Do extroverts create stronger passwords?

by

Amit Maraj

A thesis submitted in partial fulfillment of
the requirements for the degree of

Master of Science

in

Computer Science

University of Ontario Institute of Technology

Supervisor: Dr. Miguel Vargas Martin

April 2018

Abstract

We investigate the relationship between personality types and the strength of created and selected passwords. For this purpose, we conducted an experiment on Amazon's Mechanical Turk, with 510 participants. Participants were given a pre-questionnaire that included, among others, three binary questions: "Password Awareness", "Security Training" and "Account Hijacking", which were used to predict participants' exposure to passwords in the past. Our results suggest that participants with higher levels of *Extroversion*, tend to create stronger passwords, if they were not required to change an online account password in the past (e.g., due to a security incident). In contrast, participants with lower levels of Extroversion tend to create stronger passwords (though not significantly), if they had been required to change an online account password in the past. These results indicate that there is a distinct relationship between the Extroversion personality dimension and the way we create passwords, whether it be in a familiar situation or not. Though password strength, as investigated, is the criterion of the aforementioned tests, it is worth mentioning that Extroversion cannot be deemed a predictor in this domain. We also investigated the relationship between personality and several password characteristics such as the total length, letters, digits, and symbols used within a password. To this end, we note that for participants who have had to change an online account password for the first time, *Extroversion* was directly correlated with creating and selecting shorter passwords, *Openness* was directly correlated with creating passwords containing fewer letters, but more numbers and symbols, and *Con-*

scientiousness was directly correlated with creating passwords containing fewer symbols. These results conclude that there is a distinct correlation between the construction of passwords and personality when participants are required to change an online account password for the first time. This thesis presents the detailed observations and findings from our experiment, discuss potential considerations for contradictions, and identify related future research.

Acknowledgements

I would like to thank my supervisor, Prof. Miguel Vargas Martin, for the patient guidance, encouragement and advice he has provided throughout my time as his student. I have been extremely lucky to have a supervisor who cared so much about my work, and who responded to my questions and queries so promptly.

I would also like to thank Prof. Matthew Shane of University of Ontario Institute of Technology and Prof. Mohammad Mannan of Concordia University for their guidance and expertise in unfamiliar research domains.

Additionally, I would like to acknowledge and thank the Natural Sciences and Engineering Research Council of Canada (NSERC) for their financial support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Personality, though it has been investigated to influence a plethora of psychological online behaviours, has had a relatively non-existent history in digital authentication research. The influence of personality traits has been investigated with regard to a plethora of online behaviours, but not much in authentication research. Anecdotally, there are individual differences in the way users create passwords for authentication. However, to what extent (if any), the strength of online passwords is influenced by one's personality remains unexplored, despite password research still being very active (see e.g., password strength [41, 76], password habits [25, 27], and effects of strength meters on passwords [71]).

There have been many studies in the psychometric field of personality types, providing ample evidence that our personality traits significantly influence our interaction with inanimate objects; see e.g., [16, 39, 44, 57, 75]. Interesting findings in this domain include correlations between personality and internet usage [44], personality and smartphone usage [16], and even personality and personal motivation [3, 39].

Traditionally, passwords have been researched in the strength and guessability domain [27,41,71,76,77] with an upward trend toward memorability [5,20,66,73,80], which con-

cludes that users tend to choose easier-to-remember passwords that include names, short words, dates, and patterns resulting in easier to guess passwords. The general consensus of most password research is an equilibrium between security and memorability, with a tradeoff for either. Based on previous experiences, users tend to gravitate toward making passwords similar to what they are familiar with such as including a number or symbol in their passwords if told that was the right thing to do in the past. More generally, studies on password selection, memorability and usability conclude that people choose poor passwords [1, 9, 58]. Users tend to choose short passwords and derive them from personal information that is easily guessable, although no solid research attempts were put forth to investigating personality-based influences on chosen and selected passwords.

We have conducted an online experiment, where 510 participants from across the world supplied various levels of data including situational-based created passwords and personality profiles. This data collected is discussed in Section 3.1.3.

One particular personality trait - Extroversion - has been linked to patterns of online behaviour, paving the grounds for the hypotheses in this study. For instance, Chittaranjan et al. [16] showed that Extroversion is negatively correlated with the amount of internet usage, which was further reinforced by Landers et al. [46] yielding the same result. Krämer et al. [44] found that extroverts were more willing to be experimental and try something completely different in a new situation.

Hypotheses. Based on the above studies, we put forward the following hypotheses:

1. When answering "No" to "Account Hijacking"[1], those who demonstrate a higher level of Extroversion are more likely to create a stronger password.

2. When answering "Yes" to "Account Hijacking", those who demonstrate a lower level of Extroversion are more likely to create a stronger password.

[1] "Account Hijacking" is a binary question asked to participants to determine whether they have ever had to change a password for an online account in the past as a result of a security breach.

The primary contribution of this research is the exploration of relationship between user's psychological traits and their password choices. We have conducted an online experiment, where 510 participants from across the world provided (through Amazon Mechanical Turk) situational-based created/selected passwords, as well as responses related to their personality profiles. Our major findings include: extroverts tended to create stronger passwords if they were not required to change their passwords due to a previous security incident (r = 0.184, p = 0.023, N = 154); while introverts tended to create stronger passwords (though not significantly) if they were required to change their passwords due to a previous security incident (r = -0.099, p = 0.062, N = 356).

Furthermore, we conducted exploratory analyses on the influence of personality traits on certain password characteristics (e.g., length and character distribution), which yielded several findings suggesting *Extroversion*, *Openness* and *Conscientiousness* have direct correlations with how participants constructed their passwords. Overall, our work opens the possibility of improving password security though personalization, by taking certain personality traits into consideration.

1.1 Motivation

Passwords are the leading form of digital authentication and repeatedly has always been preferred by users for authentication methods. Although much is known about passwords and what constitutes to a memorable and strong one, very little is known about its relationship with our psychological profile. This is of huge interest as discoveries in this area could open various doors of understanding such as how psychological state affects password strength, how a certain personality trait may respond to security incidents, whether specific personnel may be more or less likely to be trusted with sensitive data and so on.

Perhaps one of the more notable possibilities is the integration with trust. If a company

is able to identify a personality type of an employee with confidential clearance, they may be more likely to invest in extended security training for said individual should they notice their personality is more likely to create a weaker password. This could help mitigate risk for the company and save huge sums of money.

Furthermore, companies and websites may have the opportunity to implement a more tailored password creation form if they have a password meter specific to the user's personality. Of course, this is stipulated on the website knowing the personality type of the user, but an implication such as this could mean more accurate password creation. Noted at the weakest point of entry, user password-centric breaches can be reduced. Potential implications of this research can include helping make systems more secure by addressing psychological-based security loopholes when considering security designs. For example, if a company is thinking about implementing certain security standards into their infrastructure, conducting an experiment such as a fake security breach can help strengthen the way certain personality types create future passwords for their company account. Whether this is an ethical solution is up for discussion, the concept remains a thorough example.

1.2 Background

In this section, we provide further background information regarding several items related to our experiment including: the password meter we used for ranking passwords, a brief introduction to the Big-five personality traits, and the specific test we used for personality type determination.

1.2.1 The Zxcvbn Password Strength Measure

Several tools exist for evaluating and ranking the strength of specific passwords. We chose to use Zxcvbn [77] due to its simplicity and reliability (any decent strength meter can be

used in its place). Zxcvbn rates a password's strength on a scale from 0-4:

- **0:** *too guessable* – risky password;
- **1:** *very guessable* – protection from throttled online attacks;
- **2:** *somewhat guessable* – protection from unthrottled online attacks.
- **3:** *safely unguessable* – moderate protection from offline slow-hash scenario.
- **4:** *very unguessable* – strong protection from offline slow-hash scenario.

All the passwords created and obtained within this experiment were run through Zxcvbn for a standardized strength score, which would remain consistent within the life cycle of the experiment and analysis.

Along with the score, another field of particular interest generated by Zxcvbn is the "Offline Slow Hashing" which is Zxcvbn's simulated offline attack with multiple attackers using a slow hash function. This provided insight as to how long it took to crack each password in more granular detail so we could further normalize the selected passwords, which is explained further in Section 3.1.3 below.

1.2.2 Big 5 Personality Traits

Modern day personality research is commonly grounded in the work of McCrae [49] and Costa [19], and widely used in psychological research today. Developed in part as an revision to the seminal work by Cattell [12–15], who developed a relatively complex taxonomy of individual differences that consisted of 16 primary factors and 8 second-order factors, the strength of Costa and McCrae's *Big-5 Personality Characteristics* is its stability, consistency over the lifespan, and ability to predict important psychological and behavioral variables across a wide range of contexts and situations. Moreover, contemporary research suggests that each of the Big-5 characteristics appear to be universal (across countries and

cultures), and may have unique grounding in distinct biological underpinnings, further suggesting that they measure real, important, predictive personality features [48]. However, repeated attempts by researchers to replicate his work were unsuccessful [24, 68, 69] and, in each case, researchers found that a 5-factor model accounted for data quite well.

This model was investigated further in four studies [7, 28, 33, 52]. Borgatta's findings are noteworthy because he obtained five stable factors across five methods of data gathering. Norman's work is especially significant because his labels (Extroversion, Neuroticism, Agreeableness, Conscientiousness, and Intellect) are used commonly in the literature and have been referred to, subsequently, as "Norman's Big Five" or simply as the "Big Five" [2]. Borgatta deemed the fifth dimension as Intellect or Intellectence, which further changed to Openness as the fifth dimension instead, explained in more detail below. Openness is the chosen adoption for this study.

In the past 2 decades, the views of many personality psychologists have converged regarding the structure and concepts of personality. Although there are many personality tests that gauge several aspects of one's personality, they all root back to observing the Big-five personality traits. The 5-factor model was thus selected for this study as it helps discretize personalities on a quantifiable spectrum. Another common scale, the Myers-Briggs [59], observed 16 different indicators, which were too many discrete options for the purposes of this research. The goal was to find similarities between core, overarching dimensions of personalities.

The five personality indicators in Big-5 are as follows:

1. **Extroversion** relates to one's degree of outgoingness, particularly in social situations. High levels of Extroversion are characterized by excitability, sociability, talkativeness, assertiveness and high amounts of emotional expressiveness. People who are high in Extroversion tend to be outgoing and gain energy from social interactions. People who are low in Extroversion (i.e., more introverted) tend to be more

reserved and insulated.

2. **Agreeableness** relates to one's level of cooperativeness and concern for others. It relates to attributes such as trust, altruism (selflessness), kindness, affection as well as other pro-social characteristics (e.g., helping others). People who are high in Agreeableness tend to be more cooperative, while those with lower scores tend to be more competitive, more manipulative and more aggressive.
3. **Conscientiousness** relates to one's goal-directedness, thoughtfulness, and impulse control. Individuals high in Conscientiousness tend to be more organized, more responsible, and more mindful of details. Individuals low in Conscientiousness tend to be less rule-oriented and more irresponsible.
4. **Neuroticism** relates to one's level of emotional stability. High levels of Neuroticism are characterized by sadness, moodiness, and emotional instability. Those with low Neuroticism tend to be more emotionally stable and emotionally resilient.
5. **Openness** relates to one's ability to see connections between divergent concepts, and willingness to consider other perspectives. Individuals high in Openness tend to be imaginative and insightful, with a broad range of interests. Individuals low in Openness tend to be more traditional and may struggle with abstract thinking.

