse.

RQ1: Can password reuse notifications originating with a password manager improve

user comprehension regarding the problem of password reuse, its associated risks?

RQ2: Can password reuse notifications originating with a password manager improve the likelihood of users taking appropriate corrective action, replacing affected passwords with strong, unique ones?

I additionally consider what kinds of interactions with password managers can reinforce messaging, improving the effectiveness of reuse notifications.

RQ3: Within the context of a password manager, can the channel or task associated with password reuse notifications impact their effectiveness?

This research is preliminary and exploratory, but positioned to make a contribution to the fields of usable privacy and security, and authentication. My research questions are designed to validate the approach of alerting users to password reuse, and communicating associated risks, via the dialogues of password managers. My experimental results are meant to provide an early concrete test of the intuitions of past researchers, and may inform the development of a fully interactive prototype application or plugin, and testing with partners in academia and industry in future. This research also has the potential to extend existing frameworks for supporting the development of reuse notifications across domains, as I further isolate effective factors, and consider complex interactions in my analysis of study results.

3.2 Rationale

This thesis seriously and scientifically explores the efficacy of communicating password reuse and associated risks to users via password managers. The project is informed by the intuitions of past researchers, and there is a common sense appeal to addressing reuse via security-focussed applications like password managers, but scientists are not inventors, and are not guided by hunches. Gut feelings and common sense are not sufficient evidence to direct resources toward the development and testing of a new application or extension. The strategies such a tool might embody need to first be validated in a rigorous way.

The research described in this thesis takes up the call of Golla et al. [21], shifting domains and investigating an alternative channel for communicating password reuse to the users of online accounts. It adapts the framework developed and validated by these researchers, extending it modestly by incorporating explicitly concepts from Porter

Felt et al. [50], and addressing elements from industry leading password managers to enhance external validity [41–47]. My work validates an approach to addressing password reuse not previously investigated, while also seriously considering the most recent findings concerning how and when interaction with password managers has been shown to enhance the security of users, and considering the impact password manager-based interventions can have during key activities [29].

The study conducted in support of this research has been shaped by recent, important developments in the fields of usable privacy and security, and authentication [21, 29, 50]. It is motivated by the understanding that addressing problems like password reuse through the right kind of notifications can offer potential benefits with regard to accessibility as well by enhancing the security of users with fewer resources [30, 48, 49], and reflects my larger goal of supporting good security decision making for all users. The results of my experiment are meant to help point the way forward, identifying opportunities to improve upon or augment existing messaging and strategies for addressing password reuse.

3.3 Methodology

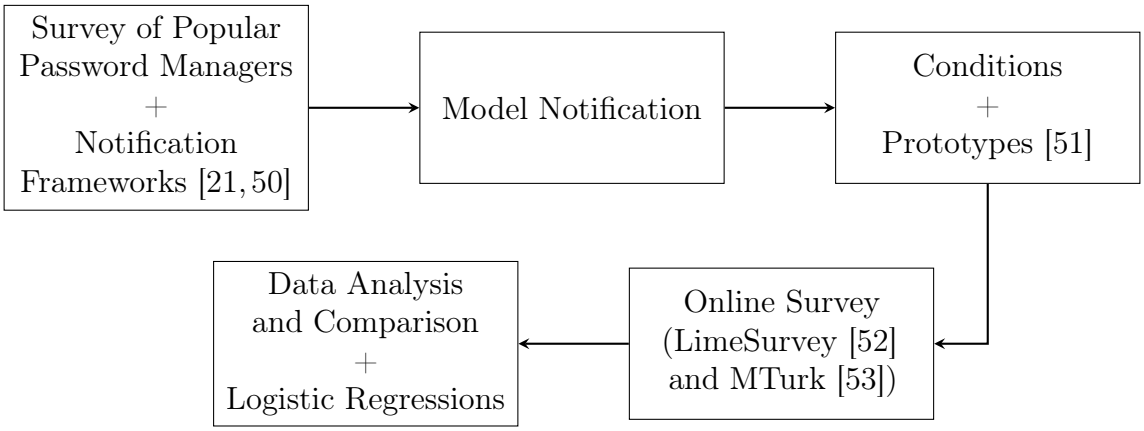


Figure 3.1: Experimental Methodology

This figure depicts the design and development of the thesis experiment, from the creation of the model notification and prototypes, to data collection and analysis.

To determine whether password reuse notifications originating with a password manager could improve user comprehension regarding the problem of password reuse and its associated risks, and effectively encourage users to replace duplicate passwords

with strong, unique ones, I executed an experiment testing a model reuse notification via an online scenario-based survey. Participants encountered a dialogue in the context of a prototype password manager, and reported attitudes and opinions reflecting their understanding of both the model notification and password reuse more generally. Like Golla et al. [21], I employed a factorial design in my experiment. I modified a model notification in an attempt to isolate the impact of three experimental variables. I conducted a preliminary, between subjects study with the goal of generating data to better asses how users understand and respond to password reuse messages originating with password managers.

My experimental design reflects a desire to efficiently and thoroughly validate a set of best practices and extend past work into a new, adjacent domain. As a result, my approach closely parallels that of the team of Golla et al. [21], who proposed a framework for reuse notifications and called for researchers to investigate opportunities to communicate password reuse via password managers and browsers. To facilitate comparison, improve consistency, and avoid the introduction of confounding factors, I have necessarily implemented some of the strategies adopted and validated by others. The methodology outlined in this section reflects the practices of past researchers[1], as well as established practices from the fields of human computer interaction, and usable privacy and security concerning how to best conduct online research while guarding the validity of results.

Like these researchers, my work relies upon participants' self-reporting of their attitudes, intentions, and understanding. This is common in early investigations in the field of human computer interaction, but can present challenges when interpreting experimental results and generating inferences. There has been some skepticism expressed in particular with regard to the validity of self-report data where matters of security and privacy are concerned [54]. More recently, however, a team from the University of Maryland lead by Redmiles has presented findings which suggest that while responses related to security messages may show some bias, they do indeed correlate to participants actual responses in the field [55]. Though self-report data does not map perfectly to real-world responses and security behaviours, it has been shown

[1]It should again be briefly noted that I have benefited from the patience and support of peers and mentors who conducted the study to which my work responds [21], receiving guidance concerning the development of my survey instrument in particular.

via a large-scale systematic study to provide a reliable indication of attitude, intent, and action [55]. Knowing this to be the case, I argue that this kind of data is more than sufficient to support a preliminary investigation, focussed on the comprehension of notifications and associated risks. My experimental design has the added benefit of providing rapid results while maximizing convenience and control.

3.3.1 Survey

My thesis and experiment relied on a single, crucial instrument: an online survey hosted locally on the graduate server of Dalhousie University's Faculty of Computer Science. This survey tool was the source of all experimental data. Following recruitment, study participants were directed via a static link to a survey site designed specifically for this research project. They were presented with a scenario by means of a short introductory text, and asked to imagine themselves to be the users of a new password manager (*NewPass*). They were then shown one of eight high fidelity prototype images depicting an onscreen notification associated with a password related task, and guided through an online survey consisting of 35 short answer and multiple choice questions designed to capture their understanding of a security problem related to password reuse, as well as relevant demographic information.

Knowing that the value and validity of my experimental data would be largely contingent on the strength of the survey instrument, I ensured its development and testing reflected best practices and established methodologies. I deployed strategies designed to limit response bias and guard the validity of results, and implemented a set of best practices specific to survey work in privacy and security focussed research [56]. Established best practices were observed in constructing and conducting the survey [57], and following the approach of Golla et al. [21] my survey design reflected long-standing practices from the field of social psychology to limit biases related to social desirability in participant responses [58].

My survey instrument resembles in many ways the survey of Golla et al. [21]. This is by design. To address my research questions (in particular **RQ1** and **RQ2**), I required a point of comparison or control. I chose the expedient path of treating past results (from a very similar experiment, conducted under very similar conditions, with the same sampling frame) as this point of comparison, and used them to provide

context for observations and inferences. My experiment and survey questions were developed in consultation with past researchers [21], and I have captured data amiable to comparison, at least in the context of a preliminary investigation. My survey instrument was additionally designed to provide data specifically related to participants' attitudes toward and experiences with passwords and password managers.

The survey questions are presented in their entirety in an appendix to this thesis (Appendix B: Survey Instrument). Experimental questions ask participants to describe what a password reuse notification is telling them, to rate their agreement with statements about the notification (via 5-point Likert scale), or to report how they might respond to it. A combination of multiple choice and short answer questions designed to generate quantitative and qualitative data, provide measures of consistency in the responses of a given participant, and offer unobtrusive attention checks have been included. Security knowledge, attitudes regarding passwords and password managers are likewise captured, along with questions related to past experience with data breaches and account compromise, past experience using password managers.

Questions addressing security and privacy can result in biased responses. As a result, these questions take a number of forms in my survey, and some are indirect [59]. Others reflect extensively validated approaches from the field of usable privacy and security. In an attempt to stay within a reasonable question budget, I eschewed the approach of Lyastani et al. [29], who used questions from Westin's Privacy Segmentation Index [60] in an attempt to capture participants' attitudes related to privacy and security, using a more modern psychometric tool; the Security Behavior Intentions Scale (SeBIS) of Egelman et al. [61, 62]. From the SeBIS question set, I included only those related to passwords. Combining data from these four questions with responses to questions focusing on participants' attitudes regarding the utility of passwords in supporting personal security goals from the survey of past researchers. This permitted me to focus in on the most relevant security related attitudes supported by respondents, and also to compare my sample against that of Golla et al. [21].

To improve analyses, and provide the opportunity to fully explore interactions between experimental variables and potentially confounding subject variables, I administered a secondary set of questions related to demographic information about age, gender identity, education, technical knowledge and expertise. I chose to limit

questions focussed on technical ability and knowledge, to keep the survey from running longer than necessary, and to avoid fatiguing participants. Instead of using a validated method like the questions of Hargittai and Hsieh [50, 63], which wold have required six more questions, I stuck with the parsimonious approach of past researchers. Like Golla et al. I have treated data about education attainment and employment (particularly employment in fields related to IT, computer science, or engineering) as providing measures which can proxy technical expertise [21].

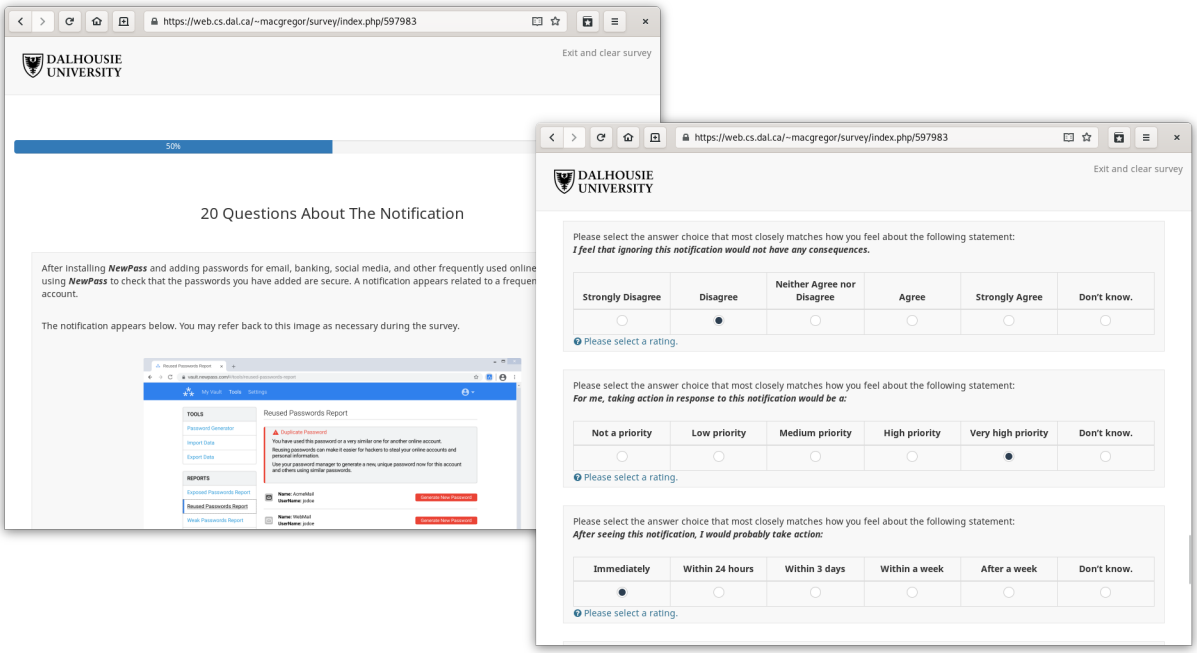


Figure 3.2: Survey Presentation

The survey was administered, and participants' interaction with the survey site instrumented, using tools provided by LimeSurvey GmbH [52]. LimeSurvey provides extremely high quality survey software under GNU General Public License (GPLv3), and its open core model permits researchers to create and customize their own local instances without any interaction with either the company or its partners. I used the LimeSurvey software to develop a custom site to host the scenario-based survey, to collect and manage experimental data securely. LimeSurvey is built using PHP, with AJAX for interaction, and supports MySQL and SQLite databases. This greatly simplified creating an instance on the Hector graduate server of Dalhousie University's

Faculty of Computer Science. I was able to use local resources, secured by Dalhousie's ITS group, to securely store data by pointing the application at my own database on the Faculty's SQL server. Without compromising security (the confidentiality of participant data), I was able to do more customization my survey site, access better block and question randomization options, provide better mobile optimization and a more modern experience for participants than I might have, had I used the University's Opinio survey tool [64]. LimeSurvey offered the additional benefit of being able to more readily manage interaction with participants recruited from online services. The application simplified the task of generating completion codes for participants recruited via Amazon's Mechanical Turk (MTurk) service [53], managing persistent random identifiers while dropping or ignoring possible source of personally identifying information (PII), and permitted me to write scripts to handle the random assignment of participants to experimental conditions, among other things. Finally, the software provided significant control over the kind of data each question would generate (nominal, ordinal, ratio) and offered direct export to formats compatible with both the GNU PSPP [65] and IBM SPSS [66] statistical suites. An example of the kind of screens displayed to users via the survey site is presented in Figure 3.2.

The survey instrument and site were developed and tested over a period of one month, prior to the recruitment of participants. Members of my lab group ($N = 16$) piloted the survey extensively and repeatedly as part of an iterative development process. The design of both the survey proper and its presentation were improved. I was able to test scripts controlling the assignment of participants to experimental conditions, and well as those managing the flow of a given survey task (adding or omitting questions reflecting past responses as participants progressed through the survey site's screens). An early export of representative response data was used to ensure the correct formatting data and syntax files used by the statistical software, the validity of my proposed approach to statistical testing, and the accurate instrumentation of questions to capture timing data.

3.3.2 Conditions

I employed a $2 \times 2 \times 2$ factorial design in which three independent variables (each with two levels) were manipulated, resulting in eight experimental conditions. The

experimental conditions encountered by survey participants reflect modifications to a model password reuse notification. Significant research informed the development of this model. I engaged with prior work from the fields of warnings and alerts, and usable privacy and security, considering established frameworks and practices for capturing attention and communicating risk [67–70]. Warning and alerts, and related dialogues are well-trod territory. I felt it important to be conversant in the arguments from primary sources, like recent foundational works of Bravo-Lillo and colleagues [67, 68], Akhawe and Porter Felt [69], Egelman and Schechter [70], to ensure the validity of my approach. Golla et al. acknowledged a similar body of research when attempting to synthesize a notification reflecting the best of the state-of-the-art messages they surveyed, and prior work necessarily informed their set of best practices for developing password reuse notifications [21].

This thesis represents a concerted attempt to extend the work of Golla et al., focussing on the effectiveness of password reuse notifications generated by and delivered via password managers. I adopted the best practices proposed by these researchers in developing my experiment and model. Golla et al. identified five practices they argued should employed in the development of password reuse notifications. The approach was validated by the results of their linked studies [21]. The best practices might be paraphrased:

Password Reuse Notification Best Practices [21]

1. Reuse notifications should be explicit. Making the cause of password reuse as plain as possible.
2. Reuse notifications should force a password reset. Making corrective action mandatory.
3. Reuse notifications should strongly encourage the replacement of similar passwords. Making the cross-site nature of the problem clearer and improving security.
4. Reuse notifications should strongly encourage the adoption of other security-enhancing technologies (e.g. 2FA). Making attacks related to reuse more difficult and improving mental models.
5. Reuse notifications should be delivered via immediate and trusted communications channels (e.g. email and push), more than one when possible. Making a notification appear more valid.

I directly reflected these best practices in the development of my model password reuse notification, and the conditions survey participants encountered. I made a reasoned decision not to include a 2FA invitation, after members of the original research team suggested this factor was perhaps one that showed statistical, but not practical significance in their analysis [21], but addressed each of the others in my design.

Along side this set of best practices, I also considered the guidelines of Porter Felt et al. for improving users' comprehension of security risks [50], viewing this work to run happily parallel to that of Golla et al. in many ways. One of the major problems discussed in the work of Golla et al. [21], and further explored in this thesis, is the persistent failure of reuse notifications to be read correctly and understood by users. I sought to develop a model password reuse notification that reflected the approach of past researchers, while being mindful of an original source (cited in their published findings) that had in part informed their investigations and approach to designing notifications [21]. When attempting to address and improve users' comprehension of complex security problems through warnings, alerts, and other dialogues, in an effort to support informed security decision making and improve long-term outcomes, Porter Felt et al. suggest a parsimonious approach that might be paraphrased:

Addressing Comprehension [50]

1. Address the source of a threat explicitly. Making clear where a security problem is located (e.g. local machine, browser, website of a service provider).
2. Address what data is at risk. Providing context, making clear what possible negative outcomes could be associated with inaction.
3. Address the possibility of false positives. Making the validity of notifications plain.

Incorporating these two related approaches I built my model password reuse notification, sourcing phrases and dialogues from market-leading password managers. These I modified and combined to suit the particular needs of my study. The applications surveyed in advance of my research included the stand-alone applications 1Password [41], BitWarden [42], Dashlane [43], Keeper [44], and LastPass [45]. I additionally examined the password managers included in two major modern browsers: Google Chrome [46]

and Mozilla Firefox [47]. Prior to the start of my thesis research proper, I completed an extensive survey of these applications focussing on usability, dialogues and notifications. I attended particularly to password reuse: both how reuse was measured (e.g. did applications identify partial or merely exact reuse) and how it was communicated to users. I found a number of common factors during my coding of the data collected, as well as opportunities to improve upon existing offerings. My results will be more fully described in a secondary publication. For the model reuse notification, I borrowed from and synthesized password reuse dialogues and notifications from the password managers tested, to improve participant acceptance and enhance the validity of my study. I looked to the security audit tasks, challenges, and alerts from each of the seven applications drawing out related features and taking care to capture those that could directly address elements from the framework of Porter Felt et al. [50] and the best practices of Golla's research team [21].