These traits appear to be universal as evident from studies across countries and cultures. Based on these studies, psychologists now believe that the five personality dimensions are not only universal, they may also have biological origins [48].

1.2.3 Mini-IPIP Personality Test

Investigators often want to measure a wide range of constructs in research; however, completing a large packet of questionnaires can be a boring or irritating task for participants. This might end up producing transient measurement errors (e.g., [61]) because participants

are in a negative mood, or because they respond carelessly due to frustration with the length of the assessment. Moreover, to the extent that it is even mildly unpleasant to participate in research, long questionnaires may increase the likelihood that participants will decide not to complete the study, will drop out of subsequent data collections in longitudinal studies, or will refuse to partake in future studies entirely [22]. As a web-based application, the experiment we conducted was intended to be as short and concise as possible. The motivation to investigate another test yielded the Mini-IPIP (International Personality Item Pool).

To evaluate the Big 5 personality indicators, we asked participants to complete the Mini-IPIP, which is a well-validated, 20-item short form of the 50-item IPIP [30]. The Mini-IPIP has been shown to be psychometrically stable, practically useful, and with stable test-retest reliability across studies that spanned from days to several months. The Mini-IPIP scales show consistent and acceptable internal consistencies [22].

Chapter 2

Related Work

Personality is an important determinator of one's general behaviours and habits, and may provide insight into one's trustworthiness, as well as one's impulses and goals [7,24,33,52, 69]. Personality also influences how we interact with others, both inside and outside our immediate social circle. Discussed below, personality also has a major influence on one's interaction with digital security.

2.1 Personality

Common convergence in research has brought current personality traits to the Big 5 personality indicators - Openness, Conscientiousness, Extroversion, Agreeableness and Neuroticism. However, this does not explain the intricate behaviours associated with each trait.

Openness, also commonly referred to as "Openness to Experience" has become attributed to the likelihood of obtaining a leadership position, likely due to the ability to entertain new ideas and think outside the box [47]. Douglas et al. [23] also noted that "Openness is also connected to universalism values, which include promoting peace and tolerance and seeing all people as equally deserving of justice and equality." This individ-

ual traits are especially valued in positions of higher status. Schretlen et al. [62] notes that Openness is also linked to knowledge and skills and due to exposure to new experiences, is one of the only traits that is less likely to change over time - rather would lead to gains in knowledge and skills. Some smaller traits encompassed by Openness include creativity, originality and a negative correlation to conservative political attitudes [64]. Openness is weakly related to Neuroticism and Extroversion, and is mostly unrelated to Agreeableness and Conscientiousness [53]. Openness is often deemed as the trait least likely to change over time, and perhaps most likely to help an individual grow. Openness is a trait that was highly considered in our research for cases where participants have been exposed to password creation in the past, due to the collection of experiences, knowledge and skills open people demonstrate.

Considered a stricter trait, Conscientiousness is highly attributed valuing order, duty, achievement, self-discipline and consciously practicing deliberation [56]. Though Openness is thought to help one achieve growth, Conscientiousness helps that person view their achievement and set their limits. Judge et al. [38] also showed Conscientiousness having a positive correlation with intrinsic and extrinsic career success, further augmented by Soldz and Vaillant [64] who found that Conscientiousness was also an indicator of one's adjustment to life's challenges along with the maturity of one's defensive responses. This indicates that the more Conscientiousness an individual is, the more prepared they are to tackle novel obstacles that come their way. Conscientiousness was found to correlate somewhat negatively with Neuroticism and somewhat positively with Agreeableness, but had no discernible relation to the other factors [53]. Conscientious individuals tend to demonstrate goal-based traits, sparking the potential for investigation within this trait to some degree in our research. In a situation where a conscientious person is faced with a new scenario such as creating a password for the first time, they may excel.

Extroversion shares very similar traits to Openness such as being a strong predictor

of leadership. Roccas et al. [56] finds that those high in Extroversion are likely to value achievement and stimulation, while often demonstrating traits of being assertive, active, sociable, and shunning self-denial in favour of excitement and pleasure. Extroversion is also synonymous with the unlikeliness to value tradition or conformity. This makes those who demonstrate a higher level of Extroversion more likely to try something new and exit their comfort zone. Soldz and Vaillant's [64] study also reports that high Extroversion positively correlates with high income, conservative political attitudes, early life adjustment to challenges and social relationships over a lifetime. The same long-term study also found that Extroversion generally remained the same in people over the course of a lifetime - Extroverts tend to stay Extroverts and visa versa with Introverts. Extroversion was also shown to be an excellent predictor of effective functioning and general well-being [54], positive emotions [74], and overconfidence in task performance [60]. Extroversion is weakly unrelated with Neuroticism, and weakly related to Openness [53]. Due to the high correlation with achievement and experiences, Extroversion was considered a viable trait for investigation in our research.

Roccas et al. [56] reported that overarching values for individuals who demonstrate a high level of Agreeableness are benevolence, tradition and conformity, while placing less importance on power, achievement or the pursuit of selfish pleasures. Agreeable individuals are more concerned with the well-being of others and less about themselves. Those high in Agreeableness are also more likely to have positive peer and family relationships, model gratitude and forgiveness, attain desired jobs, live long lives, experience relationship satisfaction, and volunteer in their community [54]. Agreeableness is somewhat related with Extroversion and Conscientiousness, while being somewhat unrelated with Neuroticism. Although interesting to look into how an agreeable individual would perform in a situation where a security incident would affect those around them more than themselves, we did not foresee this trait having much of a correlation with our research outcomes.

The last trait within the 5 dimensions is Neuroticism, which indicates mostly negative traits when observed at a higher level. Judge et al. [37] found that Neuroticism relates negatively to self-esteem, general self-efficacy and individual locus of control. People who demonstrate a higher level of Neuroticism tend to be close minded and do not get along well with others for the most part. Neuroticism is also highly correlated with emotional instability and vulnerability to stress and anxiety. A long-term study done by Soldz and Vaillant [64] concluded that Neuroticism was also negatively correlated with smoking cessation and healthy adjustment to life. Neuroticism correlates somewhat negatively with Agreeableness and Conscientiousness, in addition to a weak, negative relationship with Extroversion and Openness [53]. Neuroticism was not considered for our research as it requires a social and temporal aspect of research to accurately measure. It would be interesting to look at how Neuroticism would affect password suggestions as a facet of this trait is close-mindedness.

Two of the more popular measures for assessing the Big-five come from the Big-five Inventory (BFI) and the Revised NEO Personality Inventory (NEO PI-R). These two assessments gained popularity as they are the most reliable and valid measurements. Originally proposed in 1993 by Lewis R. Goldberg [29], the BFI was created to measure not only the five dimensions, but the 40 facets around them as well. The 44 question test has been used extensively in psychology research and is still quite popular. The original NEO-PI was introduced in 1985 by Paul Costa, Jr. and Robert McCrae [17], which has been revised 3 times [18,48,49] over in recent years to keep up with changing times. Originally developed to assess three main dimensions: Neuroticism, Extroversion and Openness, the NEO-PI has since been expanded to include a NEO Five-Factor Inventory (NEO FFI), which contains 60 items and measures just the overall domains instead of all facets. The latter of the tests is of particular interest to this our research as we hypothesize influences from the overarching dimensions rather than the granular facets.

2.2 Passwords

To help develop secure systems, competition historically has been devising new ways to attack the security of the system. At the same time, new techniques to resist the attacks are also developed. This competition has been in integral in the development of authentication over the years.

An underlying goal has been to provide password security at minimal inconvenience to the users of the system. For example, those who want to run a completely open system without passwords, or to have passwords only at the option of the individual users, are able to do so, while those who require all of their users to have passwords gain a high degree of security against penetration of the system by unauthorized users.

"A password system must be able not only to prevent any access to the system by unauthorized users (i.e., prevent them from logging in at all), but it must also prevent users who are already logged in from doing things that they are not authorized to do. The so-called "super-user" password on the UNIX system, for example, is especially critical because the super-user has all sorts of permissions and has essentially unlimited access to all system resources." [50]

Although implemented for remote-access systems, the UNIX system was the first consumer/commercial system to implement a password file containing the actual passwords of all the users. Originally, passwords stored within the UNIX system were in plaintext. They have since changed this to implement hashing for improved security of these passwords.

Text passwords have dominated human-computer authentication since the 1960s [79]. Although many password cracking studies have been done to support the claim that passwords are the sole weak point of security systems [4, 21, 50], there is still no consensus on the actual level of security provided by passwords or even on the appropriate metric for measuring security. Password authentication has existed for several decades and it is likely

to remain one of the top authentication mechanisms also in the future [6, 45].

2.3 Password Security

Along with passwords come security considerations. Due to decades of competition between password attacking and prevention, resulting research has proved that with modern technology, the difference between a weak and a strong password can be the difference between an inherit security incident and a successful mitigation strategy.

So far, large-scale password data has arisen only from security breaches such as the leak of 32 M passwords from gaming website RockYou in 2009 [21, 76]. Password corporations have typically been analyzed by simulating adversarial password cracking, leading to sophisticated cracking libraries but limited understanding of the underlying distribution of passwords.

Claude Shannon [63] defined the term entropy in information theory as "a statistical parameter which measures in a certain sense, how much information is produced on the average for each letter of a text in the language. If the language is translated into binary digits (0 or 1) in the most efficient way, the entropy H is the average number of binary digits required per letter of the original language." While Shannon was alluding to English text strings, the term entropy became widely used in cryptography as a measure of the difficulty in guessing or determining a password or a key [11]. Although common security practice is not aligned with using Shannon's entropy to benchmark passwords, it is worth noting as a precursor to newer methods of password strength estimations.

Estimating the entropy of a password depends on the number of options available for each character. If a password was binary and composed only of zeros and ones chosen randomly, there will be 2^n possible values of that password, where n is the number of bits in the password. The password, in this case, has n bits of entropy. As a general rule,

the entropy of a randomly chosen password is calculated as n^l where n is the number of available characters, and l is the length of the password. For example, if a password was based on the standard English keyboard, where there are 95 printable characters, the space available for each password character is 95. The entropy of an 8 characters long password that is randomly chosen based on a standard keyboard will be $95^8 \approx 6.6 \times 10^{15}$, which is almost equivalent to 2^{52}. In this case, the password is said to have 52 bits of entropy.

Considering the aforementioned, it is a well known fact that user-chosen passwords are somewhat predictable. There is very little literature providing a solid answer to the following question: given a number of guesses, what is the probability that a state-of-the-art attacker will be able to break a password. Passwords have an inherint trade-off between usability and security: while strong passwords are hard for attackers to guess, they are on the other hand also difficult for the user to remember, which drives a lot of the motivation behind creating a more memorable, but weaker password. Dell'Amico et al. [21] compared and evaluated the effectiveness of currently known password attacks using various datasets of known passwords, including over 50,000 real passwords. It was found that with the absence of a password policy (a set of rules designed to enhance computer security by encouraging users to employ strong passwords and use them properly see e.g., at least 1 digit, at least 1 uppercase letter), users tend to create weaker passwords. This was observed using a variety of password guessing techniques including dictionary attacks, mangling using dictionaries and context free grammars, and Markov chain-based strategies. It is also noteworthy that the password guessing techniques used in this study decreased roughly in performance as the size of the explored password grew in length. This shows that people tend to respond in a more security-centric fashion when provided with a sense of guidance, whether it is linked to psychological traits remains undetermined within this study, but would make for interesting future investigation.