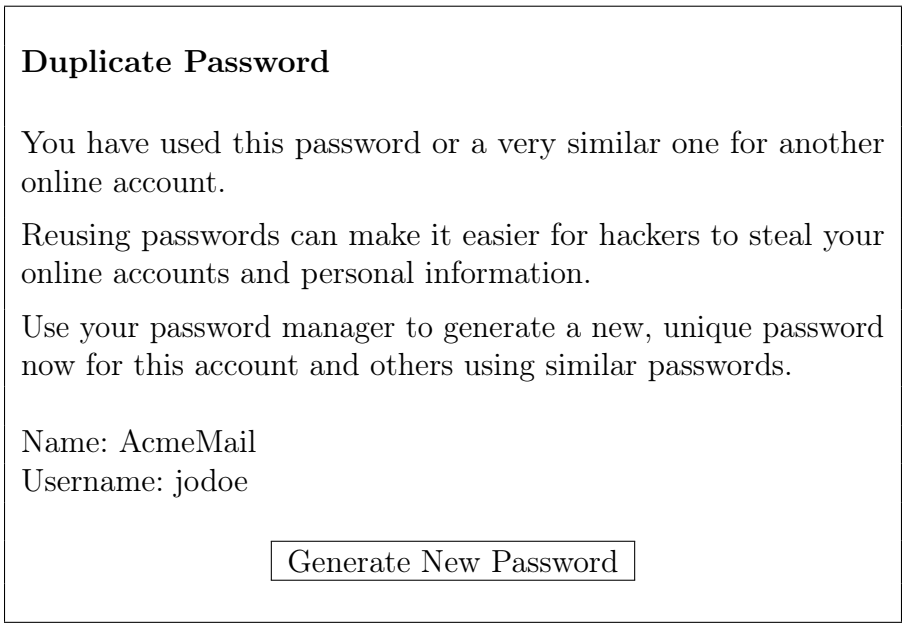


Figure 3.3: Model Notification

A generic version of my model password reuse notification is presented in Figure 3.3. Direct, imperative language is employed throughout the model notification. Password reuse is clearly identified as the cause of the notification, with partial reuse explicitly included. The notification directs users to generate a new, unique password, providing no alternative course of action. The danger of password reuse is outlined in brief,

lay terms, and the data at risk is clearly identified. Related accounts (those relying on similar credentials) are referenced, and users are directed to change associated passwords as well. The possibility of false positive alerts is implicitly addressed by the context of the password manager, which must necessarily store credentials for multiple accounts. Delivery channel is addressed by the fist of three experimental variables, though it is always assumed to be presented in a timely and direct fashion.

Implementing a $2\times2\times2$ factorial design, the model notification was modified slightly in order to operationalize three experimental variables. All possible combinations of these varied factors were tested, resulting in a total of eight conditions encountered by study participants. Each of the experimental variables is designed to help isolate factors which might result in observed difference between my own results and those of past researchers[2].

	Delivery Channel
Login Alert:	Active alert during simulated login task
Security Audit:	Application dialogue during simulated security audit task
	Reuse
Partial:	"You have used this password or a very similar one for another online account... and others using similar passwords."
Exact:	"You have used this password for another online account... and others using this password."
	Remediation
Generate:	"Use your password manager to generate a new, unique password now..."
/ :	"Create a new, unique password now..."

Figure 3.4: Notification Factors Varied

This figure displays the factors varied across my study's eight conditions at a glance, and each of the experimental variables is described in detail in the following section.

[2]The scope of this experiment, and the variables manipulated, where again discussed with members of the research team of Golla et al. [21]

V1: (Delivery Channel) – The first variable addresses the best practices of Golla et al. [21], facilitating an exploration of how the channel of communication impacts the effectiveness of password reuse notifications in a new domain. This experimental variable also reflects the observations of Lyastani et al. [29], who found that the security benefits associated with password manager use where different depending on the tasks users engaged in and how they interacted with the applications. Study participants were presented with conditions featuring one of two simulated tasks. An audit task (reflecting the challenges, check-ups, and diagnostic tools provided by all of the password managers surveyed) was presented via high fidelity prototype images depicting a web application. A simulated login task, interrupted via a dynamic, active alert (most similar to those generated by the LastPass browser extension[3]) features a generic email login page.

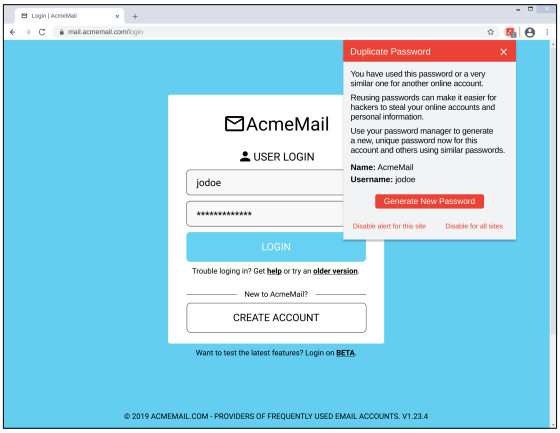
V2: (Reuse) – The second variable relates to mental models and the nature of password reuse. Partial reuse (known to significantly increase the vulnerability of users [16–18], while being poorly understood [5, 21]) was named explicitly for participants encountering one set of conditions, while a simple description of exact reuse featured in an alternative set of conditions.

V3: (Remediation) – The third variable again responds to the best practices developed by Golla et al. [21], affording the opportunity to validate the past observation that effective password reuse notifications should force users to accept the most appropriate corrective action. Study participants were presented with conditions in which they were directed to use a password manager to generate a strong, unique password to replace a weak, reused one. In another set of conditions, this instruction was omitted, and participants were simply told to change the affected password.

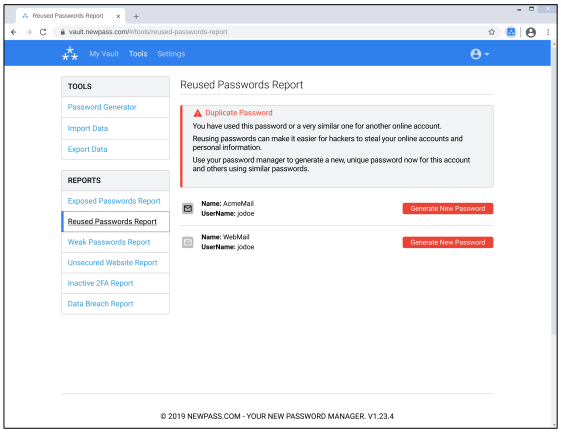
With variables and conditions established, I proceeded to create eight high fidelity prototype images to present to survey participants. The prototypes were designed to be maximally acceptable to participants, and embody elements observed, captured, and coded during my survey of market-leading password managers. Each of the applications included in my survey engaged users principally via the browser; either

[3]While all of the password managers surveyed featured dynamic alerts and dialogues that would interrupt an enrolment or account set up task, inviting users to generate long random passwords when the `autocomplete == "new password"` attribute is detected, only LastPast generated a similar alert when a reused password was being submitted as part of a login attempt.

leveraging plugins, extensions, and web applications [41–45] or a browsers' own built-in features [46, 47]. From the perspective of the user, this was a distinction without a difference. I argue that seriously considering password managers as a means to communicate password reuse results in addressing both password managers and browsers as communication channels simultaneously. As a result, each of my prototype images features a browser frame for context. I selected a current, default Google Chrome theme to enhance validity. At the time of my experiment, Google's Chrome browser commanded greater than 60% of the global browser market share, with the next most popular applications holding only about 17% [71–73]. I first engaged in a rapid prototyping effort, using Chrome's Devtools [74] to manipulate and combine existing elements presented via the browser to password manager users during login and security audit tasks, capturing my results as screenshots. The resulting images were used in my first pilot studies, and would inform the development of my final, high fidelity prototypes. The high fidelity prototypes were produced using the vector-based illustration, development, and prototyping tool Figma [51]. Figma is a tool commonly used in studies conducted by researchers in the fields of human computer interaction, and usable privacy and security. It can output a variety of high fidelity formats, provide for user interaction.



Condition 0
Login Alert - Partial - Generate



Condition 4
Security Audit - Partial - Generate

Figure 3.5: Model Experimental Conditions
This figure depicts a pair of representative prototype images.

The final resulting prototype images are of a very high quality. The prototypes represent a synthetic state-of-the-art, selecting and combining factors from my survey of password managers. They include design elements, use pallets and icons, common across modern web applications. To avoid priming and limit response bias, all notifications feature generic branding. While the prototypes reflect some of the design choices made by real-world providers of password managers and webmail applications, the visual branding presented is that of two fictional companies: *NewPass* and *AcmeMail*. The prototypes used in my experiment supported usability (I wrote scripts to ensure survey participants could zoom in on prototype elements and the text of the reuse notifications, with no loss of image quality), and guarded validity.

Figures representing the model password reuse notification as it appeared to participants experiencing Condition 0 and Condition 4 are presented in this chapter (Figure 3.5), and all prototypes are reproduced in an appendix to this thesis (Appendix C: Prototypes).

3.3.3 Participants

Study participants were recruited via Amazon's Mechanical Turk (MTurk) service [53]. This is a practice common in the fields of human computer interaction, and usable privacy and security, where data from online survey tasks is frequently used to validate frameworks, scales, and test prototypes. It is an expedient and affordable tool for recruiting a large and reasonably diverse sample, increasingly leveraged in usability research [75]. Avoiding a convenience sample drawn from university populations, I used MTurk to recruit participants from a sampling frame exactly matching that of past researchers [21]. Like Golla et al., I recruited adult (18 years of age or older) Mechanical Turk workers, located in the United States, with a 95%+ task approval rating. Matching these factors was an important measure in guarding the validity of data and preserving the opportunity to compare experimental observations against those previously published.

I recruited a total of 297 study participants, observing established practices from the fields of human computer interaction, usable privacy and security, and other human-centred disciplines to ensure my experiment was conducted ethically, my participants compensated appropriately, and that my data would be of sufficient quality to support

the comparisons and statistical tests my research required [75–79]. In place of a traditional recruitment notice, I published a task as a "requester" via MTurk's Human Intelligence Task (HIT) system. The HIT was designed to closely align with current best practices for conducting ethical research with human subjects via Amazon's services, to avoid coercion or compromise, and preserve choice. The full text of the HIT is reproduced in an appendix to this thesis (Appendix A: HIT).

I hewed as closely as possible to the guidelines published by Danielle Navarro of the University of New South Wales [78], an expert in conducting ethical research using Mechanical Turk[4]. I also attended to practices outlined by the Committee for the Protection of Human Subjects, at the University of California, Berkeley [79], who have published documents concerning how to align MTurk research with U.S. Department of Health and Human Services requirements [80]. Active and informed consent was required as a condition of participation in the survey experiment, and participants were instructed that they could end their survey at any time, without forfeiting a promised honorarium. Each participant was paid $1.00 Canadian in recognition of their contribution to the research effort, as well as being entered into a draw to receive one of three bonus payments of $50.00 Canadian. After submitting a unique identifier (their Mechanical Turk Worker ID) at the beginning of the survey task, all participants received a completion code that could be used to request their payment and be entered into the draw. There was no special monetary incentive to complete the survey. The promise of an up-front honorarium disincentivized participants from implementing a "satificing" strategy [58] in completing the survey task or rushing through the survey to secure a payment with minimum effort.

The thesis experiment was conducted over the course of one week in January 2020. 38 distinct batches of human intelligence tasks were generated programatically using the MTurk service, to provide for more granular management of participants, and ensure that requests for honoraria were approved efficiently. All communications with survey participants, as well as the payment of honoraria, were mediated by Amazon's services, with persistent random identifiers (Mechanical Turk Worker IDs) passed

[4]I was fortunate to have the opportunity to correspond with Dr. Navarro when designing my study, to discuss ethical issues not previously addressed in the literature. In particular, Dr. Navarro helped to address the possible impacts of providing survey completion codes to participants in advance of a research task.

to the survey software and shown to me via MTurk dashboards. Multiple passive methods were used to prevent repeat participation in the survey experiment while preserving the anonymity of participants. IP addresses and other sources of personally identifying information were discarded by the LimeSurvey software. A first-party cookie, JavaScript embedded in the source code of the HIT published, a custom MTurk "qualification", and the Mechanical Turk service's own defaults were leveraged in combination to prevent any attempted repeat participation. Each new participant to load the survey site was assigned a unique number by the LimeSurvey software. A counter was incremented, and a script using a simple modular operation assigned participants to one of the eight experimental conditions in round robin fashion. Given the method of recruitment, this pseudo-random assignment to conditions is considered sufficient to guard the validity of data.

3.4 Statistics

The survey instrument was designed to generate both quantitative and qualitative data. Questions capturing participant responses via a 5-point Likert scale resulted in ordinal data, while most others were of a multiple choice format, and resulted in discrete, categorical measures. Both of these kinds of data are amenable to analysis by logistic regression. To test whether responses to these questions varied significantly by condition, relative to the password reuse notification encountered, I generated a series of regression models. I used a pair of similar statistical software packages to support my analyses: GNU PSPP [65] and IBM SPSS [66]. GNU PSPP (a mature, free software statistics suite) was used to calculate frequencies, descriptive stats, and perform tests establishing the independence of categorical variables. IBM SPSS (a more full-featured tool) was used to fit ordinal and multinomial regression models, and to test the assumption of proportional odds (ensuring the validity of ordinal models).

Before performing statistical regressions, I calculated frequencies and generated descriptive statistics to explore the overall shape of the data collected. The goal was to ensure that all of the assumptions for the tests I was running were met. I then calculated coefficients of correlation to establish that participant groups were similar enough in their composition to compare the responses of their members across conditions. This could be assumed, with a large number survey participants pseudo-randomly

assigned to condition groups, but I wanted to ensure there was no concentration of any potentially confounding subject variable before beginning my analysis proper. These statistics could in effect provide an early indication of multiple collinearities that might in turn impact the validity my regression models. I calculated a Pearson Chi-Square statistic for each combination of independent (predictor) variables to establish their independence.

Responses to multiple choice questions yielded sets of binary outcomes (e.g. participants identified password reuse as the cause of a notification or they did not). Simple analyses via binary (binomial) logistic regression could be performed, with the notification condition treated as the independent variable. Ordinal logistic regression models were similarly calculated where the dependent variable being measured took a range of values (e.g. participants identified the level of concern a notification elicited or how quickly they might act to take corrective action after receiving a notification via Likert scale).

My research is preliminary, and exploratory in nature. I preformed multiple statistical tests on the data, in an attempt to identify which factors most strongly influenced participants' responses, and this required an acknowledgement of the risks associated with taking multiple measures and testing multiple hypotheses. Similar to Golla et al. [21], I chose to set $\alpha = 0.05$ and apply a Holm-Bonferroni correction [81], as a way to reduce the likelihood of type I errors and reduce the family-wise error rate (FWER).

I favoured simplicity in the regression models generated, attempting to isolate factors before reporting significance. In an approach closely reflecting that adopted by past researchers [21], a variety of models were considered, and factors eliminated to arrive at the most parsimonious explanation possible. These multinomial models took into account the independent variable of the notification condition, the individual variables adjusted by condition (**V1**, **V2**, **V3**), as well as covariant subject variables such as experience with account compromise, data breach, the use of password managers. All independent variables (experimental and subject alike) were treated as categorical. In the case of the experimental variable, the seventh and final notification condition was always used as a reference, the number of respondents encountering each condition being approximately equal. In the case of each subject variable, the commonest measure

became the reference value when calculating coefficients of regression and odds ratios. I sought to isolate the notification condition as the most meaningful factor, by testing null models absent this variable and comparing p-values with those including it. Only in cases where the notification was itself significant could the significance of related factors (subject variables) be meaningfully reported. As in past studies, I share both p-values and odds ratios where significant relationships are indicated by the regression models.

The coding of participant responses to open-ended, short answer questions posed a challenge. Absent a team to collaborate in the coding process, and without other expert readers to reinforce or moderate my observations, I looked for recurring themes in qualitative data. When possible, I mapped these to existing frameworks for understanding users' relationships to passwords and authentication technologies, attending particularly to the usability benefits outlined by Bonneau, et al. [3]. This prior work helped name common themes and concerns related to effort, scalablity, recovery and lock-out expressed by respondents with regard to their likely responses to password reuse and prescribed mitigation. As in prior research [21], I report only on recurring themes present in multiple responses captured. While qualitative responses provide an opportunity to explore some of the rationale offered by study participants for the attitudes they report, they are of limited explanatory value in this thesis. A more thorough coding exercise in advance of future publications will improve the predictive and explanatory value of this kind of data.

Chapter 4

Results

This chapter presents the results of my thesis experiment. A brief, lay summary provides an introduction to the overall shape of the data collected and the broad significance of observed patterns (4.1 Summary). Demographic and sample-related statistics are then described in detail to better identify study participants (4.2 Sample). Data reflecting differences in the perception and comprehension of password reuse notifications, related to experimental conditions and subject variables, as well as interactions between these factors are fully analysed (4.3 Perception and Comprehension) to lay the groundwork for this work's final discussions and the presentation of my conclusions.

4.1 Summary

My thesis experiment resulted in a total of 297 completed surveys. Data from this pool of respondents was collected between 14 and 22 January 2020 via an online survey tool. A conservative approach to recruitment and the administration of the experiment resulted in only a few participants at a time interacting with the survey site during early phases of the study, so the tool could be observed as load increased, and minor course corrections could be made to improve the experience of study participants. The goal was to make the survey instrument as effacing as possible, so that participants could complete their tasks without encountering issues related to lag or misbehaving scripts. Two small changes were made over the course of the experiment that are not likely to have impacted the data collected significantly. On the first day of the survey, a single typo was corrected, and the validation of Mechanical Turk Worker IDs (not part of the survey proper) was relaxed slightly to reflect greater than anticipated variation in the length of valid ID strings[1].

[1]While the structure of MTurk worker IDs is fixed, they are not uniformly 14 characters as some sources suggest. Future researchers should take this into account when validating strings via scripts employing regular expressions or similar methods.

Following the data collection phase, I performed an early, cursory analysis of the responses captured. The purpose of this analysis was to identify and discard data that appeared unsuitable for inclusion in statistical analyses, and likely to impact the validity of observed trends. The survey site was highly instrumented. The first factor considered during the weeding out of unsuitable responses was related to timing: participants' overall time spent interacting with the survey site. All data from survey tasks completed more than one standard deviation faster than the mean completion time were discarded. I made a reasoned decision to retain responses associated with surveys that took significantly longer than the mean completion time, provided they passed other tests. To do otherwise might have unduly biased results by removing participants from the study who might have taken longer due to some subject variable (measured or unmeasured), and harmed the validity of results. After removing data generated by participants determined to have spent too little time interacting with the survey and prototypes to have generated usable data, I considered the content of responses. I discarded data associated with surveys that were incomplete; where there was obvious disagreement between a participant's responses (e.g. responding that taking action in response to a notification would be a "Very high priority", but indicating that action would be taken "After a week".); where responses to open-ended, short answer questions were nonsensical, off topic, or likely submitted via script[2].

The responses of 260 participants were retained for analysis. This translated to >=30 participants per experimental condition, and a sample similar (in terms of the distribution of demographic and subject variables) to that accessed by past researchers [21]. The groups of participants assigned to each of the study's experimental conditions were not significantly different from one another (i.e. participants encountering the first prototype were no more likely to have experienced data breach or used a password manager than those seeing the next). Finding neither correlations nor collinearities between independent variables, and being confident that both sampling and assignment resulted in data suitable for comparison and the fitting of regression models, I began an in-depth analysis of experimental data.

An observation of central importance to my thesis, and speaking directly to the first

[2]In at least three cases it seems likely that an automated tool was used to complete and submit the survey. Short answer responses appear to be pasted text related variously to mobile push notifications, mobile messaging applications, accessibility.

research question (**RQ1**), is that participants in this study appear to have been better able to identify password reuse than past respondents encountering reuse notifications related to data breaches [21]. A related observation is that study participants reporting an intention to change affected passwords also appear to express a better understanding of meaningful mitigation strategies than in prior experiments. I found little evidence of variation in participants' responses that could be definitively linked to the notification conditions they encountered, but was able to isolate factors associated with the individual variables adjusted by condition (**V1**, **V2**, **V3**) that appeared to have an impact on participants' reported attitudes and understanding. While no statistically significant effect could be measured isolating the notification factor or its components as the primary drivers of observed difference, I did find that some subject variables (in particular experience with data breach) resulted in participants appearing more receptive to password reuse notifications. I was able to observe some interesting trends, making observations of potentially practical significance concerning notification factors which tended to increase concern (communicating risk and improving validity).

My results, and the statistical tests informing them, are further unpacked in the sections that follow. A detailed presentation of descriptive statistics and regression tables is provided in an appendix to this thesis (Appendix D: Detailed Statistics)

4.2 Sample

I accessed a sample substantially similar to that engaged by past researchers [21], using Amazon's Mechanical Turk service [53] to recruit study participants from the same sampling frame in order to facilitate comparisons between my own observed results and the those described in prior work. Participant responses to demographic questions and questions capturing subject variables show little deviation between the two study groups for the most part.

Of 297 total participants, the responses of 260 were retained for analysis. The great majority of these participants were between the ages of 25 and 44 (70.4%), with 7.7% younger and 21.9% older. There was a modest contrast between the sample I accessed and that of past studies [21], with just 40.8% of respondents identifying as female (a difference of about 8%). Similar to previous samples, more than one half (52.7%) of all participants had completed two- or four-year post-secondary degrees, and about one

third (31.2%) reported an education or employment in fields related to IT, computer science, or engineering that could reasonably proxy technical expertise.

A quarter (26.2%) of participants reported that their online accounts had been accessed by someone else without permission. Of those respondents, 10 knew the person who compromised their account(s) personally, where the remaining 58 did not. Better than one third (35.4%) of participants had prior experience with data breaches, either having been notified by account providers that data had been compromised or investigating themselves, using auditing tools and online services like haveibeenpwned [82]. This is a significantly reduced proportion respondents in comparison to the 53.2% figure reported by Golla et al. [21], and is further examined in the following chapter (§5 Conclusion).

Finally, 34.6% of participants reported being either past or current users of password managers. Market-leading third-party password managers (i.e. 1Password, Bitwarden, Dashlane, Keeper, Lastpass) were all represented. Of those participants who reported using password managers, 36 reported using their browser's built-in password manager (with most mentioning either Chrome or Firefox by name). Password manager use was not captured or considered by past researchers, nor were the attitudes and behaviours of participants regarding passwords as addressed by questions from the Security Behavior Intentions Scale (SeBIS) [61, 62]. 70.8% of participants reported creating passwords that exceeded the minimum complexity required by the sites they used, with 65.4% claiming to include special characters even when not explicitly required to by a password policy. Greater than three quarters of respondents (75.8%) said they used different passwords for different accounts, and 31.5% reported changing their online account passwords more often than services required.

4.3 Perception and Comprehension

Participants successfully identified password reuse. Study participants were instructed to choose from among ten potential causes of the notification they encountered in the scenario-based survey, selecting all that applied. The question closely reflected one included in the previous experiment of Golla et al. [21]. Past researchers found that only a minority of respondents could correctly identify password reuse as the cause of the notifications they were presented with. Golla et al. saw just 44.7% of

participants across conditions able to name reuse as having resulted in the messages they saw, with that number rising modestly to 57.9% for the most effective notification tested. In contrast, 92.7% of those taking part in my study identified that "[reusing] the same or similar passwords for multiple online accounts" was the root cause of notifications, regardless of experimental condition. A number of conditions were more successful, with Condition 3 (*Login Alert - Exact - /*) resulting in 100% of respondents naming password reuse as the cause of their having received the notification. Condition 6 (*Security Audit - Exact - Generate*) was the poorest performer in this regard, with just 83.3%.

The proportion of participants selecting password reuse as the cause of the notification they encountered in my study did not vary meaningfully between experimental groups (by notification condition). A binary logistic regression model could not be fit, and notification group had no real predictive value. The operationalization of **V3** (*Remediation*) appeared to make some impact, and approached significance when the experimental variables when considered in isolation. Participants seeing notifications instructing them to "change" affected passwords were more likely to select password reuse as the cause of the notification, but not reliably.

Very few participants identified causes that were not in principal aligned with password reuse. Fewer than 5% of respondents misidentified the notification they saw as being related to a hacking attempt (4.6%) or number of allowed login attempts exceeded (2.3%). A minority of participants (13.1%) thought the notification seen reflected a "weak" password that needed to be changed. Importantly, only eight study participants, just 3.1%, thought that any of the prototypes could be a false positive message, checking "NewPass showed this notification by mistake" as one of their selections.

Though the notification condition itself did not significantly impact participants identification of password reuse, the overall result is promising. When weighed against the findings of past researchers [21], my experimental results provide a strong indication that notifications delivered via the password manager, and explicitly referencing "duplication" and "reuse", might better communicate password reuse than state-of-the-art messages originating with individual online service providers.

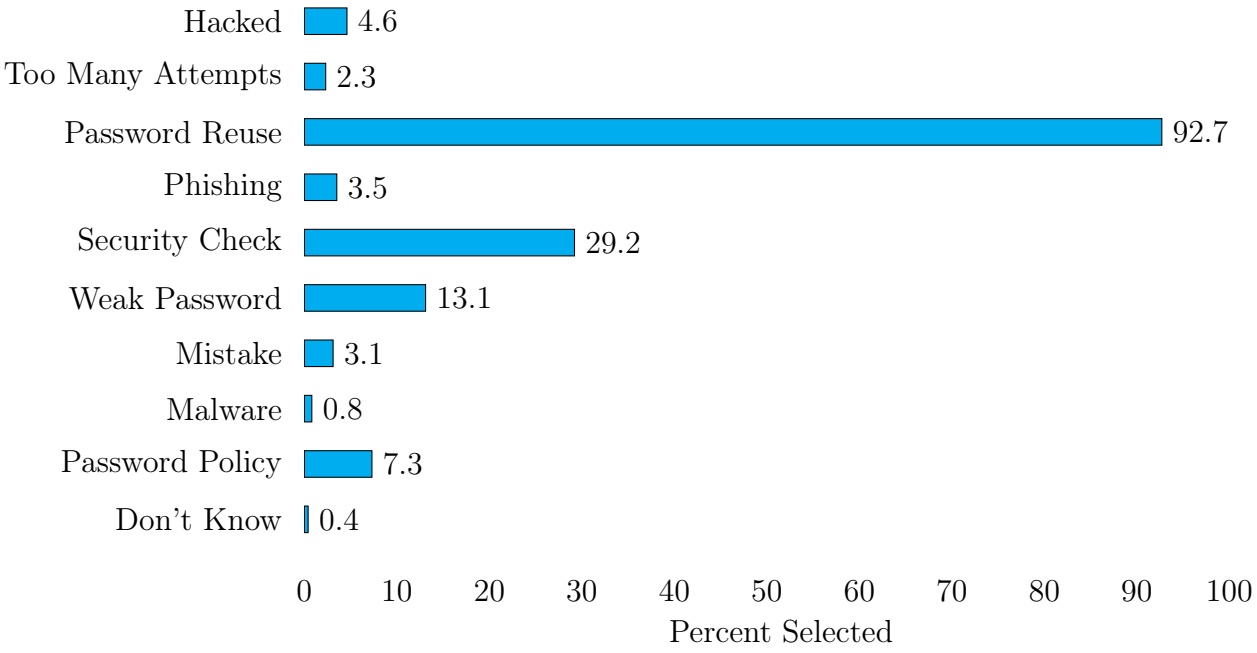


Figure 4.1: Cause of Notification

This figure depicts participants' understanding of what may have cuased them to see the notification. Ten potential causes were presented, and respondents were ask to select all that they felt applied. The percetage of participants selecting a given cause is reproted across all conditions.

4.4 Changing Passwords

Participants intended to change reused passwords. Study participants were asked what they would do if they saw the experimental notification about a "frequently used email account [they] had with a real company" (the fictional, *AcmeMail*). Three options were presented, of which respondents could select only one. Better than three quarters (78.1%) of respondents across conditions indicated that they would change an affected primary password. In response to a second, similar question about what they would do about other affected accounts reusing the same password, a cumulative majority (81.2%) indicated that they would change these passwords as well.

Neither set of responses was observed to vary significantly by experimental group. Numbers remained high, regardless of the notification condition encountered. No multinomial logistic regression model could be fit. Notification group had no real predictive value, nor did any of the experimental variables when considered in isolation.

Likewise, no subject variable could be identified as meaningfully contributing to respondent's decisions to change affected passwords.

The responses to questions about changing affected passwords to address reuse appear to vary markedly from those captured by past researchers. Golla et al. [21] reported that greater than 90% of their participants said they would change a primary password in response to a reuse notification from an online service provider. Just 35.2% of the subjects in the study they conducted would change related passwords used for other accounts. Some possible reasons for these observed differences are explored in the following chapter (§5 Conclusion), and the likely implications discussed.

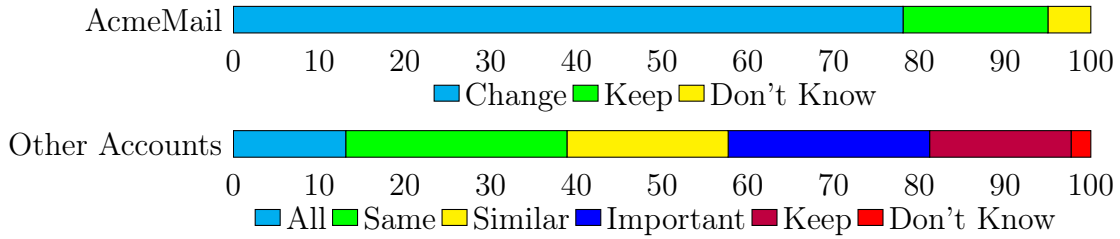


Figure 4.2: Intention to Change Passwords

This figure depicts participants reported intention to change passwords for *AcmeMail* ("a frequently used email account") and for other accounts. A variety of options were presented for managing the credentials associated with other accounts: participants could elect to change all of their other passwords, passwords the same or similar to their *AcmeMail* password, to change only important passwords, or to keep their passwords unchanged. Cumulative percentages are reported.

Participants would change reused passwords in ways likely to improve security. Study participants who indicated that they would change a password were presented with follow-up questions not visible to those who intended to ignore the notification they saw and keep their affected passwords the same. In both the case of *AcmeMail* and other accounts, study participants were instructed to select one of five options indicating what sort of a strategy they would implement when replacing passwords. A related and significant observation is that those respondents who identified that they intended to change passwords, would largely do so in ways likely to enhance their security and limit password reuse. In the case of the primary account (*AcmeMail*), 54.9% reported that they would change their password to "something completely unrelated to the old password" (explicitly eschewing minor modifications to

existing strings or exchanging an affected password for one associated with a different, existing online account). Another 32.7% responded that they would change their password, using a new one "generated by a password manager or browser". For other accounts, the story was much the same. 54.3% of participants responding to this question would select a completely unrelated password as their preferred replacement, with 30.5% identifying an intention to use a password manager or browser to generate a new password.

Participants' responses to questions about how they would change reused passwords appear to vary in meaningful and measurable ways. Again, there were no multinomial regression models that could be fit to predict outcomes based on the notification condition, nor individual variables. The operationalization of **V3** (*Remediation*) gave the impression of having some effect when analysing participants' strategies for changing their primary password (*AcmeMail*) and considering the experimental variables when considered in isolation. Those instructed to use an application to "generate" a new password appeared less likely to report a remediation strategy incorporating reuse, but this trend did not raise to the level of statistical significance. This will be unpacked in the following chapter (§5 Conclusion) and discussed in relation to other findings. The subject variable of password manager use was observed to significantly impact reported remediation strategies in both scenarios. For the primary, email account (regression $p = 0.001$), respondents were much more likely to report an intention to generate a new password with a password manager or browser than to simply try to create a unique one if they have previously used a password manager ($p < 0.001$, $odds\ ratio = 3.61$). The results were similar for other accounts as well (regression $p = 0.001$). Those who had previously used a password manger were again more likely to generate new passwords with an application than attempt to create secure ones by other means ($p < 0.001$, $odds\ ratio = 3.77$).

These findings are a departure from those reported from past researchers [21], who found that those reporting an intention to change reused passwords would mostly do so in ways that left them still vulnerable to password reuse attacks. Just 1.4% of past participants claiming they would change the password of an affected primary email account would choose a wholly unique password as a replacement, with 9.7% saying they would use an application to generate a new password. This contrast is

certainly worthy of discussion, especially in light of the overwhelming majority of past participants saying they intended to change reused passwords. It seems possible my experimental results indicate that notifications delivered via the password manager could impact users' comprehension or at least result in greater consideration of the problems and the likely cost of mitigations than state-of-the-art messages originating with individual online service providers.

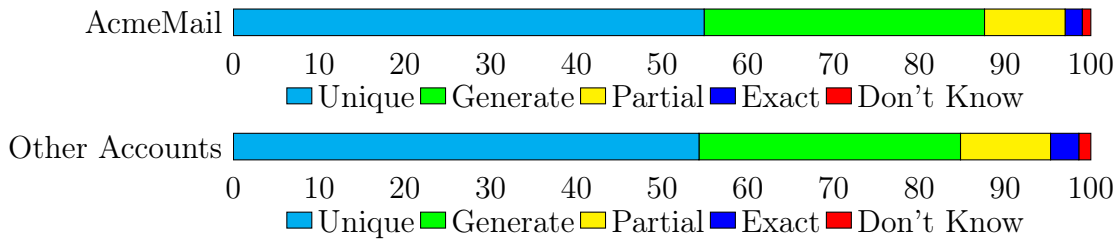


Figure 4.3: Strategies for Changing Passwords

This figure depicts participants strategies for changing reused passwords for *AcmeMail* ("a frequently used email account") and for other accounts. A variety of possible options were presented: participants could change an affected password for something similar to the old one (partially reusing an existing string), create a new password unrelated to the old one, use a password already associated with another online account (exactly), or generate a new password with an applications. Cumulative percentages are reported.

4.5 Considering The Notification(s)

A series of questions asked participants to address their perception of the notification they encountered directly, recording their agreement with statements about the notification and how they might respond to it on a 5-point Likert scale. The resulting data were of particular interest. They reinforce and provide context for other observations, and may also reflect the validity of the prototypes and the overall experiment.

Participants accepted the notifications, and felt the cause was clear. The vast majority of study participants responded positively to questions designed to determine if the experimental notification they encountered was perceived as valid or likely to represent a false positive. Better than half (53.8%) of all participants, across conditions, registered strong agreement with the statement, "I would expect a real password manager to display notifications like this one when necessary", and an additional 36.5% agreed. Respondents further reported that, were they to receive

a notification like the one seen about a real email account, they would "believe that it was accurate or correct". A cumulative 90.7% responded either that they strongly agreed or agreed with this. These results taken together suggest a level of acceptance that could reflect high validity.

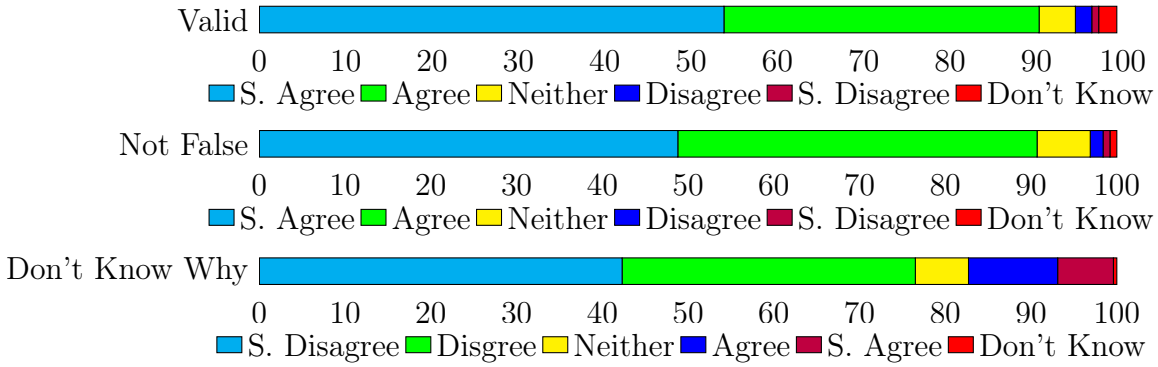


Figure 4.4: Notification Impact and Acceptance

This figure depicts a set of responses to questions concerning the impact of notifications and participants' acceptance of the notifications they encountered. The questions instructed participants to rate their agreement with satements concerning:
· The validity of the notification they saw (would they expect a real password manager to display a notification like this).
· The likelihood of notifications being interpreted as false (would they believe a notification like this to be accurate and correct).
· The likelihood of the cause of a notification being correctly identified (would they be confused and not know why a notification had been generated).

When presented with an instruction to rate their agreement with the statement, "If I saw this notification about a frequently used email account I have with a real company, I wouldn't know why I saw this notification", greater than three quarters (76.5%) of study participants registered some level of disagreement. Respondents were presented with a 5-point Likert scale, running from "Strongly disagree" to "Strongly agree", with both a true neutral option and the option of answering with "Don't know". 42.3% reported disagreeing strongly with the statement, while another 34.2% disagreed. This negatively framed, personalized question reflected the first multiple choice responses identifying the cause of the notification encountered. Interestingly, while an overwhelming majority (92.7% of respondents) successfully identified password reuse as a cause in the earlier question, the responses to this question concerning comprehension and the understanding of cause more closely align with response rates

to questions related to changing passwords and taking action. Again, 78.1% of study participants across conditions indicated that they would change an affected primary password.

The proportion of participants registering disagreement with the statement about not knowing why a notification was presented did not vary meaningfully between experimental groups (by notification condition). Notification group had no real predictive value, nor did any of the experimental variables when considered in isolation, and no ordinal logistic regression could be made to fit the data. The subject variable of experience with data breach came close to having a significant impact on participants responses, falling just short. Respondents who reported past experience with a data breach appeared more likely to understand the cause of a notification, regardless of the prototype they encountered.

Most respondents understood the notifications. Study participants were asked to report whether or not they agreed that the notification to which they were exposed explained what was going on with primary email account (*AcmeMail*) adequately. Responses were again recorded via a 5-point scale ranging from "Strongly disagree" to "Strongly agree", with neutral and "Don't know" options. 45.4% of participants registered agreement with the statement, 31.9% strong agreement. These numbers are inline with the earlier observation that relatively few respondents misidentified the cause of the notification they encountered, believing that the email account had been hacked (4.6%) or that the maximum number of login attempts had been exceeded (2.3%).

The proportion of participants responding in the affirmative varied little across conditions, with no statistically significant effects associated with the notification condition nor even with the experimental variables when they were considered individually and treated as independent predictors in ordinal regression models. Only participants' past experiences with account compromise were found to trend toward having a meaningful impact on outcomes. Though the omnibus statistic did not suggest it was a strong indicator, those reporting unauthorized access to online accounts did appear somewhat more likely to view the notifications as informative.