Passwords are used all around the web, and with the emergence of more online services,

the need for multiple passwords becomes more apparent. To observe the effectiveness of current password cracking techniques, Weir et al. [76] collected passwords from several different websites, the largest one containing over 32 million passwords, to perform one of the largest studies of its kind to data to analyze strength in passwords. Taking a deeper dive into password policies, Weir et al. mention that password policies can become a double edged sword as although they seem to increase password robustness, appending "123" at the end of insecure passwords can easily be circumvented and taken into consideration by sophisticated password crackers. A password created as such may satisfy the requirements of a password policy, but may also contain a similar amount of insecurity as the same password without the digits at the end. Results concluded various findings including:

1. As password length grew, passwords became harder to crack.
2. Passwords which included an uppercase letter became significantly harder to crack.
3. Passwords which included a symbol became significantly harder to crack.

Based on the above findings, Weir et al. suggest various password policies, and suggested passwords when users create new passwords (e.g., if a user types "password123", the system may suggest "!!pasSword123"). Although it could be seen as a security vulnerability to suggest passwords to users when using a website, this could be grounds for future work to see how certain personalities respond to such suggestions. Also considering a level of awareness around the web nowadays with authentication, seeing suggested passwords on a website could be perceived as insecure by some users. It would be of interest to see how something like this would play out with individual personality types; whether they respond positively or negatively to a password suggestion.

Kelley et al. [41] took a deeper look into password policies and the influence they may have on user created passwords. Analyzing 12,000 passwords collected under seven different password policies via an online study, Kelley et al. investigated how resistant pass-

words created under different password policies were. For each password created under the 7 policies, participants were given the following scenario:

> *"Imagine that your main email service provider has been attacked, and your account became compromised. You need to create a new password for your email account, since your old password may be known by the attackers. Because of the attack, your email service provider is also changing its password rules. Please follow the instructions below to create a new password for your email account. We will ask you to use this password in a few days to log in again, so it is important that you remember your new password. Please take the steps you would normally take to remember your email password and protect this password as you normally would protect the password for your email account. Please behave as you would if this were your real password!"*

A notable observation in the above scenario is the last sentence - *"Please behave as you would if this were your real password!"*, could skew some of the results. Although the intent to replicate a real life scenario is non-trivial for accurate results, participants may be used to creating their real passwords under different circumstances. Enforcing a new scenario may cause distortion.

Participants were also given the respective password policy that outlines the password they created for the activity. Findings here shed light on passwords created and their strength when users are guided by a password policy. Although the email scenario shown above was common to every password policy, the policy itself was unique. Results indicate that as the policy became more strict and demanding, passwords created tended to be more resistant against password cracking attempts. The password policy $basic16$, which required participants to have at least 16 characters in their password provided the greatest security against a powerful attacker, outperforming the more complicated $comprehensive8$, which

required participants to have at least 8 characters including an uppercase and lowercase letter, a symbol and a digit. To improve password security on the whole, Weir et al.'s [76] research noted above might be a potential beneficiary of this research in that instead of suggesting a specific password, the authentication system could provide a suggested password policy on the fly instead. It would be of interest to observe how a participant's reaction varies between a suggested password and suggested policy.

2.4 Password Composition

A password's strength comes from its constitution. Research has looked into how passwords throughout the web have been and which ones work best. Password composition include anything from password length, to the number of uppercase letters used.

In addition to password strength testing, Weir et al. [76] also found some noteworthy observations regarding common password compositions in all the leaked passwords used in their study. Perhaps most noteworthy, the length of the password correlated positively with the number of digits, symbols, and uppercase letters used within those passwords (i.e., a higher percentage of passwords were determined to have digits, symbols and uppercase letters the higher they grew in length). 60%, 8% and 7.1% of passwords with 10+ characters contained digits, symbols and uppercase letters respectively. Whereas 57.5%, 4.4% and 6.5% of passwords with 7+ characters contained digits, symbols and uppercase letters respectively. The length of the password was also shown to be negatively correlated with containing only lowercase letters and digits. In general, this could imply that most people who created longer passwords were generally more security aware when it comes to secure passwords. Of the passwords that were more than 7+ characters in length, the majority of the ones that contained digits (64.28%) tended to have the digits *after* the password (e.g., password123). 20.51% were only number based (e.g., 1234567), 5.95% included the digits

before the password (e.g., 123password) and only 9.24% had digits sprinkled throughout the password (e.g., passw0rd, pass123word, p1a2ssword). Because no results were shown to illustrate the difference between digit composition comparison between passwords of length 7+ and 10+, no solid reporting can be done on whether users who create password that are 10+ characters in length are more likely to sprinkle their digits throughout the password. It would be assumed that as the length of the password grows, users tend to be more security aware, so they would be more inclined to sprinkling their digits within their password. It is also worth mentioning that of the passwords that contained digits, the most frequent string of digits were "1", "2" and "123" with 10.98%, 2.79% and 2.29% of passwords containing them respectively.

Further validating the findings above in Weir et al.'s [76] research, Komanduri et al. [42] conducted a similar study on 5,000 participants. It was also found that the $basic16$ password policy yielded the greatest security out of all given password policies. Komanduri et al. also tested for participant sentiment to the provided password policies. More participants agreed or strongly agreed to the $comprehensive8$ policy being more beneficial to a strong password in contrast to the $basic16$ policy (67% vs. 57% respectively), when in reality, the $basic16$ policy provided more security. What's of perhaps the most interest in this study is participants overarching sentiment toward the two policies. More participants agreed or strongly agreed to making a password for the $comprehensive8$ policy more annoying and difficult than the $basic16$. This implies that although uppercase letters, symbols and digits may be beneficial to the strength of a password, creating a longer password may not only be stronger, but easier for users. A policy which combines both could be warranted, however the case for annoyance and difficulty could outweigh password strength. It would be of interest to investigate the personality dimensions that reacted more positively to the individual password policies. For example, more creative personality type could enjoy the $comprehensive8$ policy more than $basic16$.

2.5 Personality and Passwords

Passwords are at the root of online security and to investigate whether or not there were psychological influences behind risky password practices, Lauren VanDam from LastPass [72] partnered with Lab42 to interview 2,000 adults around the world about their password habits, their beliefs and their understanding of what secure online behaviour looks like. They found that 91% of participants know there is a risk when reusing passwords, but 61% continue to do so. They also concluded that people tend to prioritize their financial online accounts over retail, social media, and entertainment. The data collected through the LastPass survey suggest that the theory of cognitive dissonance also applies to a user's digital behaviour: you know it is bad for you, but you continue to do it anyway for example, you pick up your phone to answer an important call while driving even though it is dangerous. These findings are of particular interest as it begins introducing social aspects of our online interactions.

It was shown that 82% of participants knew that a combination of letters, numbers and symbols create a stronger password, but while users understood what a secure password looks like, they still fell short when it came to password creation. 47% of all participants used initials, fiends or family names in their passwords, and 42% of all participants used significant dates and numbers. 26%, 21%, 14% and 13% of participants used pet names, birthdays, their hometown and their school name or mascot in their passwords. Although the risks of insecure passwords are aware to the majority of users, it may not be at the forefront of their concerns during password creation. Our research aimed to look more into how password strength and composition may be influenced by personality dimensions. Interested more in the lower level of password composition (i.e., password length, number of digits, number of letters and number of symbols), the observation of higher level password composition demonstrated in LastPass's research could extend to personality types. For

example, if a specific personality trait had a higher likelihood to include dates as opposed to family names in their password.

On top of this, LastPass also found that only 29% of consumers change their passwords for security reasons – the number 1 reason people change their passwords is because they forgot it. Taking into consideration forgetfulness, this sheds a specific light on the experiment activities conducted within this research that depends on a participant's answer to "Account Hijacking", which is solely dependent on requiring a change in password in the past. Although the question was framed to revolve around requiring the need to change an online account password due to a security breach, this could be interpreted ambiguously (i.e., the participant considering a forgotten password a type of security breach). They were also unable to find any correlation between two very different personality types and password behaviours. These personality types did not seem to impact online behaviour, but to drive rationalizations of poor password habits.

Although not immediately related to personalities, 39% of respondents within the LastPass survey mentioned that they create more secure passwords for personal accounts over work accounts. Though most businesses make it very clear that the first line of defense for businesses in protecting themselves from attacks is informed users, the two password creation activities within this research are strictly declared to be for personal accounts. The LastPass survey makes no reference to how the other 61% of respondents answered this question, so no solid conclusions can be derived from this. Although similar, our research is focused on identifying personality characteristics specifically the Big-five personality characteristics (Section 1.2.2) that predict how strong of a password they will choose. The ability to do this may help identify individuals at greatest risk for creating weak passwords, and help identify methods for encouraging these individuals towards stronger passwords.

2.6 Personality and Security

Halevi et al. [34] found that certain personality traits may influence security and privacy related user-behaviours online. Participants with higher levels of Neuroticism responded to phishing emails that touted prizes of some sort. Also, participants who scored high on the Openness factor tended to both post more information on Facebook and have less strict privacy settings, making them susceptible to privacy attacks. Whitty et al. [78] found that younger people and individuals who scored high on self-monitoring (participants who were more likely to observe and regulate their expressive behaviours) were more likely to share their passwords.

The security of computer systems often relies upon decisions and actions of end users. Ultimately, the final state of a computer system is up to the end user. Conducting a novel neuroscience-based study, Neupane et al. [51] reported on measuring users' security performance and underlying neural activity with respect online security incidents. It was discovered that a high degree of correlation ($p = 0.0002$) in brain activity within the decision-making regions were activated when attempting to detect phishing attacks and when presented with malware warning. It was also found that a high functional connectivity among the core regions of the brain was established while participants performed the phishing detection task. The regions of the brain that were highly engaged when trying to identify a fake website implied that participants had a more difficult time dealing with fake websites as opposed to the real ones. The fake websites may have posed more of a challenge to participants as they may have had to spend more time thinking about different attributes, sometimes recalling from memory. When posed with inherit security risks, this shows that people respond in a similar manner on a neural level. However, it does not suggest an appropriate evaluation of participant performance on the experiment activities. Where neural activity is the low level of human activity, personality can be considered higher level. Our

research investigated this higher level.

A study done by Thorpe et al. [67] introduces a difference in graphical password creation (passwords created with a series of provided images rather than characters) when participants are provided with the usable images in different manners. Participants were provided with 72 passwords in a grid fashion, but with a curtain in front of them before allowing them to select. In the case where the curtain drew from left the right (i.e., showing images on the left first), participants tended to select images on the left of the grid in more situations for a part of their passwords rather than images on the left. In the case where the curtain drew right to left (i.e., showing images on the right first), the opposite was true. Participants were also restricted from selecting images for their passwords until the curtain was fully drawn, thus alleviating the sense of urgency or time in participants selecting their password. This presents the idea of using visual cues to help encourage strong password creation, rather than just a policy of some sort. It seems as though Openness and Conscientiousness, being meticulous, calculative, collected and organized may correlate with the outliers in this study suggesting that they may have taken more time to evaluate all the options after the drawn curtain, rather than the familiar, first shown images.

Exposure to various online accounts with ranging levels of required security and password requirements create biases within data, which is of particular interest to this research. A study done by Landers et al. [46] investigated 117 undergraduate students and their internet usage in relation to their personality. It was found that total internet usage amongst the participants negatively correlated with Agreeableness and Conscientiousness, along with having a strong negative correlation with Extroversion. Being research that is primarily rooted within the same personality scale, Agreeableness, Conscientiousness, and Extroversion could become specific points of interest within an online behavioural domain. This study by Landers et al. makes no mention to Intellect and Neuroticism having any influence on internet usage amongst participants, but as any personality trait, these two affect

the way we behave on any level. The lack of findings for Intellect and Neuroticism in this study could be because they play little to no role on technology-related usage as observed in Table 4.11. Of course, this is not a solid finding, but has been a repeated trend in several findings in our research.

Usage of the web goes hand in hand with internet usage. Halevi et al. [35] investigated the approach to cyber-security within different cultures and personalities, which points to the suggestion that certain personality traits affect a user's cyber-security related behaviour when dealing with web-based forms and personal information. This falls in line with research done by Gratian et al. [32], who found that characteristics such as financial risk-taking, rational decision-making, Extroversion, and gender were found to be significant unique predictors of good security behaviours.

2.7 Personality and Behaviour

Personality indicators such as Extroversion and Introversion can be correlated with more abstract personal behaviours, which is especially useful for this study. A study done by Vernon et al. [75] found that different types of humours are correlated with the big 5 personality indicators such as affiliative (humour used to bring people together) and self-enhancing (humour directed toward yourself, even in bad situations) humour being correlated positively with Extroversion and Openness, with a specific Openness to new experiences. Barrick [2] determined that Conscientiousness showed consistent positive relations with job proficiency, training proficiency, and personnel data all within the job performance spectrum. Their study also revealed that Extroversion was a valid positive predictor for training proficiency and social interactions with managers and sales.

2.8 Personality and Technology

With personality types playing a large factor with general and targeted internet usage, investigation between socially-directed platforms within the web and personality types have also resulted in interesting conclusions. One study in particular done by Krämer et al. [44] found that there was a strong correlation between levels of Extroversion and participants' likeliness to be more experimental with their profile pictures such as choosing a photo with a "different style" (e.g., black and white or altered colours) rather than a realistic colour picture. This is of interest as being more experimental can have a similar effect on password creation. Extroverts being more creative may find it easier to come up with and memorize a more complicated password.

Besides internet usage within introverts, another study conducted by Chittaranjan et al. [16] on participants with self-reported personality traits yielded interesting results in relation to smartphone usage. Beside further reinforcement of a strong correlation between internet usage and Introversion, this study found that application usage, number of calls, and number of SMS logs had a positive correlation with Introversion – further proving that Introversion as a personality trait and its correlation to internet usage also translates to mobile usage.

Our research aims to investigate the relationship between the Big-five dimensions of personality and user behavior regarding digital authentication. To the best of our knowledge, the Big-five personality traits ("Openness", "Neuroticism", "Conscientiousness", "Agreeableness" and "Extroversion") have never been investigated for their influences in password choice.

Chapter 3

Experiment

3.1 Experiment Design

The tool used to collect data from a wide diversity of locations was Amazon's Mechanical Turk (MTurk), which is an online labour market created by Amazon to assist "requesters" in hiring and paying "workers" for the completion of computerized tasks.

3.1.1 Infrastructure

The experiment was developed as a web application to take advantage of MTurk's requirements. The web application warranted data collection in a seamless manner by sending the required data to a database, which was stored on UOIT's premises to abide by UOIT's Ethics Research Board guidelines under our approved protocol for this research. The web server did not require a large amount of computation, so a minimal cloud-based server was provisioned to host the website.

The web application was developed using HTML5, CSS3, JavaScript, and jQuery, which utilized AJAX requests to send and retrieve data to and from the on-site database. The experiment ran for roughly two hours, at which point the 500 participant set limit was

hit. Upon completion, all data was backed up locally and to an external database. The discrepancy between the 500 participant limit and the 517 total participants is due to the extra experiments conducted internally as a pilot test before publishing to MTurk.

Design

We used Amazon's Mechanical Turk (MTurk) to recruit our participants. Upon clicking the experiment URL hosted on MTurk, participants were forwarded to a web application we designed, which consisted of several different parts: a questionnaire, password ranking test, password scenarios, and a personality test; see Fig. 3.1. We elaborate these parts below.

Preliminary Questionnaire.

Every participant was greeted with an initial questionnaire, which consisted of several general demographic questions. Three questions worth noting include:

- *"Security Training"*: Have you ever had any security training in the past? (this includes any type of security for example, law enforcement, computer, etc.)
- *"Password Awareness"*: Have you ever had any password security awareness training? (e.g., learning the differences between a weaker and stronger password)
- *"Account Hijacking Involvement"*: Have you ever been required to change your password as a result of an account compromise in the past?

These questions were used to segment the data in Section 4 where the results are discussed in further detail.

Password Ranking Test.

The initial test given to the participants was a password ranking test where five passwords were given to each participants. Every password was randomly selected from the pool of passwords prepared for this experiment, one from each rank given by Zxcvbn; see Section 3.1.3. Participants were asked to place each of these passwords within the following buckets: Very Weak, Weak, Normal, Strong, and Very Strong. The goal of this activity

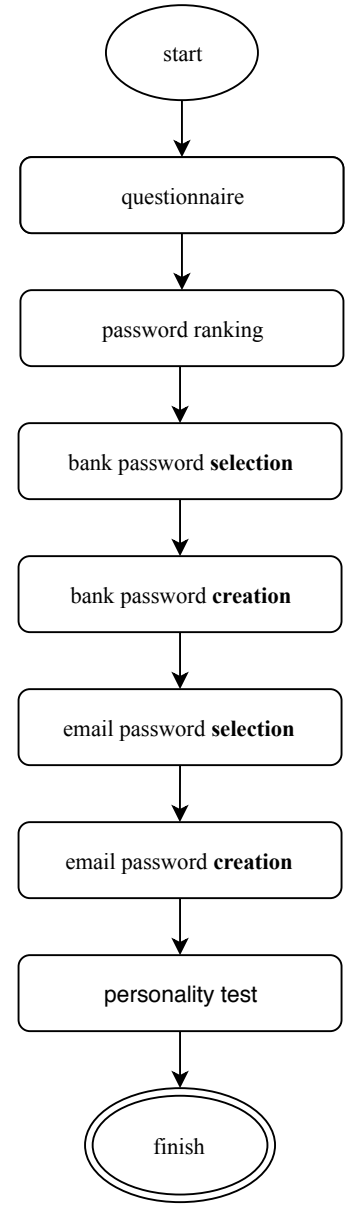

Figure 3.1: Experiment Design

was to measure the participant's ability to identify stronger passwords from weaker ones. Research done by Ur et al. [70] shows that participants' perceptions of a password's security level usually fell right in line with current password-cracking tools. To observe a participants' perception of given a given password's security level, this password ranking activity was given. No instructions about determining the differences between strong and weak passwords were given.

Bank Scenario.

There are two scenario-based tests given to each participant – the first of which is a bank scenario where participants were given the following dialogs. The instructions are different only in the last sentence (password creation vs. password selection).

- "Imagine there was a breach within your **main banking provider's** online banking platform and because of this, your bank has released a notice that says all accounts may have been ***compromised***. Your bank strongly recommends a password change for all accounts. Please **create** a new password below."
- "Imagine there was a breach within your **main banking provider's** online banking platform and because of this, your bank has released a notice that says all accounts may have been ***compromised***. Your bank strongly recommends a password change for all accounts. Please **select** a new password below."

Email Scenario.

The second scenario was email-based where participants were asked the following questions (again, the difference is between password creation vs. selection):

- Imagine your main **email service provider has been attacked** and that because of the attack, your email service provider is requesting all users change their password. ***This is your main email account and contains very sensitive information.*** Please **create** a password below.

  Note 1: Create the most secure password you feel **comfortable** using and you'll

be able to remember.

Note 2: This password should be different from the one you created in the previous step.

- Imagine your main**email service provider has been attacked** and that because of the attack, your email service provider is requesting all users change their password. ***This is your main email account and contains very sensitive information.*** Please **select** a new password below.

They are then required to create a new password and select a new password for their account (the selection and creation process was intertwined within each scenario; i.e., bank scenario select password $\rightarrow$ bank scenario create password $\rightarrow$ email scenario select password $\rightarrow$ email scenario create password). Participants were given 5 different passwords to select from in the selection scenarios, each of which were with a Zxcvbn strength estimation between 0-4 to choose from. After these 4 activities, participants would have created 2 passwords and selected 2 passwords.

Two separate scenarios were given to observe whether there was a different effect on the passwords selected/created for a participant's banking passwords versus their email password. This could help shed light on a personality's heightened interest in security, if any, depending on the scenario. Conscientious individuals for example, tend to have a stronger grasp on finances and direction. Extroverted individuals tend to enjoy a more social setting, perhaps implying that more Conscientious individuals could put a higher value on their banking account security, rather than a social email password.

The password selection activities were included alongside the creation activities to examine whether participants tended to select and create similarly-strong passwords. The limitations of the experiment from including these activities are discussed in Section 5.2.1.

Personality Test.

The last test given to participants was a short 20 question personality test based on the

Mini-IPIP Big-five personality indicator test discussed in Section 3.1.1. Every question in this test had the following options for an answer:

1. Very Inaccurate
2. Inaccurate
3. Neither Inaccurate or Accurate
4. Accurate
5. Very Accurate

The 20-question test is provided in Appendix A.1.5. Upon completion of this test, each participant's personality type is determined within the 5 personality indicator domains. The scale of Extroversion among participants is depicted in Figure 3.2.

3.1.2 Amazon's Mechanical Turk

Tasks on Amazon's MTurk (e.g., transcribing text) are typically completed within minutes and usually pay in cents rather than dollars [26]. Although payment is an important factor, self-reports indicate that workers are driven by both extrinsic and intrinsic motives (e.g., workers have reported that they complete tasks "to make basic ends meet" and because "tasks are fun" [55]), suggesting that the rewards of working on MTurk are not merely monetary.

MTurk has a growing presence in the psychological literature as a source of research participants, and researchers in some fields, such as human computation, have examined work experiences on MTurk. We apply knowledge from long-term in-person work relationships traditionally studied in industrial-organizational (I-O) psychology to the very short-term online work experiences of crowdsourcing [8]. This has given MTurk a huge degree of presence and validity within recent studies.