The observation that experience with data breaches appears to impact the perception of password reuse messages, reflect the results of past researchers [21] quite

closely, and is thoroughly discussed in the following chapter (§5 Conclusion).

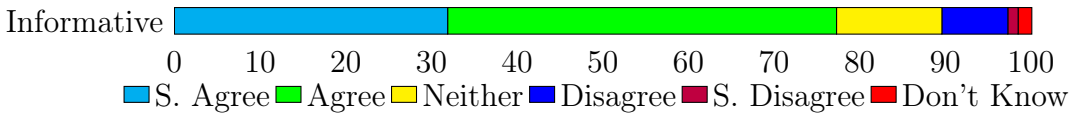


Figure 4.5: Clarity of Notification Content

This figure depicts participants' agreement with the statement, "This notification adequately explains what is going on with my email account."

Most participants could fix the problem based on the information provided. Respondents were asked to report whether they felt the notification explained to how to resolve the problem it communicated. They registered agreement with a positively framed statement via a standard 5-point Likert scale. An overwhelming majority expressed agreement with the statement, "I feel that the notification explained to me how to resolve the situation.", with 44.6% selecting "Agree" and nearly half (47.3%) selecting "Strongly agree". This result again runs interestingly parallel to data collected early in the survey, concerning participants' understanding of what caused them to see a notification, and bears further discussion.

These responses did not vary significantly with the notification condition, and no ordinal logistic regression model could be made to fit the data. While not rising to the level of statistical significance, the operationalization of **V2** (*Reuse*) did appear to have some effect. Respondents exposed to notifications making reuse explicit, and mentioning the even "very similar" passwords increased risk, seemed more likely to find a notification informative and to report knowing how to resolve the problem of password reuse being communicated.

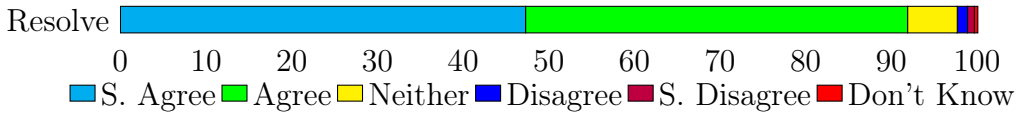


Figure 4.6: Mitigations in Notification Content

This figure depicts participants' agreement with the statement, "I feel that the notification explained to me how to resolve the situation." Respondents were instructed to rate their agreement with this statement on a 5-point Likert scale to measure the success of the experimental notifications in communicating information about changing passwords and managing password reuse.

Participants would take action, but were not sure how concerned to be. Participants were asked to rate their agreement with the statement, "I feel that ignoring this notification would not have any consequences." Their responses were captured via a 5-point Likert scale, running from "Strongly disagree" to "Strongly agree", with both a true neutral option and the option of answering with "Don't know". The result was an almost even split between respondents expressing concern (13.8% strongly disagreed with the statement, while another 36.9% disagreed) and those who were either uncertain of the consequence of ignoring a notification or indifferent to it. More than a fifth of my study's participants (21.2%) recorded a neutral response of "Neither Agree nor Disagree".

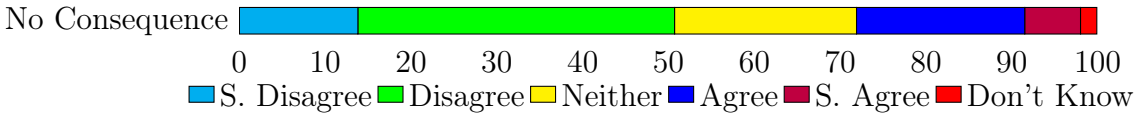


Figure 4.7: Consequences Associated with Ignoring Notifications
This figure depicts participants' agreement with the statement, "I feel that ignoring this notification would not have any consequences." Respondents were instructed to rate their agreement with this statement on a 5-point Likert scale as part of a series of questions measuring participants' acceptance of the experimental notifications and their likely impact.

Participant responses did not vary significantly with the notification condition, and no ordinal logistic regression model could be fit to the data. However, the operationalization of **V2** (*Reuse*) did approach significance, and appear to have an impact on participants' level of concern and perception of consequence when the experimental variables when considered in isolation. Notifications informing respondents they reused a password exactly across multiple online accounts were more likely to believe that ignoring the notification could have serious consequences, though the link was not definitively, statistically significant. Only the subject variable of experience with data breach was observed to be a strong predictor of whether or not participants reported feeling concerned about a notification (regression $p < 0.001$). Experience with breaches resulted in respondents being much more likely to believe that inaction would have consequences ($p < 0.001$, $odds\ ratio = 2.33$).

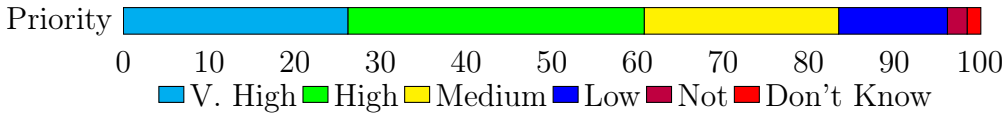


Figure 4.8: Intention to Act on Notifications

This figure depicts participants' completion of the statement, "For me, taking action in response to this notification would be a:" Respondents were instructed to select a phrase from among five scaled possiblities ranging from *Not a priority* to *Very high priority*.

The possible uncertainty communicated by respondents' assessment of the consequences associated with the notifications needs to be reconciled with their ability to correctly identify reuse as the cause of notifications, and their reported intention to replace reused passwords. Participants' willingness to take action in response to the notifications was again confirmed by asking them to rate how much of a priority responding to the notification would be for them. This question immediately followed the request for them to consider the likely consequences of inaction, and observations were in line with previous measures, but showed some possible moderation. A cumulative 60.8% said responding to the notifications would be a *High* or *Very high* priority, with another 22.7% setting it as a *Medium* priority. These results are considered in tandem and unpacked in light of explanatory responses to open, short answer questions in the following chapter (§5 Conclusion).

Chapter 5

Conclusion

This chapter presents my conclusions. A final accounting of my thesis experiment and discussion of observed results is provided, elaborating upon the likely value of my findings in context (5.1 Discussion). The concrete contributions of this work are outlined, and both benefits and limitations are addressed (5.2 Benefits and Limitations). In closing, likely avenues for further research are explored (5.3 Future Work).

Through a detailed consideration of my experimental results, I will paint a clear picture regarding the importance of my specific observations, and their potential for guiding future research in the fields of usable privacy and security, and authentication.

5.1 Discussion

The results of my thesis experiment are, on the whole, encouraging. Participants in my study appear to have been able to identify password reuse more readily than the respondents of Golla et al. [21], who encountered state-of-the-art reuse notifications related to data breaches. Study participants were additionally observed to express a better understanding of how to change reused passwords effectively to limit their vulnerability to attack than in past experiments.

5.1.1 Demographics and Subject Variables

Before the potential significance of these findings can be seriously discussed, some of the observed difference between my sample and that of past researchers must be addressed. I made every effort to access a group of participants substantially similar to that of Golla et al. [21], using Amazon's Mechanical Turk [53] to recruit from the same sampling frame, and observing best practices for engaging these online workers. Despite this, some potentially important differences between the two subject groups were observed. In particular differences in the reported genders of study participants, and their past experience with data breach need to be addressed.

While there was some difference in the proportion of respondents identifying as female (40.8% in my study v. 48.4% in the study of Golla et al.), gender identity was not found to have any significant predictive value in analysis or that of past researchers [21], and was ignored in my own statistical tests. This factor is very unlikely to account for observed differences in the dependent variables tested, and the variation in sample populations appears to simply reflect fluctuations in the overall demographics of Mechanical Turk workers during the times when the two studies were conducted [83].

In contrast, whether or not participants reported past experience with data breaches was observed to have a significant impact on the level of concern notifications elicited, and was found by past researchers to have a significant impact on respondent perceptions of notification validity [21]. The difference in the proportion of respondents reporting this kind of experience (35.4% in my study v. 53.2% in the study of Golla et al.) might account for some of the differences between my results and those of past researchers. It is also possible however, that study participants simply responded to the question capturing experience with data breach differently. While the exact same question was posed in both studies (it followed the experimental questions and preceded demographic questions about age, gender identity, etc. in both cases), there is reason to think that priming and recency effects could have biased the responses captured by Golla et al., the whole of their experiment having been focussed on reuse messages generated by online service providers in response to data breaches [21]. This possible priming, and the impact of communicating a credible threat related to a breach as opposed to a theoretical one related to local data is more fully explored in the discussion of subjects' reported intention to change reused passwords.

5.1.2 Identifying Cause

One of the most exciting results from my thesis experiment is the observation that study participants were overwhelmingly able to identify password reuse as the cause of the notifications they encountered. The experiment, survey, and sampling frame all closely reflect past experiments, supporting the inference that shifting domains and communicating reuse via a password manager can meaningfully impact user comprehension of password reuse. Participants viewing reuse notifications originating

from a single service provider were seldom able to correctly identify the cause of a warning in past studies [21]. The great majority of my participants (92.7% of respondents, across experimental conditions) could identify password reuse as the root cause of the notifications they saw in the context of a password manager (an application with explicit access to credentials for multiple sites).

There is good reason to accept that the responses participants provided to this question as valid, and not merely a function of the question having been presented early in the survey (shortly after respondents read the scenario and examined the prototype). The responses to Likert scaled questions at the end of the survey task appear to closely reflect the results observed. 92.7% of respondents selected "You reused the same or similar passwords for multiple online accounts." when asked to identify what might have caused them to see a notification. A cumulative 91.9% either agreed or strongly agreed with the statement that the notification they had seem explained the problem they were facing and how to resolve it correctly later in the survey, with similar numbers additionally indicating a high level of acceptance of the prototype images and belief that the warning these prototypes communicated would be accurate and not false. While far from definitive, these observations tend to support one another and add some weight to the suggestion that the notifications presented could have an impact on participants thoughts regarding password reuse in some way.

Though there was no statistically meaningful variation observed in the proportion of participants identifying the cause of password reuse across conditions, some did appear to have more of an impact than others. There were additionally patterns in way the conditions appeared to perform that need to be discussed.

It is possible that Condition 3 (*Login Alert - Exact - /*) outperformed others by chance. The notification condition was not observed to have a statistically significant impact on participants' attitudes or understanding of password reuse in any of my tests. It is notable, however, that this condition, which resulted in 100% of respondents naming password reuse as the cause of their having received the notification, is also a brief and straight forward version of the dynamic alert. This version of the model notification is parsimonious, and eliminates additional information regarding partial reuse that might be interpreted as equivocation or be otherwise misaligned with participants' intuitions about reuse. The addition of the words, "or a very similar one"

could be seen as weakening the notification or communicating uncertainty, and appear to downplay the severity of the issue. This aligns with a possibly practically significant trend showing a reduced tendency to interpret notifications as having consequences when partial reuse was made explicit. Those in which only exact reuse was mentioned, and which instructed participants to simply "create" a new password were also the most readable, with high measures of readability and low grade level equivalent on the Flesch-Kincaid scales [84, 85] (increasing readability by 5 points, and reducing grade level from 10 to 9). This simplicity cannot be discounted as a possible cause of what appears to be an improvement in comprehension.

5.1.3 Resolving the Problem

Most study participants reported that they would change a reused password in response to notifications like the ones encountered during the survey. This was encouragingly true regardless of whether a primary email account (*AcmeMail*) or other related online accounts were considered. 78.1% of respondents said that they would change an affected primary password, and a cumulative 81.2% reported that they would change other passwords as well. This observation contrasted interestingly with the results of Golla et al. [21], who posed an almost identical pair of questions in their own study.

Past participants reported that they would change a reused password for a frequently used email account in much greater numbers (more the 90% selecting this option regardless of the message seen). A difference of more that 10% needs to be unpacked. It is possible that the participants recruited by Golla et al. [21] were better motivated to act by the data breach messages that they were exposed to, and that password reuse messages originating with password manager fair less well by this measure. It is also possible that past participants were more susceptible to a good subject bias, and that this coupled with the desire to present a correct the socially desirable response to a security focussed question influenced the number of positive responses recorded. Neither of these hypotheses seems likely. It has been established that the survey instruments employed by both myself and past researchers were substantially similar, that participants were drawn from the same sampling frame, treated similarly, and that demographic and subject variables were proportionally similar. This wold

tend to discount the idea that one group would be much more susceptible to bias than the other. I also think it unlikely that stand-alone breach messages were more effective at communicating password reuse or encouraging participants to report an intention to act to mitigate the problem identified. A consistent proportion of my respondents (around 80%) said they would change passwords for all affected accounts. The participants of past researchers did not appear to understand that additional action was required to improve security, and that the passwords of other affected accounts also had to be changed. Only about a third (35.2%) of these participants registered an intention to change any of their related passwords, where 81.2% of mine were prepared to do so.

I see the consistency in my participants' responses as a positive indicator of increased comprehension. They also demonstrated a much better understanding of how to change their passwords to effectively address the risks associated with password reuse. Participants in my study who indicated that they would change password identified mitigation strategies incorporating wholly unique new passwords and using password managers and browsers to generate secure random strings overwhelmingly, whereas the respondents of Golla et al. largely reported an intention to make modifications that simply amounted to more reuse [21]. The proportion of respondents indicating an intention to change all of their affected passwords in the early questions presented in my survey also runs closely parallel to the number of participants indicating that experimental notifications were informative (explaining what was going on with the primary email account) and that they would know why a notification was presented if encountered in the wild. All of these results considered in combination tend to support a greater level of confidence in the validity of the observed response to questions about changing passwords, indicating strong comprehension.

I am inclined to view my study's participants as rational. There is evidence in the literature that users can and do make informed choices about when to act on security advice, adopt new technologies or change behaviours, and when not to [11, 14]. Unlike in past studies, my participants were presented with notifications originating with a password manager, an application with direct access to credentials for multiple online accounts. The notifications encountered during the study could have been correctly interpreted as indicating increased risk without communicating any information about

the likely availability of password data to malicious parties and potential attackers. In effect, participants choosing not to indicate an intention to change their passwords could have been refusing to act on what they perceived to be a theoretical threat. The participants of Golla et al. [21] saw notifications explicitly related to data breaches, making the risk communicated potentially more concrete.

The responses to open, short answer questions, tend to support the inference that my participants were acting (or believed themselves to be acting) rationally when refusing to change reused passwords. Past researchers believed many subjects refusing to change reused passwords to be expressing a kind of belief in their own invincibility, misunderstanding the risks associated with password reuse [21]. My own subjects may be expressing credulous faith in the security of online service providers, but the commonest justification for choosing not to opt to change reused passwords were expressions in the local security of account credentials. Participants referenced their systems for saving and accessing passwords, the use of mnemonics in creating new passwords. One wanted to discuss their "extensive password system" (P88). Participants also expressed a belief that they correctly adhered to other security advice concerning safe browsing, and weighing costs and benefits associated with changing passwords if they were not sure they would be attacked. P56 talked about being "careful online" and a concern that making changes would be a waste of time and energy. The effort associated with changing passwords was repeatedly mentioned when participants enumerated potential costs (aligning with the past findings of Herley and van Oorschot [2] as well as Bonneau et al. [3]). These kinds of responses indicate an imperfect understanding of the overall shape of the threat posed by password reuse, but a good understanding of the content of the experimental notifications, perhaps highlighting something to be better addressed in future models. In contrast, many of those participants that would change passwords appeared to draw on past experience or other security knowledge to justify their own decisions. P13, who had experience with both unauthorized account access (e.g. hacking) as well as data breach, summed up the views of many of these respondents eloquently by likening the chances of being subjected to an attack to playing an "unlucky lotto".

Considering evidence from all of the experimental questions, and allowing for the open responses to provide a moderating effect, there appears to be good reason to

thing that there is some promise in password managers as vehicles for effectively communicating password reuse to users.

5.1.4 The Impact of Experience

I found little evidence to suggest that the variation in participants' responses to the survey questions was influenced by the notification condition they encountered. I did, however, observe that the subject variables of password manager use and experience with data breach had a significant impact in certain cases. Users of password managers were much more likely to report an intention to change reused passwords by generating a new one with a password manager or browser than to simply say they would create a unique one. Participants reporting previous experience with data breaches were inclined to take the experimental notifications they encountered significantly more seriously and interpret them as having greater consequences. In both cases I interpret my observed results as reflecting the common sense inference that experience colours users understanding of reuse notifications. First-hand knowledge of the common features of password managers could be reasonably expected to result in participants selecting the option to "generate" a strong, unique new password. Likewise, whether or not a past data breach could be definitively linked to an account compromise or attempted attack, it is reasonable to assume the experience would have an impact on future perceptions of risk in related domains.

The results of Golla et al. [21] depart slightly from my own in this regard. While both I and past researchers found experience with data breaches to be a variable with significant predictive value, I was not able to fit a regression model for any hypothesis that suggested a technical background (as evinced by education or career) made a meaningful impact[1]. This difference may reflect nothing more than the greater sample size captured in the previous experiments resulting in a more sensitive set of statistical tests. It might be a further indication that the changing of domains from stand-alone data breach messages to notifications associated with a password manager has some impact on the way information about password reuse is received and perceived.

What my results and those of past researchers have in common is that they strongly indicate is that subject variables can and do impact how password reuse notifications

[1]Past researchers did not consider password manager use a factor in their analyses.

are interpreted by users. This seems to align with what some of the most current research in the field of authentication focussed on password managers suggests (that different kinds of users benefit from different kinds of supports when adopting and using password managers) [86], and may have interesting implications for future work.

5.1.5 Revisiting the Research Questions

In light of the results presented in the previous chapter (§4 Results) and discussed here, the research questions which this thesis and its experiment were designed to answer need to be formally revisited.

The two research questions at the heart of my thesis focussed on whether password reuse could be effectively communicated to users via notifications originating with an application understood to have explicit access to multiple credentials, for multiple online accounts. I argue that there is a strong indication that this is indeed the case.
RQ1: My experiment resulted in strong evidence suggesting that password reuse notifications originating with a password manager did indeed appear to have the potential to meaningfully improve user comprehension regarding the problem of password reuse. Participants in my study were almost universally able to identify password reuse and understood how to resolve it, and my experimental results differed markedly from those of past researchers who ran a similar experiment measuring responses to state-of-the-art reuse messages related to data breaches [21].
RQ2: The proposed mitigations shared by study participants further suggested that password reuse notifications originating with a password manager could improve the likelihood of users taking appropriate corrective action, replacing affected passwords with strong, unique ones. In stark contrast to past results [21], participants in my study reporting an intention to change reused passwords would employ strategies likely to enhance security and address password reuse across affected accounts, leaving them less vulnerable to targeted guessing attacks.
RQ3: I saw no evidence to suggest that the channel or task associated with password reuse notifications significantly impacted their effectiveness. Neither the notification condition nor **V1** (*Delivery Channel*) were observed to have any real predictive value. There was some slight clustering of effects around conditions featuring an *Active Alert*, but nothing worthy of reporting. This may be a function of my experiment, and the

limited level of interaction my prototypes could support. The idea of examining the possibility that a dynamic, active alert interrupting a login could uniquely capture users attention and motivate them is still something that could be incorporated into future experiments. Similar approaches are employed by market leading applications [45], and have been long validated in the field of warnings and alerts [69, 87].