Although in person samples have long been a reliable source of data collection, recent

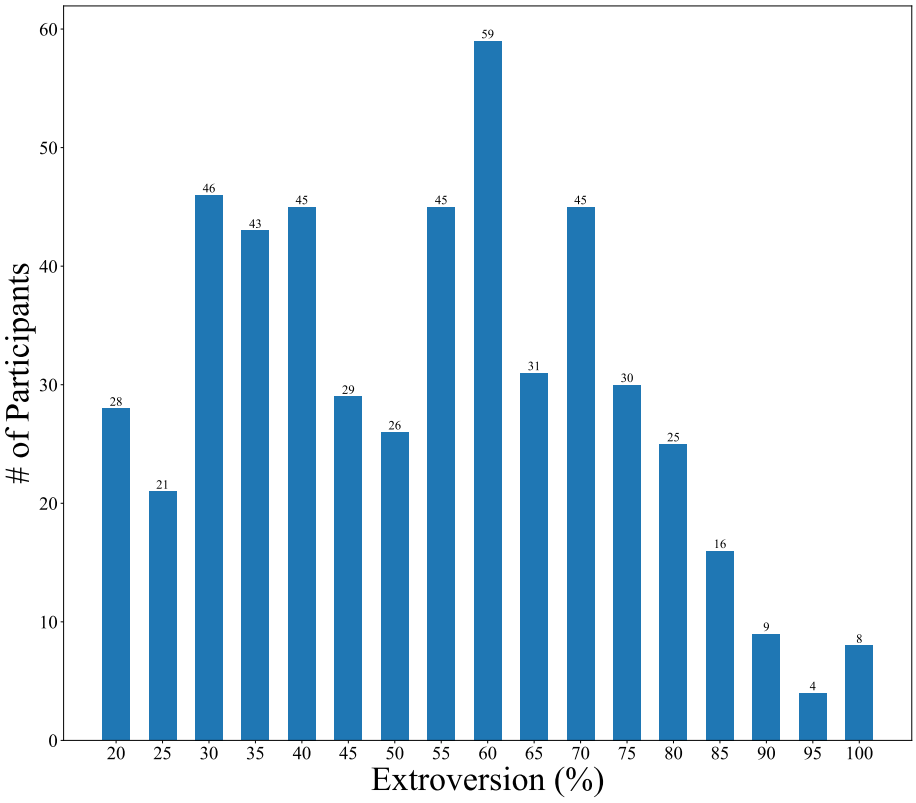

Figure 3.2: Extroversion/Introversion in Participants

evidence suggests that collecting data via the Internet, although far from perfect, can reduce the biases found in traditional samples [31]. Buhrmester M. et al. [10] conducted a study in 2011 to examine how MTurk samples compare with the diversity of standard Internet samples. 3,066 participants from over 50 countries around the world yielded a very demographically diverse distribution of people from around the world with varying cultures and backgrounds. This was found to be significantly more diverse than typical American college samples.

Participants were recruited from several English speaking countries including The United States, Canada, UK, and Australia. This increased the chance that all participants were fluent in English to understand the questions and scenarios within the experiment.

Demographics.

Demographic analysis was done on the 510 participants which remained after the data filtration step discussed in Section 3.2. More than 140 participants were over the age of 40. The majority of participants was distributed normally between the ages of 20 and 40. Out of all the participants, 55.5% were male and 44.5% were female. Roughly 10% of all the participants were left handed. Although the slight majority of participants answered to an occupation within "Business, Executive, Management, and Financial" (15.6%), "Computer Science and IT-Related" (14.5%), "Education, Training, and Library" (11.4%), and "Healthcare Support" (6.8%) occupations, more than 160 of the total participants responded to being in "Other" (33.5%) occupations.

The "Password Awareness" question was answered at almost a 50% rate between "yes" and "no". 70.2% of participants answered "no" to the "Security Training" question, whereas 69.8% of participants answered "yes" to the "Account Hijacking" question.

3.1.3 Passwords

All data used for the analysis portion of this work has been generated from the experiment through the several activities as described in Section 3.1.1. For the password ranking and password selection activities, *real* passwords were used (password origins discussed below).

The passwords used within this experiment were drawn from the 2012 LinkedIn password leak where 6.5 million passwords were exposed [40]. Out of these, a random subset of 22,000 passwords were taken and brute force attacked to recover the true plaintext passwords. The tool used for this is called Uniqpass, which includes over six billion entries within a rainbow table used to brute force hashed passwords. These passwords were then ranked by Zxcvbn (strength between zero and four).

To streamline the types of passwords that were provided to participants during the activities, several filtration steps were taken:

- 50 passwords were randomly taken from every rank.
- Passwords that were not in "English" were discarded.
  - **Note:** Many passwords with rank four in Zxcvbn mostly consist of numbers and letters, in seeming meaningless order; we deem those as non-English.
- Only passwords that fit the criteria in Table 3.1 were taken from every rank.

Passwords used by the LinkedIn leak in this research seemed to be composed of words from various dialects such as English and Spanish. To ensure only English-based passwords were provided to participants, each password was manually checked. The reasoning behind this was to involve only English speaking countries through MTurk's settings and thus only English based passwords.

By the end of this filtration, we had 98 passwords in total between all the ranks. All of the resulting passwords which ranked 0-3 from Zxcvbn were comprised of English words and numbers, whereas the selected passwords ranked 4 consisted of numbers and letters in

Rank	Slow hash time*	Composition
0	$<$ 1 second	8–12 letters
1	1–5 seconds	7–8 letters, 0–1 digit
2	1–60 minutes	7–8 letters, 0–1 digit
3	5 hours – 5 days	4–8 letters, 2–4 digits
4	5 months – 5 years	8–10 letters, 2–5 digits

Table 3.1: Selected Password Guidelines
** "Slow hash time" is the simulated time it takes the Zxcvbn algorithm to crack a password using an offline attack with a slow hashing function, such as bcrypt, PBKDF2 and scrypt.*

seeming meaningless order.

3.2 Data Pre-processing

The data captured within this experiment came in a variety of numeric and textual format. Pre-processing was required to clean and convert the data to ensure it was usable in the analysis phase. The following section explains this process along with results found in correspondence with our hypotheses.

Data pre-processing can often have a significant impact on the generalization of data, especially for algorithms used for analysis on the data. Usually the removed instances of noise have excessively deviating instances that have too many null feature values. These excessively deviating features are also referred to as outliers. Out-of-range data is the most difficult problem to detect. This is for cases in which the data for particular parameters does not contain a meaningful value (e.g., personality dimension with 0%).

Due to the concise nature of the required data for this experiment, much data pre-processing was not required, with the exception of noisy and unfinished responses.

We used a few questions for sanity-check of participants' responses. Some participants

also decided to exit the experiment before finishing. This left holes in the entire data for certain records, which could not be used. Based on this filtration, we discarded responses from seven participants. In the end, we kept responses from 510 participants. We discuss the filtration and data pre-processing steps below.

The symbolic, logical learning statistical algorithms used to perform analysis on most data are able to process symbolic, categorical data only. However, real-world problems such as the one that exists in this research involve both symbolic and numerical features. Therefore, there is an important issue to discretize numerical (continuous) features [43]. Moreover, in real-world data, the representation of data often uses too many features. but only a few of them may be related to the target concept. There may be redundancy, where certain features are correlated, which drives the motivation to create some new features from existing data for analysis purposes. A few of these features including "Ranked Score Distance", "Ranked Score +/-", and "Average Score" are discussed below.

Due to the complex nature of the data, various sub-objects and arrays were nested to handle the variety of data types for the experiment activities in JSON format. The resulting data was broken into several files, which represented the required activity-individual data in a tabular fashion as seen in Fig. 3.3.

These files were then filtered through, and every list of strings were converted to respected integers (e.g., "Yes" became 1 and "No" became 0). This was done on every string-related answer apart from the participants' justifications. For cases where the answers spanned more than just Yes/No, respective integers were assigned (e.g., "Not very likely", "Somewhat likely", "Likely", and "Very likely" would become 0, 1, 2, and 3 respectively). This process is called One-hot encoding and is required for data analysis software and is largely used in machine learning applications.

Two valuable metrics were introduced to score how each participant performed on the password ranking activity. Although the analysis was done on each individual bucket within

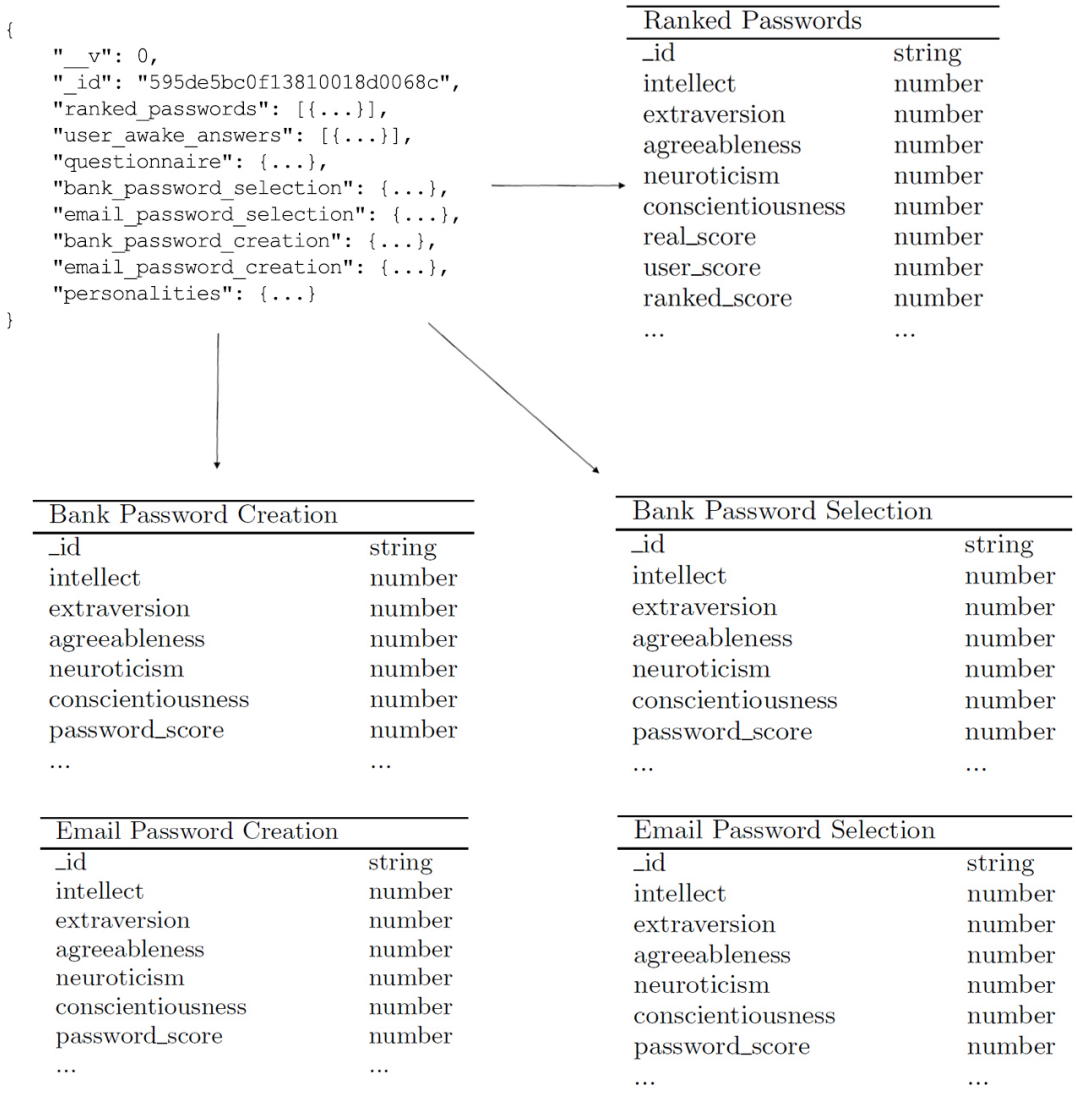

Figure 3.3: Data Pre-processing from document to tables.

the activity, we wanted to have a metric to measure overall performance.