While this research is preliminary and exploratory, the results are promising. My observations appear to provide some early validation for the approach of alerting users to password reuse, and communicating associated risks, via the dialogues of password managers. These results reflect the the intuitions of past researchers, and can help others feel more confident moving forward with additional testing, potentially via user studies and more fully interactive prototype applications or plugins.

5.2 Benefits and Limitations

This thesis describes an early effort to explore a problem of current interest and importance to researchers in the fields of usable privacy and security, and authentication. The study I conducted is one of just a handful looking at password reuse notifications and their role in addressing a common vulnerability. It is the first to take up the challenge of examining the efficacy of addressing password reuse and improving user comprehension via notifications originating with password managers.

The goal of this research project was to provide an expedient exploratory analysis of factors. With this in mind, I accepted certain compromises. In a fashion reflecting the work of Golla et al. [21], I gathered a convenience sample of Mechanical Turk workers, collecting data via a scenario-based online. This provided a quick source of data and made possible a number of comparisons that were important to the validity of the research project. It also meant that like past researchers I had to consider the possibility that participants' responses could reflect biases tied to self-reporting and a desire to provide correct the socially desirable answers. My experimental design, recruitment document, and the survey instrument itself were all constructed in an effort to reduce or counteract these biases, and like Golla et al. [21] I chose to interpret my results as reflecting likely attitudes and actions, without perfectly representing them [55]. For a study like the one I conducted, this is sufficient. The most significant trends in the observed data are strong enough to justify the claim that additional,

more nuanced investigation into the applications and factors examined is warranted. These tools were appropriate for a preliminary investigation.

Maintaining an appropriate scope in my thesis research was a challenge. Again, some of the compromises I accepted are only forgivable in light of the early, exploratory nature of the work presented. Rather than duplicating the experiments of Golla et al. [21], I chose to attempt to hold as many variables as possible constant between my own study and that of past researchers, and to report tentative observations based a broad comparison of descriptive statistics. While I have presented a preponderance of evidence in support of the modest claims made in this thesis, it should be acknowledged that my own study and that of the previous research team were conducted at different times, with different samples, and that there are likely many variables that could not be wholly controlled for between these two studies. It is also worth noting that the validity of prototypes was never independently tested in either my own work or that of the team of Golla et al. [21]. Subject responses to questions concerning the acceptance of prototypes provide indications of validity, and in both cases a systematic approach to the survey of state-of-the-art notifications and frameworks was made. There is no guarantee, however, that assumptions and design decisions made by myself or past researchers did not have an impact on the data collected.

The scope of my thesis project (a necessarily independent venture) might have prevented me from capturing more significant results. Realities associated with scale, time, budget, and the fact that I was working independently, resulted in me capturing a smaller sample size than past researchers, being able to run fewer statistical tests, and limited the kind of coding that could be done to aid in the interpretation of open, short answer questions. The contributions of this thesis are quite reasonable. However, particularly when the need to explore multiple hypotheses is considered, a larger sample, and the more powerful and sensitive tests that it could facilitate might have made an impact.

While I was fortunate to have access to a lab group willing to pilot my study, I had to design an experiment that could be completed in isolation. I took steps to limit the possible impacts of my own biases, but was not able to seek the opinions of co-investigators to validate my choices. I was conservative in my approach, opting to discard survey data when observations called into question the reliability of a

respondent. I acted consistently, attempting to avoid bias, but it is possible that some of my own assumptions could have coloured or shaped some of the data retained for analysis. Many of the questions presented in my survey had been previously validated, and in turn reflected best practices. This might reduce the likelihood of my misinterpreting observed disagreement between values recorded for related questions designed to indicate inattention or attempts on the part of respondents to implement a "satificing" strategy [58] in completing the survey task, but the reliability of participants still had to be considered to guard the validity of reported results and observed trends. My own biases might have been better limited if a group had approached the problem of sanitizing survey data cooperatively.

5.3 Further Work

This thesis represents a modest contribution to what will hopefully be a growing discourse around how best to address password reuse while focussing on usability and improving users' comprehension of this complex, cross-site problem. I have presented early results offering tentative validation of past researchers intuition that "ecosystem-level" solutions leveraging password managers and browsers might be worthy of investigation [21], and my observations are in turn well-positioned to inform and direct future work in the fields of usable privacy and security, and authentication.

I have established that password reuse notifications originating with a password manager present a opportunity to meaningfully improve users' ability to recognize reuse, and potentially their comprehension of underlying issues as well. I believe further research, and greater investment is warranted. The approach of communicating reuse via password managers having now received early validation, studies increasing ecological validity should be undertaken. There is reason to believe the that work associated with a user study featuring a more fully interactive prototype would not be wasted. The acceptance of browser-based password managers is high, they are convenient, with few barriers to adoptions and available to most users. Leveraging an open source, browser-based option to create and test a prototype embodying some of the features from this study, and incorporating a model notification similar to the one generated following my own survey could be a meaningful next step. A study of this kind could generate some of its data by directly instrumenting participants devices or

browsers, limiting the impact of biases related to self-reporting, and painting a clearer picture or what impact task and channel might have if any on the comprehension of password reuse and its appropriate mitigations.

A more in depth user study would necessarily be conducted with a relatively small number of subjects. Work like this might be combined with research efforts aimed at better understanding and isolating those factors improving the efficacy of password reuse notifications among sub-populations of users with different needs and experiences. Both I and past researchers [21] observed strong evidence suggesting that the effectiveness of notifications was impacted by subject variables like password manager use and experience with data breach. This is in line with current thinking regarding the need to personalize some aspects of user security. An attempt to isolate those factors that could make password managers more effective tools for addressing reuse by providing personalization, would be timely. Determining how to better leverage users' experiences (how to best help novices, experienced users, those who have never experience an account compromise or even an attempted attack) would address a number of open questions concerning how to best support the users of password managers raised by Lyastani et al. [29] as well as Pearman et al. [86]. It myst be said that any future work aimed at capturing opinions, attitudes, and intentions, should include a significantly greater sample size to ensure appropriate sensitivity in statistical tests.

This thesis has confirmed and validated some assumptions about the role password managers and related applications can play in communicating password reuse to the users of online accounts. In doing so, my work has also help to shine a light on a pair of avenues ripe for further exploration. This is a meaningful contribution to a field and a research community of importance.

Bibliography

[1] D. M. Ritchie and K. Thompson, "The UNIX time-sharing system," in *Proceedings of the fourth ACM symposium on Operating system principles*, ser. SOSP '73. New York, NY, USA: Association for Computing Machinery, Jan. 1973, p. 27.

[2] C. Herley and P. C. van Oorschot, "A research agenda acknowledging the persistence of passwords," *IEEE Security Privacy*, vol. 10, no. 1, pp. 28–36, Jan 2012.

[3] J. Bonneau, C. Herley, P. C. Van Oorschot, and F. Stajano, "The quest to replace passwords: A framework for comparative evaluation of web authentication schemes," *Proceedings - IEEE Symposium on Security and Privacy*, pp. 553–567, 2012.

[4] C. Herley, P. C. van Oorschot, and A. S. Patrick, "Passwords: If we're so smart, why are we still using them?" in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2009, vol. 5628 LNCS, pp. 230–237, iSSN: 03029743.

[5] S. Pearman, J. Thomas, P. E. Naeini, H. Habib, L. Bauer, N. Christin, L. F. Cranor, S. Egelman, and A. Forget, "Let's Go in for a Closer Look: Observing Passwords in Their Natural Habitat," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. New York, NY, USA: ACM, 2017, pp. 295–310, series Title: CCS '17. [Online]. Available: http://doi.acm.org/10.1145/3133956.3133973

[6] B. Ur, F. Noma, J. Bees, S. M. Segreti, R. Shay, L. Bauer, N. Christin, and L. F. Cranor, ""I Added '!' at the End to Make It Secure": Observing Password Creation in the Lab," 2015, pp. 123–140. [Online]. Available: https://www.usenix.org/conference/soups2015/proceedings/presentation/ur

[7] S. Gaw and E. W. Felten, "Password management strategies for online accounts," in *Proceedings of the Second Symposium on Usable Privacy and Security*, ser. SOUPS '06. New York, NY, USA: ACM, 2006, pp. 44–55. [Online]. Available: http://doi.acm.org/10.1145/1143120.1143127

[8] R. Morris and K. Thompson, "Password security: A case history," *Commun. ACM*, vol. 22, no. 11, pp. 594–597, Nov. 1979. [Online]. Available: http://doi.acm.org/10.1145/359168.359172

[9] E. Stobert, "The Agony of Passwords: Can We Learn from User Coping Strategies?" in *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '14. ACM, 2014, pp. 975–980, event-place: Toronto, Ontario, Canada.

[10] R. Shay, S. Komanduri, A. L. Durity, P. S. Huh, M. L. Mazurek, S. M. Segreti, B. Ur, L. Bauer, N. Christin, and L. F. Cranor, "Can long passwords be secure and usable?" in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '14. New York, NY, USA: ACM, 2014, pp. 2927–2936. [Online]. Available: http://doi.acm.org/10.1145/2556288.2557377

[11] C. Herley, "So Long, and No Thanks for the Externalities: The Rational Rejection of Security Advice by Users," in *Proceedings of the 2009 Workshop on New Security Paradigms Workshop*, ser. NSPW '09. New York, NY, USA: ACM, 2009, pp. 133–144, event-place: Oxford, United Kingdom.

[12] A. Adams and M. A. Sasse, "Users Are Not the Enemy," *Commun. ACM*, vol. 42, no. 12, pp. 40–46, Dec. 1999.

[13] F. Alaca and P. C. van Oorschot, "Comparative Analysis and Framework Evaluating Web Single Sign-On Systems," *arXiv:1805.00094 [cs]*, Apr. 2018, arXiv: 1805.00094.

[14] M. Fagan and M. M. H. Khan, "Why do they do what they do?: A study of what motivates users to (not) follow computer security advice," in *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*. Denver, CO: USENIX Association, 2016, pp. 59–75. [Online]. Available: https://www.usenix.org/conference/soups2016/technical-sessions/presentation/fagan

[15] E. M. Redmiles, E. Liu, and M. L. Mazurek, "You want me to do what? a design study of two-factor authentication messages," in *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*. Santa Clara, CA: USENIX Association, 2017, pp. 1–7. [Online]. Available: https://www.usenix.org/conference/soups2017/workshop-program/way2017/redmiles

[16] K. Thomas, F. Li, A. Zand, J. Barrett, J. Ranieri, L. Invernizzi, Y. Markov, O. Comanescu, V. Eranti, A. Moscicki, D. Margolis, V. Paxson, and E. Bursztein, "Data Breaches, Phishing, or Malware? Understanding the Risks of Stolen Credentials," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '17. Dallas, Texas, USA: Association for Computing Machinery, Oct. 2017, pp. 1421–1434.

[17] W. Han, M. Ni, W. Xu, Z. Li, and G. Gu, "Shadow Attacks Based on Password Reuses: A Quantitative Empirical Analysis," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 2, pp. 309–320, 2016.

[18] D. Wang, Z. Zhang, P. Wang, J. Yan, and X. Huang, "Targeted online password guessing: An underestimated threat," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '16. New York, NY, USA: ACM, 2016, pp. 1242–1254. [Online]. Available: http://doi.acm.org/10.1145/2976749.2978339

[19] D. Jaeger, C. Pelchen, H. Graupner, F. Cheng, and C. Meinel, "Analysis of Publicly Leaked Credentials and the Long Story of Password (Re-)use," in *Proceedings of the 2016 Conference on Passwords*, ser. PasswordsCon '16, 2016, p. 19.

[20] A. Das, J. Bonneau, M. Caesar, N. Borisov, and X. Wang, "The Tangled Web of Password Reuse," in *Proceedings of the NDSS Symposium*, ser. NDSS '14, no. February. Internet Society, 2014, pp. 23–26.

[21] M. Golla, M. Wei, J. Hainline, L. Filipe, M. Dürmuth, E. Redmiles, and B. Ur, ""what was that site doing with my facebook password?": Designing password-reuse notifications," in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '18. New York, NY, USA: ACM, 2018, pp. 1549–1566. [Online]. Available: http://doi.acm.org/10.1145/3243734.3243767

[22] "Gates predicts death of the password." [Online]. Available: https://www.cnet.com/news/gates-predicts-death-of-the-password/

[23] J. Nielsen and R. L. Mack, *Usability Inspection Methods*. John Wiley & Sons, 1994.

[24] J. Bonneau, C. Herley, P. C. van Oorschot, and F. Stajano, "The quest to replace passwords: a framework for comparative evaluation of Web authentication schemes," University of Cambridge, Computer Laboratory, Tech. Rep. UCAM-CL-TR-817, Mar. 2012. [Online]. Available: https://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-817.pdf

[25] J. Lang, A. Czeskis, D. Balfanz, M. Schilder, and S. Srinivas, "Security Keys: Practical Cryptographic Second Factors for the Modern Web," in *Financial Cryptography and Data Security*. Springer, Berlin, Heidelberg, Feb. 2016, pp. 422–440.

[26] R. MacGregor, "Evaluating the Android Security Key Scheme: An Early Usability, Deployability, Security Evaluation with Comparative Analysis," in *Who Are You?! Adventures in Authentication Workshop*, ser. WAY '19, Santa Clara, California, USA, Aug. 2019, pp. 1–6.

[27] J. Colnago, S. Devlin, M. Oates, C. Swoopes, L. Bauer, L. Cranor, and N. Christin, ""It's not actually that horrible"," *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, pp. 1–11, 2018. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3173574.3174030

[28] M. Fagan, Y. Albayram, M. M. H. Khan, and R. Buck, "An investigation into users' considerations towards using password managers," *Human-centric Computing and Information Sciences*, vol. 7, no. 1, p. 12, Mar 2017. [Online]. Available: https://doi.org/10.1186/s13673-017-0093-6

[29] S. G. Lyastani, M. Schilling, S. Fahl, M. Backes, and S. Bugiel, "Better managed than memorized? studying the impact of managers on password strength and reuse," in *27th USENIX Security Symposium (USENIX Security 18)*. Baltimore, MD: USENIX Association, 2018, pp. 203–220. [Online]. Available: https://www.usenix.org/conference/usenixsecurity18/presentation/lyastani

[30] E. M. Redmiles, S. Kross, and M. L. Mazurek, "How i learned to be secure: A census-representative survey of security advice sources and behavior," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '16. New York, NY, USA: ACM, 2016, pp. 666–677. [Online]. Available: http://doi.acm.org/10.1145/2976749.2978307

[31] I. Ion, R. Reeder, and S. Consolvo, ""...no one can hack my mind": Comparing expert and non-expert security practices," in *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. Ottawa: USENIX Association, 2015, pp. 327–346. [Online]. Available: https://www.usenix.org/conference/soups2015/proceedings/presentation/ion

[32] R. Wash, E. Rader, R. Berman, and Z. Wellmer, "Understanding password choices: How frequently entered passwords are re-used across websites," in *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*. Denver, CO: USENIX Association, 2016, pp. 175–188. [Online]. Available: https://www.usenix.org/conference/soups2016/technical-sessions/presentation/wash

[33] A. Forget, S. Komanduri, A. Acquisti, N. Christin, L. F. Cranor, and R. Telang, "Security Behavior Observatory: Infrastructure for Long-Term Monitoring of Client Machines," Carnegie Mellon University CyLab, Tech. Rep. 14-009, 2014. [Online]. Available: https://www.cylab.cmu.edu/files/pdfs/techreports/CMUCyLab14009.pdf

[34] C. E. Shannon, "Prediction and entropy of printed English," *The Bell System Technical Journal*, vol. 30, no. 1, pp. 50–64, Jan. 1951.

[35] A. L. Durity, B. Ur, L. F. Cranor, N. Christin, M. L. Mazurek, R. Shay, S. M. Segreti, P. S. Huh, S. Komanduri, and L. Bauer, "Designing Password Policies for Strength and Usability," *ACM Transactions on Information and System Security*, vol. 18, no. 4, pp. 1–34, 2016, iSBN: 9781577357384.

[36] Y. Guo and Z. Zhang, "LPSE: Lightweight password-strength estimation for password meters," *Computers & Security*, vol. 73, pp. 507–518, Mar. 2018.

[37] W. Melicher, B. Ur, S. M. Segreti, S. Komanduri, L. Bauer, N. Christin, and L. F. Cranor, "Fast, lean, and accurate: Modeling password guessability using neural networks," in *25th USENIX Security Symposium (USENIX Security 16)*. Austin, TX: USENIX Association, Aug. 2016, pp. 175–191. [Online]. Available: https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/melicher

[38] D. L. Wheeler, "zxcvbn: Low-budget password strength estimation," in *25th USENIX Security Symposium (USENIX Security 16)*. Austin, TX: USENIX Association, Aug. 2016, pp. 157–173. [Online]. Available: https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/wheeler

[39] P. G. Kelley, S. Komanduri, M. L. Mazurek, R. Shay, T. Vidas, L. Bauer, N. Christin, L. F. Cranor, and J. López, "Guess again (and again and again): Measuring password strength by simulating password-cracking algorithms," *Proceedings - IEEE Symposium on Security and Privacy*, pp. 523–537, 2012.

[40] "John the ripper password cracker." [Online]. Available: https://www.openwall.com/john/

[41] "Password manager for families, businesses, teams - 1password." [Online]. Available: https://1password.com/

[42] "Open source password management solutions - bitwarden." [Online]. Available: https://bitwarden.com/

[43] "Password manager app for home, mobile, business - dashlane." [Online]. Available: https://www.dashlane.com/

[44] "Keeper security - best personal and business password manager." [Online]. Available: https://www.keepersecurity.com/

[45] "Share passwords with a password manager for business - lastpass." [Online]. Available: https://www.lastpass.com/

[46] "Google password manager." [Online]. Available: https://passwords.google.com/intro

[47] "Firefox lockwise - password manager - take your passwords everywhere." [Online]. Available: https://www.mozilla.org/en-US/firefox/lockwise/

[48] E. Rader, R. Wash, and B. Brooks, "Stories as informal lessons about security," in *Proceedings of the Eighth Symposium on Usable Privacy and Security*, ser. SOUPS '12. New York, NY, USA: ACM, 2012, pp. 6:1–6:17. [Online]. Available: http://doi.acm.org/10.1145/2335356.2335364

[49] E. M. Redmiles, S. Kross, and M. L. Mazurek, "Where is the digital divide?: A survey of security, privacy, and socioeconomics," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI '17. New York, NY, USA: ACM, 2017, pp. 931–936. [Online]. Available: http://doi.acm.org/10.1145/3025453.3025673

[50] A. P. Felt, A. Ainslie, R. W. Reeder, S. Consolvo, S. Thyagaraja, A. Bettes, H. Harris, and J. Grimes, "Improving ssl warnings: Comprehension and adherence," in *Proceedings of the 33rd Annual ACM Conference on Human*

Factors in Computing Systems, ser. CHI '15. New York, NY, USA: ACM, 2015, pp. 2893–2902. [Online]. Available: http://doi.acm.org/10.1145/2702123.2702442

[51] "Figma: the collaborative interface design tool." [Online]. Available: https://www.figma.com/

[52] "Limesurvey: the online survey tool - open source surveys." [Online]. Available: https://www.limesurvey.org/

[53] "Amazon mechanical turk." [Online]. Available: https://www.mturk.com/

[54] R. Wash, E. Rader, and C. Fennell, "Can People Self-Report Security Accurately?: Agreement Between Self-Report and Behavioral Measures," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. Denver, Colorado, USA: ACM Press, 2017, pp. 2228–2232. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3025453.3025911

[55] E. M. Redmiles, Z. Zhu, S. Kross, D. Kuchhal, T. Dumitras, and M. L. Mazurek, "Asking for a friend: Evaluating response biases in security user studies," in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '18. New York, NY, USA: ACM, 2018, pp. 1238–1255. [Online]. Available: http://doi.acm.org/10.1145/3243734.3243740

[56] E. M. Redmiles, Y. Acar, S. Fahl, and M. L. Mazurek, "A Summary of Survey Methodology Best Practices for Security and Privacy Researchers," Tech. Rep., 2017.