The "Ranked Score Distance" can be somewhat confusing due the inverse nature of the scoring (i.e., the higher the score, the poorer the participant performed). In the rest of this research, a higher score implies a stronger password. To ensure consistency, another metric was created to capture overall participant performance called "Ranked Score +/-". Every participant started with a "Ranked Score +/-" of 0 and for every password that was placed in the wrong bucket, the participant lost a point. For every password that was placed in the right position, the participant gained a point. This means that the lowest a participant could have received with this score is -5 and the highest is 5.

An "Average Score" metric was also created and worth noting as an average of the four passwords created and selected (i.e., two passwords from the bank and email *creation* scenarios and two passwords from the bank and email *selection* scenarios). The average was taken of all four strengths assigned to these passwords from the Zxcvbn password strength meter. Although analysis was completed using this metric as a potential criterion, no significant findings were made. This is discussed more in Section 4

Data pre-processing took a somewhat different approach for the password characteristics analysis. Feature engineering was done to create new fields for the email and bank scenarios. All the created and selected passwords (four in total per participant - two for the email scenarios and two for the bank scenarios) were taken, and 4 new fields were derived from each created email password - "Number of characters", "Number of letters", "Number of digits", and "Number of symbols". The "Number of symbols" field was excluded from analysis for the selected email and bank passwords; we exclude symbols in the passwords that were provided to the participants to select from.

Chapter 4

Analysis and Results

4.1 Analysis

The collected data was brought together as a result of the 6 main activities explained in Section 3.1.1. We took all the data and broke the analysis portion into 3 major groups for primary analysis discussed below.

4.1.1 Data analysis on password strength

Analysis on the dataset was done on IBM's SPSS [65]. After pre-processing the data, individual files were fed into the software and a bivariate Pearson correlation analysis test was run on the data of interest. As we were conducting this on two variables at a time, the bivariate test was required. r refers to the correlation strength, p is the significance, and N is the sample size.

The analysis for our primary results was split into three main sections:

1. Created Passwords

- Analysis on created email password scenario
- Analysis on created bank password scenario

2. Selected Passwords

- Analysis on selected email password scenario
- Analysis on selected bank password scenario

3. Ranked Passwords

- Analysis on each individual ranked password bucket.

Analysis for the exploratory results were broken down in a similar fashion, with the addition of performing analysis on password characteristics.

Within the created and selected password scenarios, the created and selected passwords were awarded a strength between zero and four using Zxcvbn, which was then used for analysis purposes: for example, does the strength of participants' passwords increase as they become more extroverted?

The ranked passwords were treated as five individual tests (i.e., analysis for each password bucket). The analysis was done to investigate the distance between the Zxcvbn estimated password strength and the bucket where a given password was placed into. For example, if a participant placed a password of real strength four into the first bucket, they would have a very poor score. In contrast, a participant who placed a password of real strength zero in the first bucket would have a very good score.

4.1.2 Data analysis on password characteristics

Similar to the analysis done on the password strengths, analysis for password characteristics were done on IBM SPSS using bivariate Pearson correlation analysis.

The characteristics for the passwords were then broken into two main processes: password creation and password selection; each process was further divided into two parts: email and bank scenarios. The user-created passwords had four attributes: "Length of password", "Number of Digits", "Number of Characters" and "Number of Symbols". The "Number of Symbols" field was excluded for selected passwords, as we did not provide

any password with symbols.

4.2 Results

Results for all the test which yielded significant correlations are split into two sections below. Primary Results are results which correspond with the aforementioned hypotheses for this research. Exploratory Results are results worth mentioning that did not necessarily align with the hypotheses, but are still significant.

4.2.1 Primary Results

For the email password creation scenario, we found a significant correlation between Extroversion and password strength ($r = 0.184$, $p < 0.05$), but only when participants answered "No" to "Account Hijacking" (i.e., whether participants were required to change an online account password in the past). In contrast, when participants answered "Yes" to this question, there was a mild, but non-significant, negative correlation between Extroversion and password strength; see Table 4.1. Correlations remained very similar (albeit less strong/significant) with the corresponding bank password scenario; see Table 4.2.

Trait	r (corr.)	p (sig.)
Openness	-0.041	0.614
Conscientiousness	0.065	0.421
Agreeableness	-0.042	0.602
Extroversion	0.184*	0.023
Neuroticism	-0.059	0.464

Table 4.1: Email password creation with no account hijacking: Correlation between password created in the email scenario when participants answered "No" to "Account Hijacking", N = 154. * = significant with $p < 0.05$.

Trait	r (corr.)	p (sig.)
Openness	-0.015	0.780
Conscientiousness	0.018	0.736
Agreeableness	-0.079	0.139
Extroversion	-0.099	0.062
Neuroticism	-0.028	0.600

Table 4.2: Email password creation with past account hijacking: Correlation between password created in the email scenario when participants answered "Yes" to "Account Hijacking", N = 356.

For the email password selection scenario, we found a significant correlation between Extroversion and password strength (r = 0.116, $p < 0.001$) for all cases; see Table 4.3. This indicates that when all participants are tasked with creating or selecting a new password for their online email account, participants who are more extroverted created and selected stronger passwords, with a stronger correlation existing for the created email password.

This could be indicative of the creativity trait of Extroverts coming in to play, but would also fall out of line with the expectation that a similar result would exist in the bank creation scenario, which was inconclusive; see Table 4.4.

Trait	r (corr.)	p (sig.)
Openness	0.043	0.335
Conscientiousness	0.046	0.302
Agreeableness	-0.014	0.757
Extroversion	0.116**	0.009
Neuroticism	-0.036	0.421

Table 4.3: Email password selection: Correlation between password selected in the email scenario for all participants, N = 510. ** = significant with $p < 0.01$.

Trait	r (corr.)	p (sig.)
Openness	0.033	0.681
Conscientiousness	0.022	0.785
Agreeableness	0.008	0.921
Extroversion	0.171*	0.033
Neuroticism	-0.065	0.425

Table 4.4: Email password selection with no account hijacking: Correlation between password selected in the email scenario when participants answered "No" to "Account Hijacking", N = 154. * = significant with $p < 0.05$.

Findings from the bank scenarios are reported in Appendix A.1.1.

Trait	r (corr.)	p (sig.)
Openness	-0.007	0.889
Conscientiousness	-0.017	0.755
Agreeableness	-0.056	0.289
Extroversion	0.044	0.405
Neuroticism	-0.136*	0.01

Table 4.5: Password Ranking on the "Very Weak" bucket: Correlation between password placed in the "Very Weak" bucket in the password ranking activity when participants answered "Yes" to "Account Hijacking", N = 356. * = significant with $p < 0.05$.

Note that, for brevity, we omit the tables without any significant correlation.

It is worth noting that results from the password ranking analysis were inconclusive with very little correlation between the real and ranked score as given by Zxcvbn and the participants respectively. One finding worth mentioning is discussed in the exploratory results section below.

4.2.2 Exploratory Results

Due to the diverse nature of the data collected, several exploratory analyses were also undertaken. Findings reported in this section do not qualify as primary results (as they were not directly related to our specific hypotheses), and do not necessarily pass the Bonferroni Correction test (explored more in Section 5.2). Nevertheless, we discuss these findings to assess their relevance for future research directions.

Password Strength

Password Creation and Selection. Within the bank password creation activity, when participants responded "No" to "Security Training" and "Password Awareness", there was a significant negative correlation between Agreeableness and password strength (r = -0.171, $p < 0.01$, N = 229).

Within the bank password selection activity, when participants responded "No" to "Security Training", there was a significant positive correlation (r = 0.104, $p < 0.05$, N = 358) between Extroversion and password strength.

To summarize, when participants answered "No", to "Security Training" or in other words - when participants have never had any sort of formal security training in the past, those high in Agreeableness tended to create weaker passwords in the bank creation scenario. When participants answered "No" to both the "Security Training" and "Password Awareness" questions or in other words - when participants have never had any sort of formal security training or password strength awareness training of any sort, those higher in Extroversion tended to select stronger passwords in the bank selection scenario.

"Password Awareness" was also analyzed, but no significant findings could be concluded.

Password Ranking. For password ranking, a significant positive correlation was found between Openness and the ability to rank the weakest provided password correctly (r = 0.083, p = 0.03, N = 510). This implies, those who displayed more Openness tended to rank the weakest given password correctly. Analysis on the rest of the buckets yielded no significant results, but is included in Appendix A.1.2.

Password Characteristics

Password Length. Within the bank password selection and creation scenarios, it was found that Extroversion was significantly correlated with creating shorter passwords (i.e., passwords with a lower number of total characters) when participants answered "No" to "Account Hijacking" (r = -0.180, p = 0.025, N = 154 and r = -0.241, p = 0.003, N = 154 for the bank password creation and selection scenarios respectively). Additionally, there was a significant correlation between Extroversion and creating passwords with less letters in the bank password creation scenario. This means that as extroverts tend to create shorter passwords comprised of less letters as opposed to numbers and symbols (see Tables 4.6 & 4.7).

Trait	r (corr.)	p (sig.)
Openness	-.196*	.015
Neuroticism	.086	.286
Conscientiousness	-.052	.523
Agreeableness	-.121	.135
Extroversion	-.180*	.025

Table 4.6: Bank Creation Password Length

*Correlation between the password created in the bank scenario and the length of the password when participants answered "no" to "Account Hijacking". N = 154. * = significant with $p < 0.05$.*

Trait	r (corr.)	p (sig.)
Openness	-.027	.742
Neuroticism	.151	.062
Conscientiousness	-.044	.592
Agreeableness	-.103	.203
Extroversion	-.241**	.003

Table 4.7: Bank Selection Password Length

*Correlation between the password selected in the bank scenario and the length of the password when participants answered "no" to "Account Hijacking". N = 154. ** = significant with $p < 0.01$.*

Number of Letters. Analyzing the bank password creation activity, between participants who answered "No" to "Account Hijacking", it was found that Openness was significantly correlated with creating passwords with less letters as opposed to numbers and symbols (r = -0.222, p = 0.006, N = 154). This means that participants who displayed a higher level of Openness, created passwords with more numbers and symbols as opposed to a higher number of letters (see Table 4.8).