[57] D. A. Dillman, J. D. Smyth, and L. M. Christian, *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*, 4th ed. Hoboken, New Jersey: John Wiley & Sons, 2014.

[58] J. Krosnick, "Survey research," *Annual Review of Psychology*, vol. 50, pp. 537–567, 1999.

[59] A. Braunstein, L. Granka, and J. Staddon, "Indirect content privacy surveys: Measuring privacy without asking about it," in *Proceedings of the Seventh Symposium on Usable Privacy and Security*, ser. SOUPS '11. New York, NY, USA: ACM, 2011, pp. 15:1–15:14. [Online]. Available: http://doi.acm.org/10.1145/2078827.2078847

[60] P. Kumaraguru and L. F. Cranor, "Privacy indexes: A survey of westin's studies," Tech. Rep., 2005.

[61] S. Egelman and E. Peer, "Scaling the Security Wall: Developing a Security Behavior Intentions Scale (SeBIS)," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '15. ACM, 2015, pp. 2873–2882.

[62] S. Egelman, M. Harbach, and E. Peer, "Behavior Ever Follows Intention? A Validation of the Security Behavior Intentions Scale (SeBIS)," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16. San Jose, California, USA: Association for Computing Machinery, May 2016, pp. 5257–5261.

[63] E. Hargittai and Y. P. Hsieh, "Succinct survey measures of web-use skills," *Social Science Computer Review*, vol. 30, no. 1, pp. 95–107, 2012. [Online]. Available: https://doi.org/10.1177/0894439310397146

[64] "Conduct online surveys using opinio." [Online]. Available: http://www.objectplanet.com/opinio/

[65] "Pspp - gnu project - free software foundation." [Online]. Available: https://www.gnu.org/software/pspp/

[66] "Spss software - ibm." [Online]. Available: https://www.ibm.com/analytics/spss-statistics-software

[67] C. Bravo-Lillo, L. F. Cranor, J. Downs, and S. Komanduri, "Bridging the Gap in Computer Security Warnings: A Mental Model Approach," *IEEE Security Privacy*, vol. 9, no. 2, pp. 18–26, Mar. 2011.

[68] C. Bravo-Lillo, S. Komanduri, L. F. Cranor, R. W. Reeder, M. Sleeper, J. Downs, and S. Schechter, "Your Attention Please: Designing Security-decision UIs to Make Genuine Risks Harder to Ignore." pp. 6:1–6:12, 2013, iSBN: 978-1-4503-2319-2. [Online]. Available: http://doi.acm.org/10.1145/2501604.2501610

[69] D. Akhawe and A. P. Felt, "Alice in warningland: A large-scale field study of browser security warning effectiveness," in *Presented as part of the 22nd USENIX Security Symposium (USENIX Security 13)*. Washington, D.C.: USENIX, 2013, pp. 257–272. [Online]. Available: https://www.usenix.org/conference/usenixsecurity13/technical-sessions/presentation/akhawe

[70] S. Egelman and S. Schechter, "The importance of being earnest [in security warnings]," in *Financial Cryptography and Data Security*, A.-R. Sadeghi, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 52–59.

[71] "Browser market share worldwide - statcounter global stats." [Online]. Available: https://gs.statcounter.com/browser-market-share

[72] "Browser market share." [Online]. Available: https://netmarketshare.com/browser-market-share.aspx

[73] "W3counter: Web browser market share trends." [Online]. Available: https://www.w3counter.com/trends

[74] "Chrome devtools - tools for web developers - google developers." [Online]. Available: https://developers.google.com/web/tools/chrome-devtools/

[75] M. D. Buhrmester, S. Talaifar, and S. D. Gosling, "An evaluation of amazon's mechanical turk, its rapid rise, and its effective use," *Perspectives on Psychological Science*, vol. 13, no. 2, pp. 149–154, 2018. [Online]. Available: https://doi.org/10.1177/1745691617706516

[76] P. G. Kelley, "Conducting Usable Privacy & Security Studies with Amazon' s Mechanical Turk," in *Proceedings of the 6th Symposium On Usable Privacy and Security ({SOUPS} 2010)*, Redmond, WA, 2010, p. 3.

[77] W. Mason and S. Suri, "Conducting behavioral research on Amazon's Mechanical Turk," *Behavior Research Methods*, vol. 44, no. 1, pp. 1–23, Mar 2012.

[78] D. Navarro, "Some reflections on trying to be ethical on mechanical turk," in *Australasian Experimental Psychology Conference (EPC)*, University of New South Wales, Adelaide, 2015, p. 74.

[79] C. for Protection of Human Subjects, "Mechanical Turk (Mturk) for Online Research," University of California, Berkeley, Tech. Rep., Jan. 2018. [Online]. Available: https://cphs.berkeley.edu/mechanicalturk.pdf

[80] "45 CFR 46," Feb. 2016. [Online]. Available: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/index.html

[81] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian Journal of Statistics*, vol. 6, no. 2, pp. 65–70, 1979. [Online]. Available: http://www.jstor.org/stable/4615733

[82] "Have i been pwned: Check if your email has been compromised in a data breach." [Online]. Available: https://haveibeenpwned.com/

[83] D. Difallah, E. Filatova, and P. Ipeirotis, "Demographics and dynamics of mechanical turk workers," in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, ser. WSDM '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 135–143. [Online]. Available: https://doi.org/10.1145/3159652.3159661

[84] R. Flesch, "A new readability yardstick." *Journal of Applied Psychology*, vol. 32, no. 3, pp. 221 – 233, 1948.

[85] J. P. Kincaid, J. Fishburne, R. L. Rogers, and B. S. Chissom, "Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel," NAVAL TECHNICAL TRAINING COMMAND MILLINGTON TN RESEARCH BRANCH, Tech. Rep. RBR-8-75, Feb. 1975. [Online]. Available: https://apps.dtic.mil/docs/citations/ADA006655

[86] S. A. Zhang, S. Pearman, L. Bauer, and N. Christin, "Why people (don't) use password managers effectively," in *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. Santa Clara, CA: USENIX Association, Aug. 2019. [Online]. Available: https://www.usenix.org/conference/soups2019/presentation/pearman

[87] L. F. Cranor, "A Framework for Reasoning About the Human in the Loop," pp. 1:1–1:15, 2008, iSBN: 9781605580111. [Online]. Available: http://dl.acm.org/citation.cfm?id=1387649.1387650

Appendix A

Mechanical Turk Recruitment

A Survey About Online Account Notifications (Academic Research)

Please complete a short survey for academic research.
I am conducting academic research to measure the possible impacts interaction with password management applications can have on users' understanding of common online account notifications. As a participant in this research you will be asked read a short scenario and imagine yourself to be the user of a new password management application. You will be shown a notification and asked to respond to 35 short answer and multiple choice questions concerning your understanding of the notification seen.

This HIT has been allocated 60 minutes to complete, **but should take approximately 15 minutes total.** We do not want your HIT to expire, so have allocated significantly more time than you will require.

The survey will be conducted via another website. Here are the relevant instructions:

1. When you are ready to take the survey, please click this link to open a NEW browser tab or window: https://web.cs.dal.ca/ macgregor/survey/index.php/597983
2. Before taking the survey please take the time to read and consider the consent document provided. The survey will begin when you click through an onscreen prompt.
3. You will be given a unique completion code reflecting your participation in this survey. You must enter the code into the box below when you submit this HIT. Your HIT will then be complete, and your payment will be released within 48 hours.

4. You should **keep this Amazon Mechanical Turk HIT window open at all times**. Closing this window may make it impossible for you to collect your payment.

You will be paid $1.00 CDN for completing this HIT. In addition, 3 of the 300 workers who will eventually participate in this survey will be randomly assigned a bonus of $50.00 CDN.

Appendix B

Survey Instrument

Scenario

Introduction: In the following survey, you will be asked to imagine that you have recently started using a new password manager, *NewPass*. You have installed the application and added passwords for email, banking, social media, and other frequently used online accounts. Imagine that you have added your own passwords for accounts that are important to you.
This survey should take approximately 15 minutes to complete.

Prompt: After installing *NewPass* and adding passwords for email, banking, social media, and other frequently used online accounts:

1. You begin browsing the web. A *NewPass* notification appears in the browser window while logging into a frequently used email account.
   The notification appears below.

2. You begin using *NewPass* to check that the passwords you have added are secure. A notification appears related to a frequently used email account.
   The notification appears below.

Questions

1. In your own words, please describe what this notification is telling you.
   (short answer)

2. What may have caused you to see this notification? Please check all that apply.
   - Someone hacked or attempted to hack your email account.
   - You attempted to login to your email account from a new location or device, or accidentally entered the wrong password too many times.
   - You reused the same or similar passwords for multiple online accounts.

- Someone is trying to gain unauthorized access to your email account.
- *NewPass* regularly checks the security of passwords. This is just a normal security notification.
- You have a weak password for your email account.
- *NewPass* showed this notification by mistake.
- You went to a malicious website or downloaded malicious software.
- *NewPass* requires you to regularly change your email account password (e.g. every 90 days).
- Don't know.

3. If you saw this notification about a frequently used email account you had with a real company, which of the following best describes what you would do about your password for that account?
   - I would keep my password the same.
   - I would change my password.
   - Don't know

4. Why?
   (short answer)

5. If "I would change my password" is selected in Q3. What would you use for your new password?
   - Something related to the old password, but a few characters different.
   - Something completely unrelated to the old password.
   - A password that I already use for other accounts.
   - A password generated by a password manager or browser.
   - Other

6. If "I would change my password" is selected in Q3. How would you try to remember your new password? Select all that apply.
   - Write it down (e. g. in a diary, on a sticky note).
   - Use a password manager.
   - Just try to remember it.
   - Save it on my computer (e.g. in a document).

- Save it on my phone (e.g. in a note).
- Other (short answer)

7. If you saw this notification about a frequently used email account you had with a real company, which of the following best describes what you would do about passwords on other accounts? Please select all that apply.
   - I would change all of my passwords I have on other accounts.
   - I would change my passwords only for other accounts where I use the same password.
   - I would change my passwords only for other accounts where I use similar passwords.
   - I would change my passwords only for really important accounts (e.g. bank account).
   - I would keep my passwords the same.
   - Don't know.

8. Why?
   (short answer)

9. If any of the first four responses in Q7 is selected. What would you use for your new password(s) on those other accounts?
   - Something related to the old password, but a few characters different.
   - Something completely unrelated to the old password.
   - A password that I already use for other accounts.
   - A password generated by a password manager or browser.
   - Other (short answer)

10. If any of the first four responses in Q7 is selected. How would you try to remember your new password(s) for those other accounts? Select all that apply.
    - Write it down on paper (e. g., in a diary, on a sticky note).
    - Use a password manager.
    - Just try to remember it.
    - Save it on my computer (e. g., in a document).
    - Save it on my phone (e. g., in a note).
    - Other (short answer)

11. People have different reactions and responses to notifications about their online accounts. If you saw this notification about a frequently used email account you had with a real company, how likely would you be to take the following actions? ∘ Very Unlikely ∘ Unlikely ∘ Neither Likely nor Unlikely ∘ Likely ∘ Very Likely ∘ Don't Know
    - Enable Two-Factor Authentication.
    - Update my security questions.
    - Review my recent account activity.
    - Leave my password as-is.
    - Commit to change my password more frequently in the future.
    - Use a password manager or browser to generate a new password.
    - Sign up for an account with a company offering identity theft protection.
    - Sign up for an account with a company monitoring data breaches for passwords I use.
    - Update the software my devices more regularly.
    - Add a/Change my current password, PIN, pattern, fingerprint, etc. to lock my phone.
    - Add a/Change my current password to lock my computer.

12. There are many different actions that people could take in response to notifications about their online accounts. Please select the answer choice that most closely matches how you feel about the following statements:
    If I saw this notification about a frequently used email account I have with a real company, it would improve my account security if I...
    ∘ Strongly Disagree ∘ Disagree ∘ Neither Agree nor Disagree ∘ Agree ∘ Strongly Agree ∘ Don't Know
    - enabled Two-Factor Authentication.
    - changed my password for this account to a new password that is a modification (changing a few characters) of the old one.
    - changed my password for this account to a completely new password unrelated to the old one.
    - changed my password for this account to a password I use for another online account.

- used a password manager or browser to generate a new password for this account.
- used unique passwords for each of my online accounts.
- changed all of my similar passwords on other online accounts to one new password.
- updated my security questions.
- reviewed my recent activity.
- left my password as-is.
- committed to change my password more frequently in the future.
- signed up for an account with a company offering identity theft protection.
- updated the software on my devices more regularly.
- added a/changed my current password, PIN, pattern, fingerprint, etc. to lock my phone.
- added a/changed my current password to lock my computer.

13. I would expect a real password manager to display notifications like this one when necessary.
    ○ Strongly Disagree ○ Disagree ○ Neither Agree nor Disagree ○ Agree ○ Strongly Agree ○ Don't Know

14. If I saw this notification about a frequently used email account I have with a real company, I would believe that it was accurate or correct.
    ○ Strongly Disagree ○ Disagree ○ Neither Agree nor Disagree ○ Agree ○ Strongly Agree ○ Don't Know

15. I feel that ignoring this notification would not have any consequences.
    ○ Strongly Disagree ○ Disagree ○ Neither Agree nor Disagree ○ Agree ○ Strongly Agree ○ Don't Know

16. For me, taking action in response to this notification would be a:
    ○ Not a priority ○ Low priority ○ Medium priority ○ High priority ○ Very high priority ○ Don't Know

17. After seeing this notification, I would probably take action:
    ○ Immediately ○ Within 24 hours ○ Within 3 days ○ Within a week ○ After a

week ∘ Don't Know

18. This notification adequately explains what is going on with my email account.
    ∘ Strongly Disagree ∘ Disagree ∘ Neither Agree nor Disagree ∘ Agree ∘ Strongly Agree ∘ Don't Know

19. If I saw this notification about a frequently used email account I have with a real company, I wouldn't know why I saw this notification.
    ∘ Strongly Disagree ∘ Disagree ∘ Neither Agree nor Disagree ∘ Agree ∘ Strongly Agree ∘ Don't Know

20. I feel that this notification explained to me how to resolve the situation.
    ∘ Strongly Disagree ∘ Disagree ∘ Neither Agree nor Disagree ∘ Agree ∘ Strongly Agree ∘ Don't Know

21. To your knowledge, has anyone ever gained unauthorized access to one of your online accounts?
    - ∘ Yes
    - ∘ No
    - ∘ Don't know

22. If yes is selected in Q21. Who do you think accessed your online account? Please select all that apply.
    - ∘ Someone you know personally
    - ∘ Someone you don't know personally
    - ∘ Don't know

23. If yes is selected in Q21. Please describe what happened.
    (short answer)

24. Do any of your online accounts require you to change your password regularly (e.g. every 90 days)?
    - ∘ Yes
    - ∘ No
    - ∘ Don't know

25. If yes is selected in Q24. Please describe how you were informed of this regular password change policy.
    (short answer)

26. Have you ever been notified that your information was exposed in a data breach?
    - Yes
    - No
    - Don't know

27. If yes is selected in Q26. Please describe how you found out and what happened.
    (short answer)

28. There are many different approaches that people can take to manage passwords for their online accounts. Please select the answer choice that most closely matches how you feel about the following statements:
    ∘ Strongly Disagree ∘ Disagree ∘ Neither Agree nor Disagree ∘ Agree ∘ Strongly Agree ∘ Don't Know
    - I do not change my passwords, unless I have to.
    - I use different passwords for different accounts that I have.
    - When I create a new online account, I try to use a password that goes beyond the site's minimum requirements.
    - I do not include special characters in my password if it's not required.

29. Have you ever used a password manager?
    - Yes
    - No
    - Don't know

30. If yes is selected in Q29. Which password manager did/do you use?
    (short answer)

31. With what gender do you identify?
    - Female
    - Male
    - Non-binary

- Other
- Prefer not to say

32. What is your age?
    - 18-24
    - 25-34
    - 35-44
    - 45-54
    - 55-64
    - 65-74
    - 75 or older
    - Prefer not to say

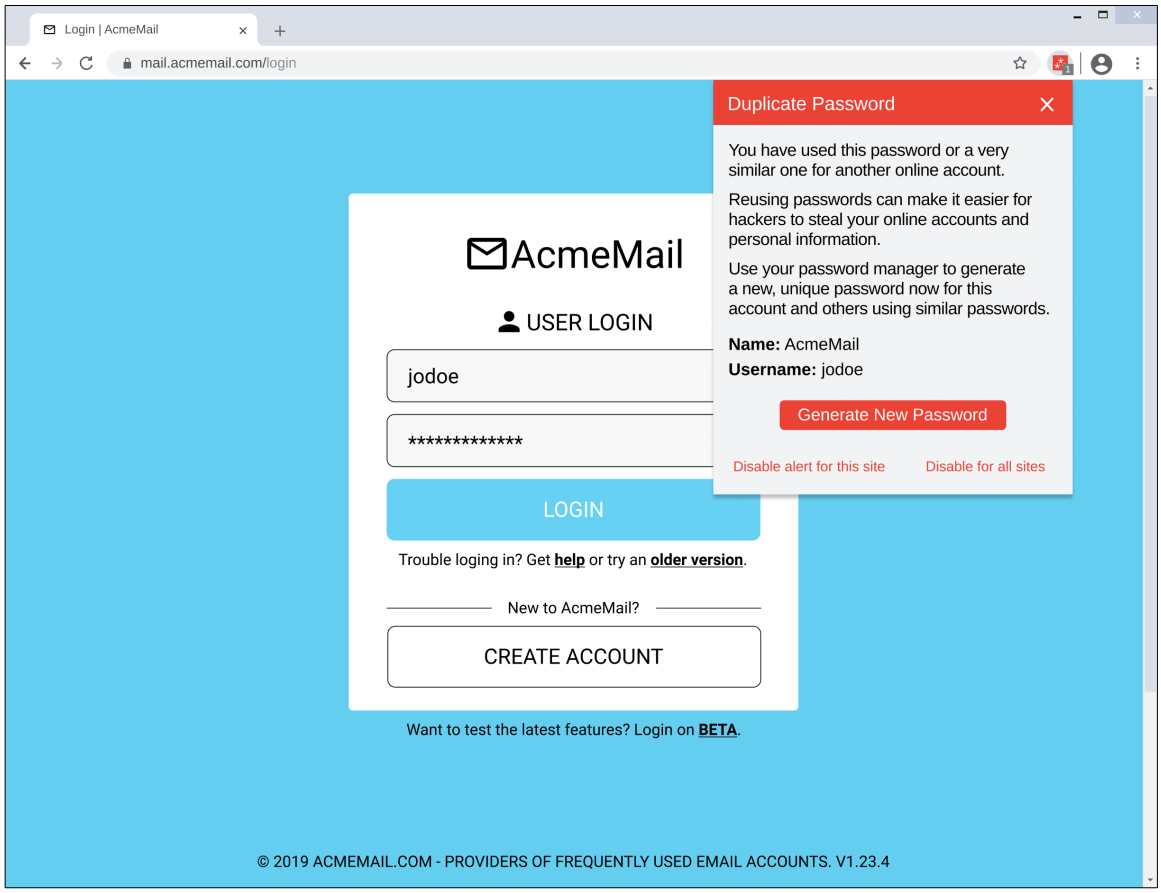
33. What is the highest degree or level of school you have completed?
    - Some high school
    - High school
    - Some college
    - Trade, technical, or vocational training
    - Associate's Degree
    - Bachelor's Degree
    - Master's Degree
    - Professional degree
    - Doctorate
    - Prefer not to say

34. Which of the following best describes your educational background or job field?
    - I have an education in, or work in, the field of computer science, computer engineering or IT.
    - I do not have an education in, nor do I work in, the field of computer science, computer engineering or IT.
    - Prefer not to say

35. (Optional) Do you have any final thoughts or questions about today's study? (short answer)

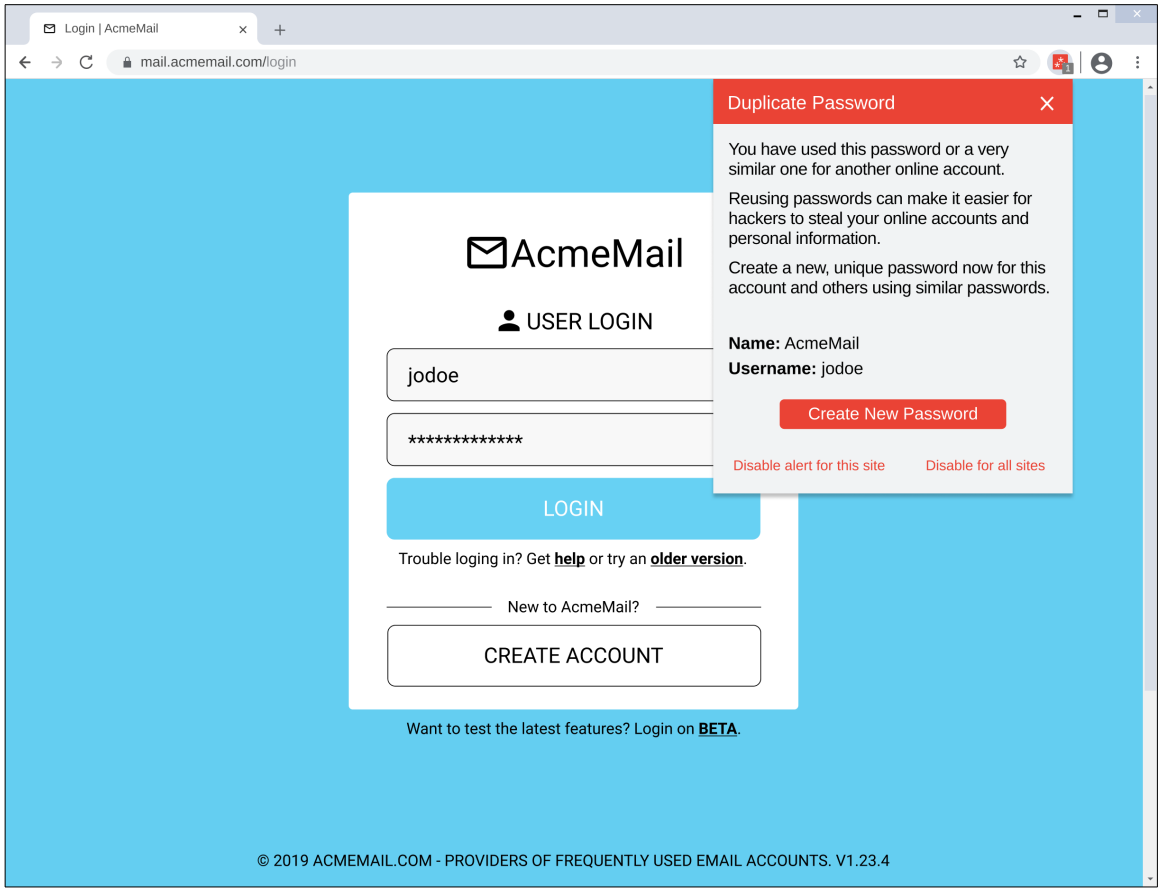
Appendix C

Prototypes



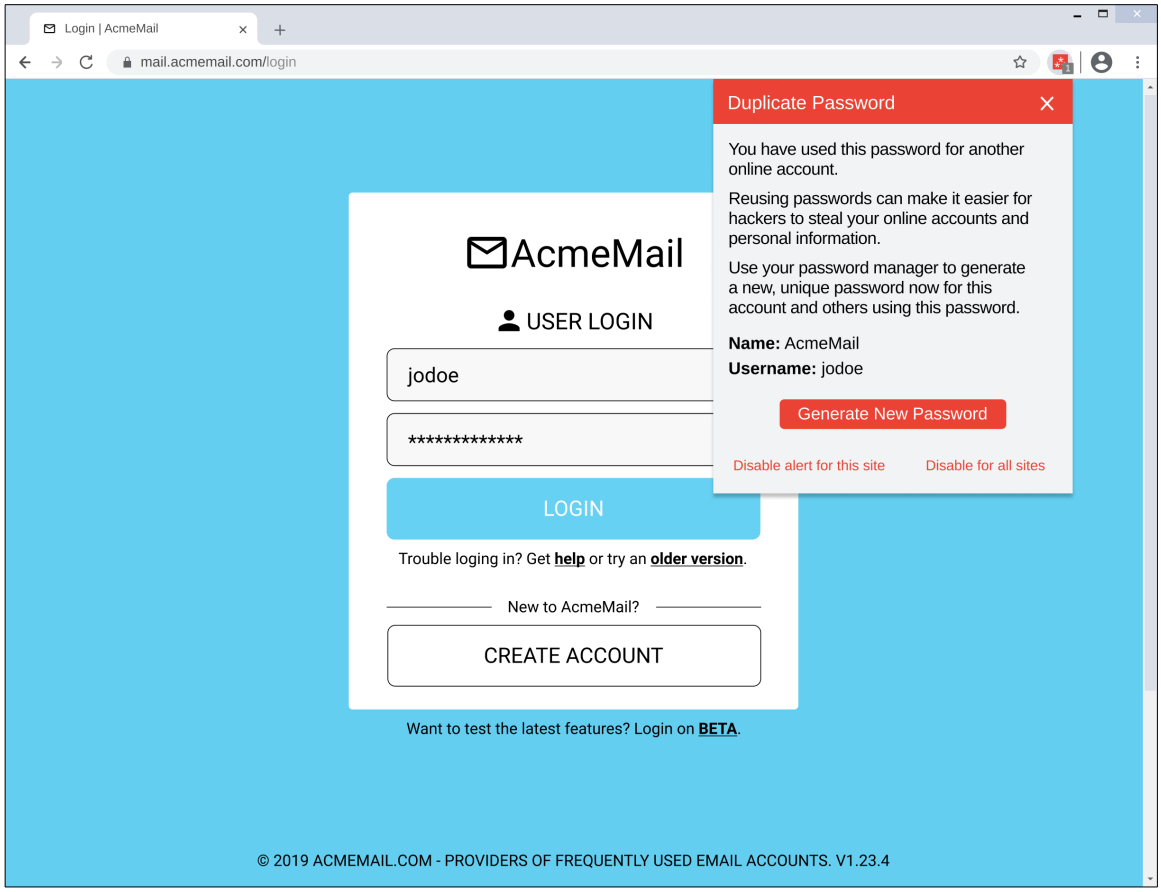
Condition 0

Login Alert - Partial - Generate



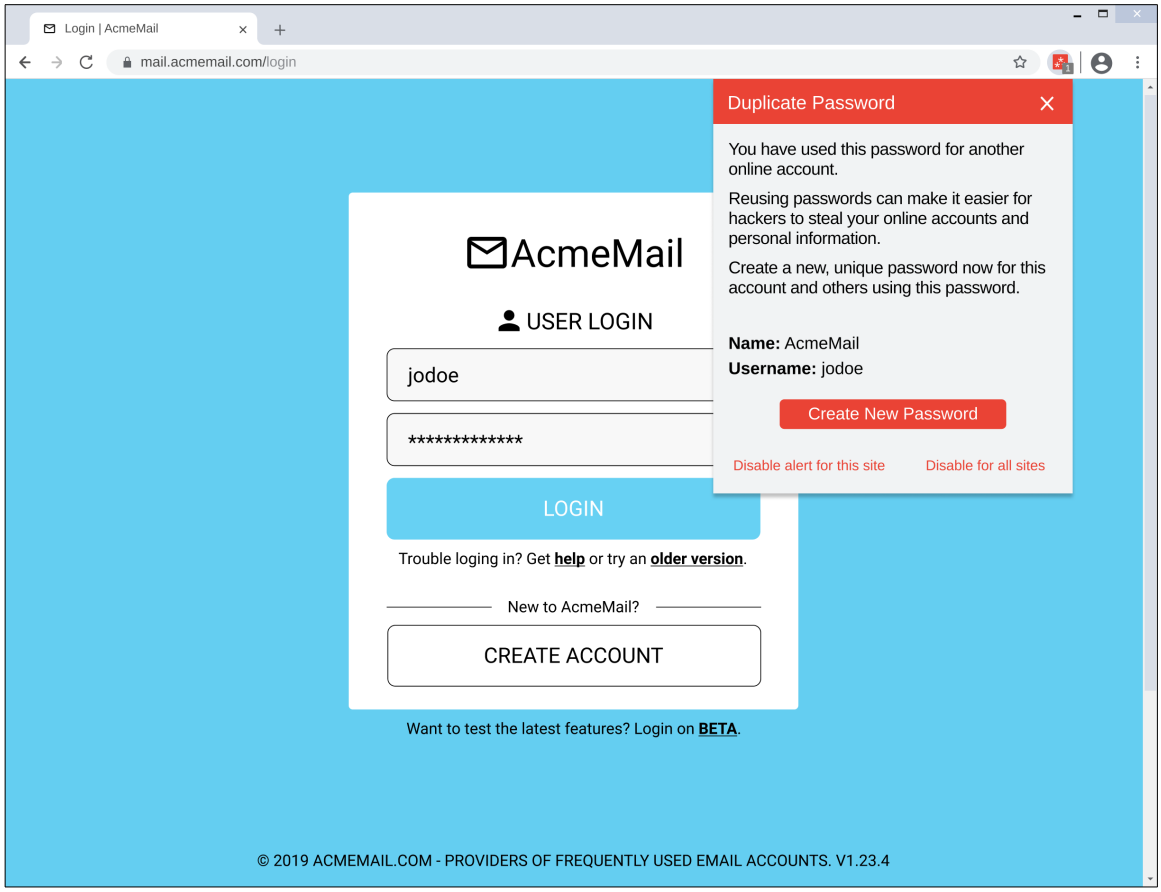
Condition 1

Login Alert - Partial - /



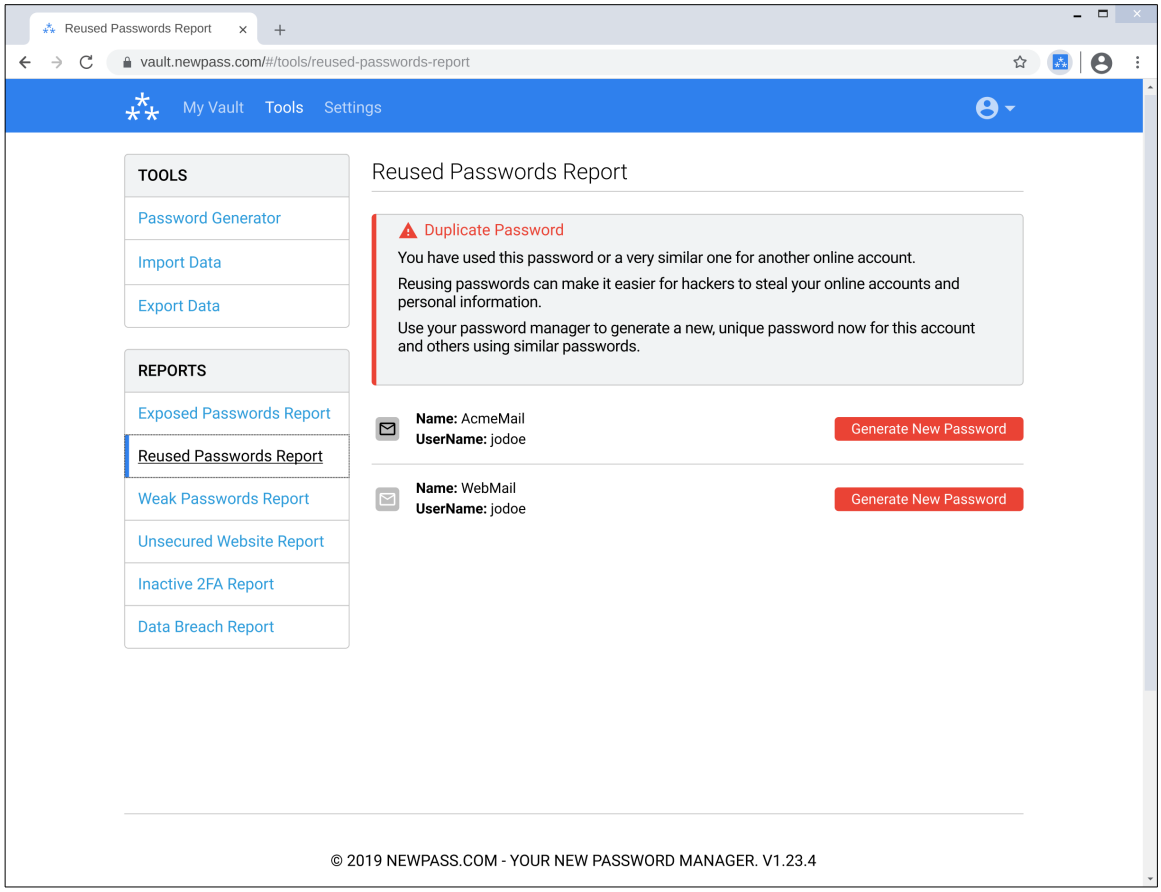
Condition 2

Login Alert - Exact - Generate



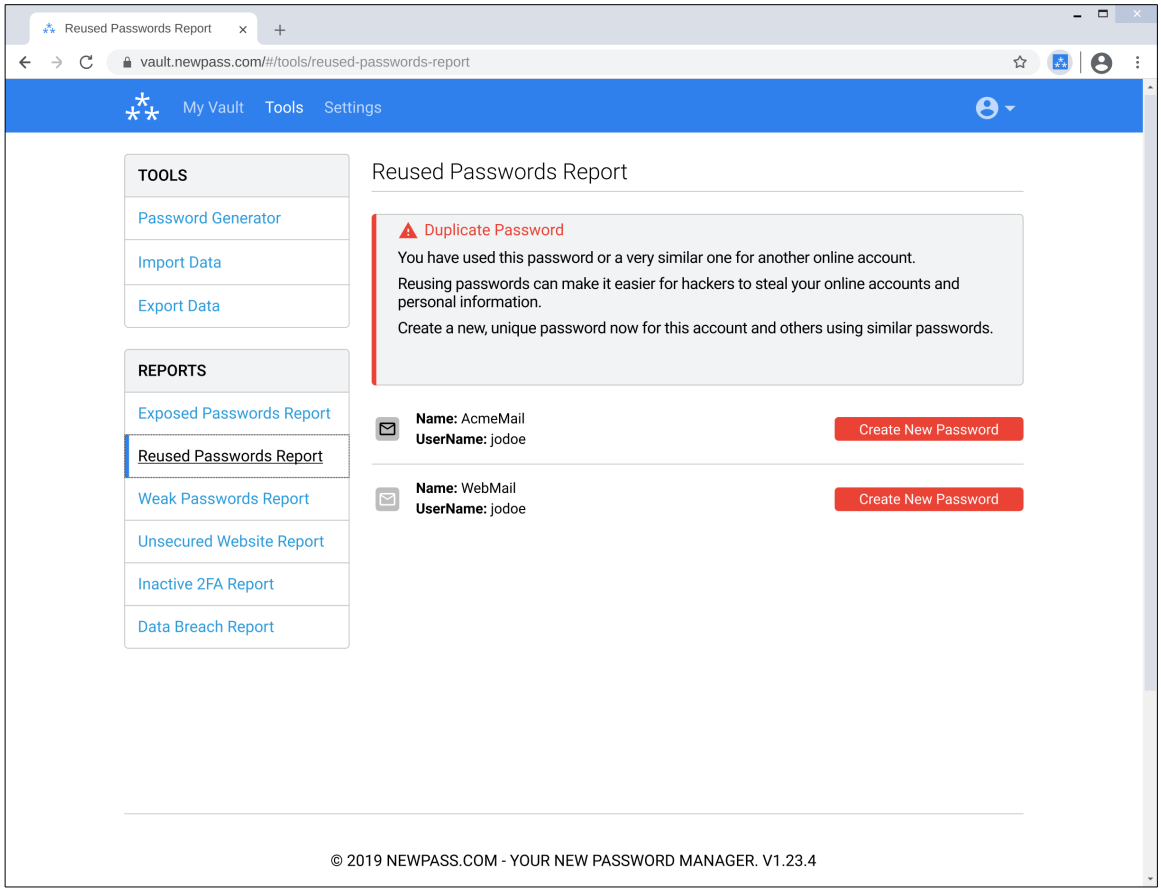
Condition 3

Login Alert - Exact - /



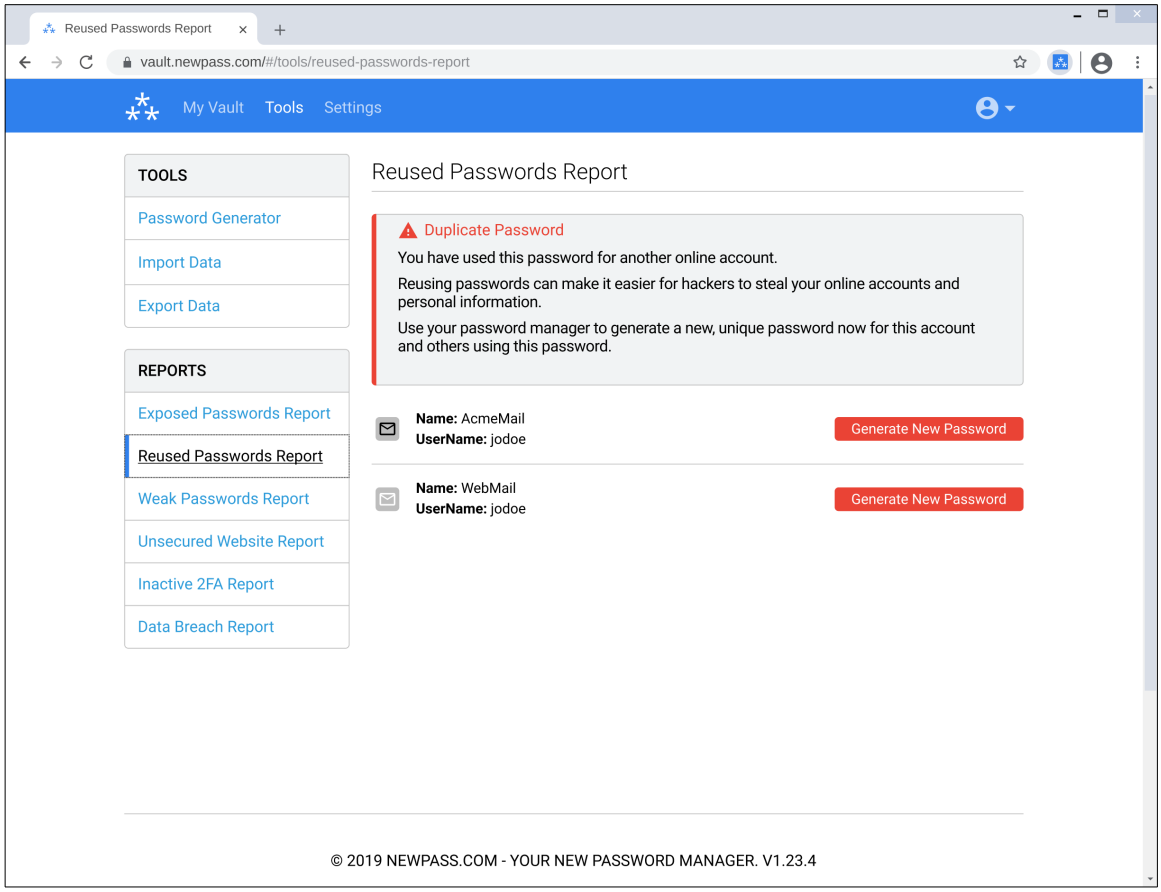
Condition 4

Security Audit - Partial - Generate



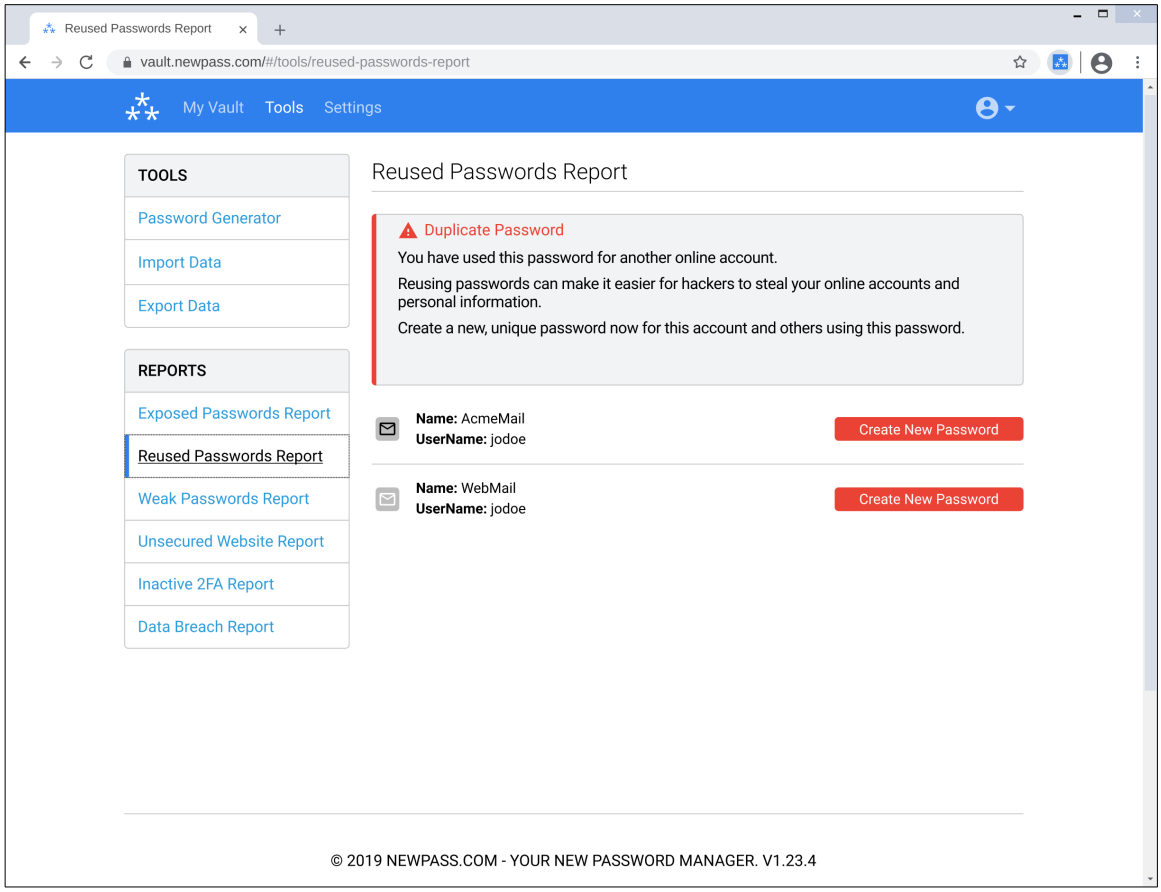
Condition 5

Security Audit - Partial - /



Condition 6

Security Audit - Exact - Generate



Condition 7

Security Audit - Exact - /

Appendix D

Detailed Statistics

Frequency Data

Condition

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	31	11.9	11.9	11.9
	1	33	12.7	12.7	24.6
	2	30	11.5	11.5	36.2
	3	33	12.7	12.7	48.8
	4	31	11.9	11.9	60.8
	5	33	12.7	12.7	73.5
	6	36	13.8	13.8	87.3
	7	33	12.7	12.7	100.0
	Total	260	100.0	100.0	

Active Alert

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	133	51.2	51.2	51.2
	1	127	48.8	48.8	100.0
	Total	260	100.0	100.0	

Partial Reuse

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	132	50.8	50.8	50.8
	1	128	49.2	49.2	100.0
	Total	260	100.0	100.0	

Generate

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	132	50.8	50.8	50.8
	1	128	49.2	49.2	100.0
	Total	260	100.0	100.0	

Tests of Correlation

Condition * Account Compromise

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	3.635[a]	7	0.821
Likelihood Ratio	3.686	7	0.815
N of Valid Cases	260		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 7.85.

Condition * Data Breach

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	8.231[a]	7	0.313
Likelihood Ratio	8.560	7	0.286
N of Valid Cases	260		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 10.62.

Condition * Password Manager Use

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	5.164[a]	7	0.640
Likelihood Ratio	5.214	7	0.634
N of Valid Cases	260		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 10.38.

Condition * Technical Expertise

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	5.261[a]	7	0.628
Likelihood Ratio	5.202	7	0.635
N of Valid Cases	260		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 9.35.

Alert * Account Compromise

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.824[a]	1	0.364		
Continuity Correction[b]	0.588	1	0.443		
Likelihood Ratio	0.826	1	0.364		
Fisher's Exact Test				0.399	0.222
N of Valid Cases	260				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 33.22.

b. Computed only for a 2x2 table

Alert * Data Breach

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1.642[a]	1	0.200		
Continuity Correction[b]	1.326	1	0.249		
Likelihood Ratio	1.646	1	0.200		
Fisher's Exact Test				0.243	0.125
N of Valid Cases	260				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 44.94.

b. Computed only for a 2x2 table

Alert * Password Manager Use

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.073[a]	1	0.787		
Continuity Correction[b]	0.020	1	0.888		
Likelihood Ratio	0.073	1	0.787		
Fisher's Exact Test				0.796	0.444
N of Valid Cases	260				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 43.96.

b. Computed only for a 2x2 table

Alert * Technical Expertise

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.425[a]	1	0.514		
Continuity Correction[b]	0.269	1	0.604		
Likelihood Ratio	0.425	1	0.514		
Fisher's Exact Test				0.592	0.302
N of Valid Cases	260				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 39.57.

b. Computed only for a 2x2 table

Part * Account Compromise

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.963[a]	1	0.326		
Continuity Correction[b]	0.706	1	0.401		
Likelihood Ratio	0.965	1	0.326		
Fisher's Exact Test				0.397	0.200
N of Valid Cases	260				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 33.48.

b. Computed only for a 2x2 table

Part * Data Breach

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	3.579[a]	1	0.059		
Continuity Correction[b]	3.105	1	0.078		
Likelihood Ratio	3.593	1	0.058		
Fisher's Exact Test				0.069	0.039
N of Valid Cases	260				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 45.29.

b. Computed only for a 2x2 table

Part * Password Manager Use

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.744[a]	1	0.388		
Continuity Correction[b]	0.536	1	0.464		
Likelihood Ratio	0.745	1	0.388		
Fisher's Exact Test				0.435	0.232
N of Valid Cases	260				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 44.31.

b. Computed only for a 2x2 table

Part * Technical Expertise

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	2.478[a]	1	0.115		
Continuity Correction[b]	2.074	1	0.150		
Likelihood Ratio	2.487	1	0.115		
Fisher's Exact Test				0.141	0.075
N of Valid Cases	260				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 39.88.

b. Computed only for a 2x2 table

Generate * Account Compromise

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.022[a]	1	0.883		
Continuity Correction[b]	0.000	1	0.995		
Likelihood Ratio	0.022	1	0.883		
Fisher's Exact Test				0.889	0.497
N of Valid Cases	260				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 33.48.

b. Computed only for a 2x2 table

Generate * Data Breach

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.006[a]	1	0.940		
Continuity Correction[b]	0.000	1	1.000		
Likelihood Ratio	0.006	1	0.940		
Fisher's Exact Test				1.000	0.522
N of Valid Cases	260				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 45.29.

b. Computed only for a 2x2 table

Generate * Password Manager Use

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.006[a]	1	0.936		
Continuity Correction[b]	0.000	1	1.000		
Likelihood Ratio	0.006	1	0.936		
Fisher's Exact Test				1.000	0.520
N of Valid Cases	260				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 44.31.

b. Computed only for a 2x2 table

Generate * Technical Expertise

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.323[a]	1	0.570		
Continuity Correction[b]	0.189	1	0.664		
Likelihood Ratio	0.323	1	0.570		
Fisher's Exact Test				0.594	0.332
N of Valid Cases	260				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 39.88.

b. Computed only for a 2x2 table

Multinomial Logistic Regressions of Significance

Password Manager Use by Q5 (Mitigation Strategy - Email)

Password Manager Use
Model Fitting Information

Model	Model Fitting Criteria: -2 Log Likelihood	Likelihood Ratio Tests: Chi-Square	df	Sig.
Intercept Only	56.560			
Final	35.916	20.644	5	0.001

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	0.000	0	
Deviance	0.000	0	

Likelihood Ratio Tests

Effect	Model Fitting Criteria: -2 Log Likelihood of Reduced Model	Likelihood Ratio Tests: Chi-Square	df	Sig.
Intercept	35.916[a]	0.000	0	
Password Manager Use	56.560	20.644	5	0.001

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