Trait	r (corr.)	p (sig.)
Openness	-.222**	.006
Neuroticism	.062	.446
Conscientiousness	-.012	.881
Agreeableness	-.110	.175
Extroversion	-.182*	.024

Table 4.8: Bank Creation Password Number of Letters

*Correlation between the password created in the bank scenario and the number of letters in the password when participants answered "no" to "Account Hijacking". N = 154. * = significant with $p < 0.05$, ** = significant with $p < 0.01$.*

Number of Symbols. Within the password creation activities (bank and email) and when participants answered "No" to "Account Hijacking", it was found that Conscientiousness was significantly correlated with creating passwords with less symbols (r = -0.310, $p <$ 0.001, N = 154 and r = -0.189, p = 0.019, N = 154 for the bank and email scenarios respectively). Although not thoroughly investigated, it could be assumed that these participants who displayed a higher level of Conscientiousness tended to create passwords with more digits and letters as opposed to a higher number of symbols (see Tables 4.9 & 4.10).

Summary of Password Characteristics Results. Demonstrated by the results shown above, it has been observed that there is only a correlation between the base composition of passwords and certain personality dimensions when participants answered "no" to "Account Hijacking" or in other words - when participants have never been exposed to creating a new password for an online account as a result of a security breach. It is found that during the bank creation scenario, created passwords were shorter and generally contained less letters as opposed to digits or symbols for participants who demonstrated higher levels of Openness. We also found that during the email creation scenario, the passwords created

by more Conscientious participants had less symbols as opposed to digits and letters. This extended into the bank creation scenario where the passwords created had significantly less symbols. It was also found that there was a very significant correlation between selecting shorter passwords and Extroversion in the bank selection scenario.

Results reported in this section are considered as tangible outcomes due to the direction they open for future research. However, because this portion of the research is exploratory in nature, false discovery must be accounted for. After applying the Bonferroni correction, the only correlation that successfully passed was that of Conscientiousness and lower number of symbols in created passwords. Requiring to adjust the new α threshold in correspondence with the number of tests run (20 for exploration due to 5 personality dimensions and 4 password characteristics), the new α threshold become 0.0025 ($0.05/20$). A few other correlations came extremely close and are worth noting because the Bonferroni correction is susceptible to Type-2 errors. Although the rest of the correlations do not pass after the Bonferroni correction, we still choose to report them as they cannot be overlooked due to the possibility of Type-2 errors after the correction.

All analysis ran only tested for linear correlations within the collected data. However, it may be the case where some more interesting findings appear when accounting for polynomial based correlations. Though no definite tests were conducted for this, visual outputs from significant password characteristics findings show that there may be a possibility for these correlations to exist (See Figures 4.2, 4.2, 4.3, 4.4 and 4.5). Although a distinct linear correlation can be found in every visualization, a slight parabolic correlation can be seen in Figures 4.2, 4.4 and 4.5, where average number of letters, average password length and average number of symbols increase as Openness, Extroversion and Conscientiousness increase beyond 85%. A summarization of all the noteworthy password characteristics results explained above can be found below in Table 4.11.

Trait	r (corr.)	p (sig.)
Openness	.022	.791
Neuroticism	.071	.382
Conscientiousness	-.310**	0.00009
Agreeableness	-.106	.191
Extroversion	-.045	.576

Table 4.9: Bank Creation Number of Symbols

*Correlation between the password created in the bank scenario and the number of symbols in the password when participants answered "no" to "Account Hijacking". N = 154. ** = significant with $p < 0.01$.*

Trait	r (corr.)	p (sig.)
Openness	.016	.844
Neuroticism	.073	.371
Conscientiousness	-.189*	.019
Agreeableness	-.090	.267
Extroversion	-.060	.463

Table 4.10: Email Creation Number of Symbols

*Correlation between the password created in the email scenario and the number of symbols in the password when participants answered "no" to "Account Hijacking". N = 154. * = significant with $p < 0.05$.*

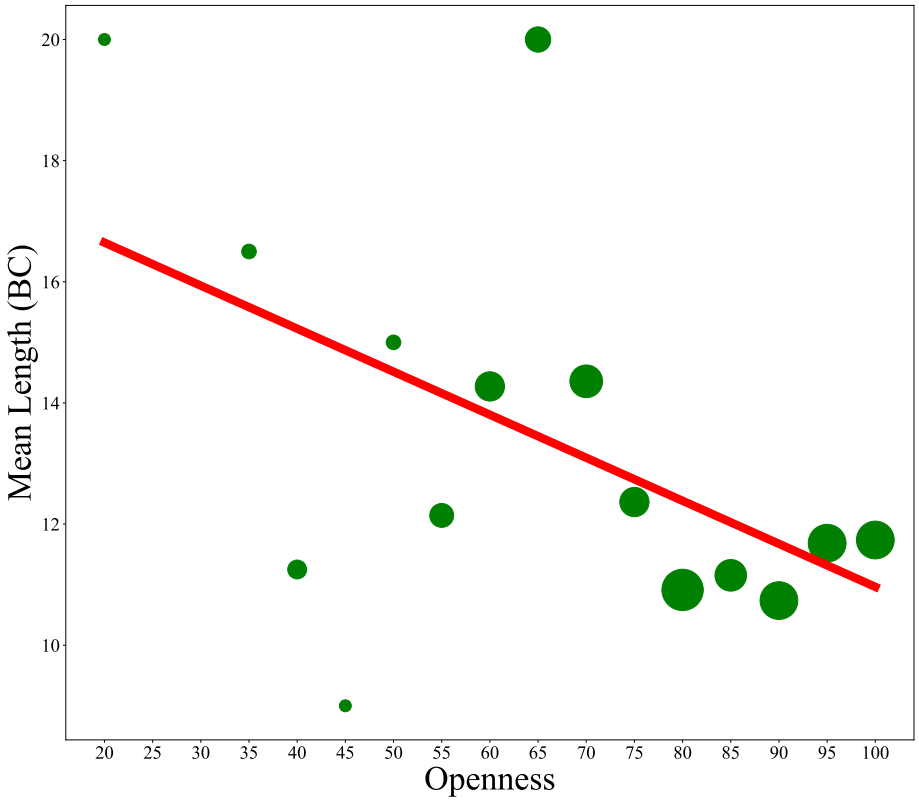

Figure 4.1: Openness vs. Password Length (BC): Y-axis: Average length of passwords created in the bank creation (BC) password scenario. X-axis: Amount of Openness demonstrated by participants out of 100%. Larger circles indicate more samples at that timestep.

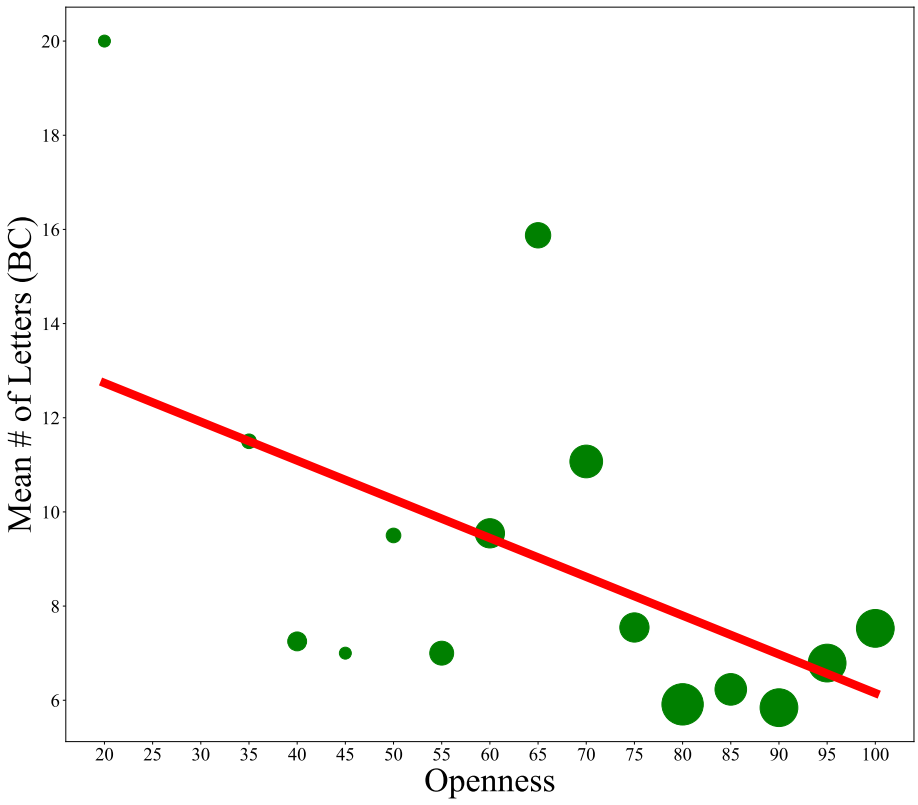

Figure 4.2: Openness vs. Password # of Letters (BC): Y-axis: Average number of letters used in created passwords within the bank creation (BC) password scenario. X-axis: Amount of Openness demonstrated by participants out of 100%. Larger circles indicate more samples at that timestep.

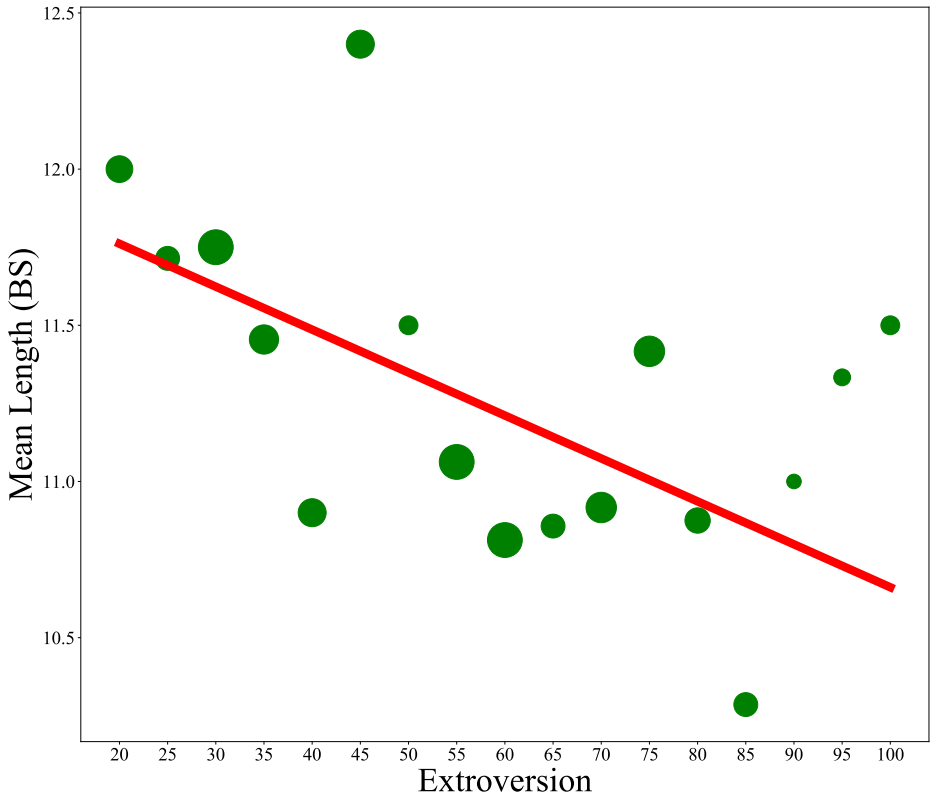

Figure 4.3: Extroversion vs. Password Length (BS): Y-axis: Average length of passwords selected in the bank selection (BS) password scenario. X-axis: Amount of Extroversion demonstrated by participants out of 100%. Larger circles indicate more samples at that timestep.