a. This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.

Parameter Estimates

What would you use for your new password?[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B): Lower Bound	Upper Bound
	Intercept	-0.477	0.300	2.526	1	0.112			
	[Password Manager Use=0]	-0.241	0.357	0.456	1	0.499	0.786	0.391	1.581
	[Password Manager Use=1]	0[b]			0				
-	Intercept	-3.367	1.017	10.961	1	0.001			
	[Password Manager Use=0]	-1.039	1.431	0.528	1	0.467	0.354	0.021	5.839
	[Password Manager Use=1]	0[b]			0				
A1 (Partial Reuse)	Intercept	-2.269	0.606	13.993	1	0.000			
	[Password Manager Use=0]	0.635	0.665	0.910	1	0.340	1.886	0.512	6.947
	[Password Manager Use=1]	0[b]			0				
A3 (Exact Reuse)	Intercept	-2.674	0.731	13.379	1	0.000			
	[Password Manager Use=0]	-1.039	1.023	1.032	1	0.310	0.354	0.048	2.627
	[Password Manager Use=1]	0[b]			0				
A4 (Generate New Unique Password)	Intercept	0.244	0.248	0.965	1	0.326			
	[Password Manager Use=0]	-1.283	0.329	15.216	1	0.000	0.277	0.145	0.528
	[Password Manager Use=1]	0[b]			0				

a. The reference category is: A2 (Create New Unique Password).

b. This parameter is set to zero because it is redundant.

Password Manager Use by Q9 (Mitigation Strategy - Other)

Password Manager Use
Model Fitting Information

Model	Model Fitting Criteria -2 Log Likelihood	Likelihood Ratio Tests Chi-Square	df	Sig.
Intercept Only	58.385			
Final	37.234	21.151	5	0.001

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	0.000	0	
Deviance	0.000	0	

Likelihood Ratio Tests

Effect	Model Fitting Criteria -2 Log Likelihood of Reduced Model	Likelihood Ratio Tests Chi-Square	df	Sig.
Intercept	37.234[a]	0.000	0	
Password Manager Use	58.385	21.151	5	0.001

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

a. This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.

Parameter Estimates

What would you use for your new password(s) on those other accounts?[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B) Lower Bound	Upper Bound
	Intercept	-0.534	0.305	3.057	1	0.080			
	[Password Manager Use=0]	-0.412	0.368	1.254	1	0.263	0.662	0.322	1.362
	[Password Manager Use=1]	0[b]			0				
-	Intercept	-2.674	0.731	13.379	1	0.000			
	[Password Manager Use=0]	-1.769	1.243	2.023	1	0.155	0.171	0.015	1.952
	[Password Manager Use=1]	0[b]			0				
A1 (Partial Reuse)	Intercept	-1.758	0.484	13.178	1	0.000			
	[Password Manager Use=0]	0.148	0.552	0.072	1	0.788	1.160	0.393	3.425
	[Password Manager Use=1]	0[b]			0				
A3 (Exact Reuse)	Intercept	-3.367	1.017	10.961	1	0.001			
	[Password Manager Use=0]	0.716	1.101	0.423	1	0.515	2.047	0.236	17.725
	[Password Manager Use=1]	0[b]			0				
A4 (Generate New Unique Password)	Intercept	0.216	0.250	0.751	1	0.386			
	[Password Manager Use=0]	-1.327	0.331	16.038	1	0.000	0.265	0.139	0.508
	[Password Manager Use=1]	0[b]			0				

a. The reference category is: A2 (Create New Unique Password).

b. This parameter is set to zero because it is redundant.

Ordinal Logistic Regressions of Significance

Experience With Data Breach by Q15 (Consiquences)

Experience With Data Breach
Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	58.044			
Final	45.464	12.579	1	0.000

Link function: Logit.

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	3.103	5	0.684
Deviance	3.497	5	0.624

Link function: Logit.

Parameter Estimates

								95% Confidence Interval	
		Estimate	Std. Error	Wald	df	Sig.	Exp(B)	Lower Bound	Upper Bound
Threshold	[Q15_SQ001 =]	-5.099	1.010	25.467	1	0.000	0.006	-7.079	-3.118
	[Q15_SQ001 = A1]	-1.315	0.222	35.151	1	0.000	0.268	-1.750	-0.880
	[Q15_SQ001 = A2]	0.601	0.202	8.807	1	0.003	1.823	0.204	0.997
	[Q15_SQ001 = A3]	1.547	0.221	49.224	1	0.000	4.699	1.115	1.980
	[Q15_SQ001 = A4]	3.041	0.291	109.571	1	0.000	20.931	2.472	3.611
	[Q15_SQ001 = A5]	4.777	0.536	79.322	1	0.000	118.689	3.725	5.828
Location	[BREACH=0]	0.844	0.240	12.357	1	0.000	2.326	0.373	1.315
	[BREACH=1]	0[a]			0				

Link function: Logit.
a. This parameter is set to zero because it is redundant.

Test of Parallel Lines[a]

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	45.464			
General	41.967	3.497	5	0.624

The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.

a. Link function: Logit.

Binomial Logistic Regressions Approaching Significance

Partial Reuse by Q2 (Cause of Notification)

Generate

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	5.079	1	0.024
	Block	5.079	1	0.024
	Model	5.079	1	0.024

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	Generate=0	1.138	0.537	4.492	1	0.034	3.119	1.089	8.931
	Constant	2.097	0.283	54.838	1	0.000	8.143		

Password Manager Use by Q2 (Cause of Notification)

Password manager Use

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	6.284	1	0.012
	Block	6.284	1	0.012
	Model	6.284	1	0.012

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	Password Manager Use=0	-1.587	0.759	4.367	1	0.037	0.205	0.046	0.906
	Constant	3.784	0.715	28.004	1	0.000	44.000		

a. Variable(s) entered on step 1: Password Manager Use.

Ordinal Logistic Regressions Approaching Significance

Partial Reuse by by Q15 (Consiquences)

Partial Reuse
Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	51.916			
Final	46.037	5.879	1	0.015

Link function: Logit.

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	2.683	5	0.749
Deviance	3.001	5	0.700

Link function: Logit.

Parameter Estimates

								95% Confidence Interval	
		Estimate	Std. Error	Wald	df	Sig.	Exp(B)	Lower Bound	Upper Bound
Threshold	[Q15_SQ001 =]	-5.869	1.011	33.715	1	0.000	0.003	-7.850	-3.888
	[Q15_SQ001 = A1]	-2.096	0.220	90.700	1	0.000	0.123	-2.527	-1.665
	[Q15_SQ001 = A2]	-0.219	0.169	1.695	1	0.193	0.803	-0.550	0.111
	[Q15_SQ001 = A3]	0.711	0.174	16.596	1	0.000	2.035	0.369	1.053
	[Q15_SQ001 = A4]	2.191	0.248	78.031	1	0.000	8.945	1.705	2.677
	[Q15_SQ001 = A5]	3.919	0.513	58.358	1	0.000	50.357	2.914	4.925
Location	[PART=0]	-0.544	0.225	5.839	1	0.016	0.580	-0.986	-0.103
	[PART=1]	0[a]			0				

Link function: Logit.
a. This parameter is set to zero because it is redundant.

Test of Parallel Lines[a]

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	46.037			
General	43.035	3.001	5	0.700

The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.

a. Link function: Logit.

Experience With Data Breach by Q19 (Wouldn't Know Why)

Experience With Data Breach
Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	47.322			
Final	41.689	5.633	1	0.018

Link function: Logit.

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	6.772	4	0.148
Deviance	6.667	4	0.155

Link function: Logit.

Parameter Estimates

								95% Confidence Interval	
		Estimate	Std. Error	Wald	df	Sig.	Exp(B)	Lower Bound	Upper Bound
Threshold	[Q19_SQ001 = A1]	0.050	0.201	0.062	1	0.803	1.052	-0.344	0.445
	[Q19_SQ001 = A2]	1.572	0.225	48.732	1	0.000	4.817	1.131	2.013
	[Q19_SQ001 = A3]	1.961	0.239	67.535	1	0.000	7.109	1.494	2.429
	[Q19_SQ001 = A4]	3.005	0.302	99.014	1	0.000	20.186	2.413	3.597
	[Q19_SQ001 = A5]	5.964	1.019	34.287	1	0.000	389.120	3.968	7.960
Location	[BREACH=0]	0.574	0.244	5.514	1	0.019	1.775	0.095	1.052
	[BREACH=1]	0[a]			0				

Link function: Logit.
a. This parameter is set to zero because it is redundant.

Test of Parallel Lines[a]

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	41.689			
General	35.022	6.667	4	0.155

The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.

a. Link function: Logit.