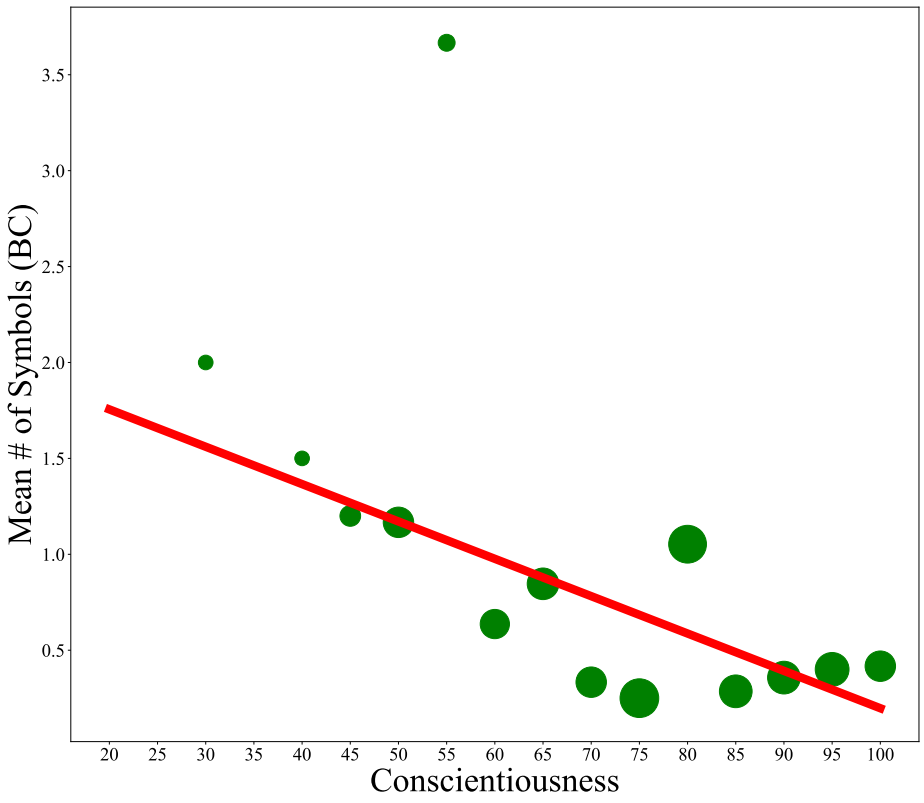

Figure 4.4: Conscientiousness vs. Password # of Symbols (BC): Y-axis: Average number of symbols used in created passwords within the bank creation (BC) password scenario. X-axis: Amount of Conscientiousness demonstrated by participants out of 100%. Larger circles indicate more samples at that timestep.

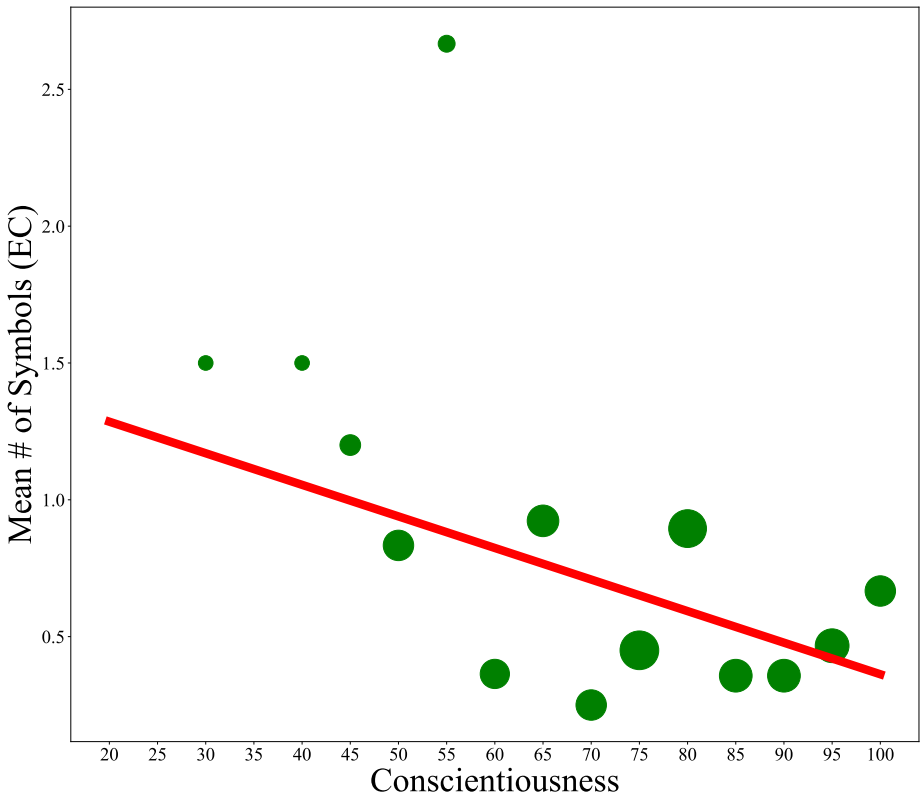

Figure 4.5: Conscientiousness vs. Password # of Symbols (EC): Y-axis: Average number of symbols used in created passwords within the email creation (EC) password scenario. X-axis: Amount of Conscientiousness demonstrated by participants out of 100. Larger circles indicate more samples at that timestep.

Trait	Password Length	# of Letters	# of Digits	# of Symbols
Openness	$\downarrow$ (BC)	$\downarrow\downarrow$ (BC)	-	-
Conscientiousness	-	-	-	$\downarrow$ (EC) $\downarrow\downarrow$ (BC)
Extroversion	$\downarrow\downarrow$ (BS)	-	-	-

Table 4.11: Password Characteristics Summary

Password characteristics results summarization table. All results are recorded for participants who answered "no" to "Account Hijacking". BC *= Bank Creation Scenario,* EC *= Email Creation Scenario,* BS *= Bank Selection Scenario.* N *= 154.* $\downarrow$ *= significant with* $p < 0.05$*,* $\downarrow\downarrow$ *= significant with* $p < 0.01$

Chapter 5

Conclusions

5.1 Ecological Validity

To ensure quality of data and participants, some design choices within the experiment were considered and are discussed within this section.

5.1.1 Data Validation

We intentionally created overlap in some of the binary questions that were asked to participants. For example, we expected that most users who answered "Yes" to "Security Training", may also answer "Yes" to "Password Awareness". It was assumed that the "Security Training" and "Password Awareness" questions would have a direct correlation in participant responses. Upon investigation, we found a strong correlation of 0.446 ($p < 0.001$) between participants who answered positively to being security trained and trained in password strength awareness.

We also expected that there may be a positive correlation between participants who answered "Yes" to "Password Awareness" and their computer skills level. This hypothesis was based on the tentative notion that participants who spend more time on the computer

may have been required to create and/or change more online passwords. Upon investigation, we found such a positive correlation (0.143 ($p < 0.01$)) between participants who answered in the affirmative to being trained in password strength awareness and having a stronger computer skill level.

5.1.2 Participant Validation

In addition to the 20-item IPIP, three additional sports-related questions were included, to make sure that participants were not simply clicking on the same answer over and over again to race through the test. Two of these questions asked participants what their favourite / least favourite sport to watch was. The last question asked participants to what degree they enjoyed watching sports. Depending on the variance of their answers for these three questions, they were automatically discarded from the filtered results. Three of the 513 participants were discarded from this test.

Of the remaining 510 participants, 51 reported being left handed, which falls in line with 10% of the population. This question was given at the beginning of the test, adding further validation that there was a random population of people, though MTurk solves many issues in relation to bots on their platform. Collecting responses from around the world through the MTurk platform, on top of the diversity of occupational responses (other than roughly 33% of participants) add another level of randomness in responses.

Due to the number of countries the experiment was outsourced to, mother tongues of the participants had to be taken into consideration. The Zxcvbn password-ranking test estimates password strength based on an English dictionary. This means that any passwords given in another language may be classified as stronger than they actually are. For example, the password "unitedstates" receives a Zxcvbn strength estimation of 0, but the corresponding Spanish "estadosunidos" receives a Zxcvbn strength estimation of 4. Because of this, we discarded all the passwords which were not in English before providing them to the

participants in the selection and ranking activities.

5.2 Discussion

Due to the subjective nature of some of the questions posed in the preliminary questionnaire and situational activities, there are variables that must be taken into consideration when reflecting on the results within this work. We discuss a few of the related issues here.

Experiment Design. The experiment was designed to give a similar experience to each participant. However, it could be observed that an earlier activity may influence the decisions of participants throughout later activities. For example, a participant may be exposed to several passwords in the password ranking activity, which they then choose to re-use in a later password creation activity. Although this was meant to be alleviated by prompting participants to create their own passwords, we acknowledge that some activities may have primed participants in a certain direction. Future research looking to replicate this experiment to some degree may want to consider randomizing the order of the activities in the experiment.

We acknowledge that the randomization of activities in the experiment could have introduced a level of randomness in created password results rather than a limitation of participant priming when always shown examples of passwords first. Another option would be to provide participants with the password ranking activity along with the scenario-based password selection activities after the creation activities.

Questionnaire. The *"Password Awareness"* question in the preliminary questionnaire is somewhat subjective in the sense that participants may answer "No" even if they know the difference between a stronger and weaker password to some extent. The reasoning behind this is the interpretation of "password security awareness training". The subjectivity in this being a formal or informal process raises consideration. For example, a participant

may have been required to create a new password conforming to creation rules such as, requiring 2 digits, an uppercase letter, and a symbol with a password strength meter in the past. The participant may not consider this as any sort of password training, however this process indirectly teaches a participant the difference between a weaker and stronger password.

Both email scenarios given to participants within the password selection and creation activities included an excerpt, "***This is your main email account and contains very sensitive information.***" Some participants may not actually use their email as a personal account to store sensitive information, hence these participants may not be able to relate to this scenario as strongly as intended.

Within both the email and bank scenarios, participants were instructed not to use the same password for both instances. The reasoning behind this was to ensure participants were treating both situations as two completely different accounts; but we acknowledge that even mentioning this could have the possibility to skew data in an unauthentic fashion. Although it is understood that we may have primed the participant to some degree and put them on the spot, what we gained from doing this was a more diverse dataset and possibly varying levels of strength within the two created passwords.

Password Ranking Activity. The "Ranked Score Distance" and "Ranked Score +/-" metrics that were developed for the password ranking activities came with a couple caveats. As mentioned earlier in the paper, the "Ranked Score Distance" is confusing as it contradicts the Zxcvbn philosophy of ranking stronger passwords with higher numbers. This could be alleviated by taking the inverse of the calculated "Ranked Score Distance" (i.e., exponentiating by -1), but this is becomes an exponential function with an asymptote at 0 - failing to capture the larger penalty for bigger errors. For example, a "Ranked Score Distance" of 8 would become 1/8, whereas a "Ranked Score Distance" of 4 would become 1/4. Before being inversed, the first score would be twice as poor as the second, but after being inversed,

it is only worse by 25%.

In this research, the "Ranked Score Distance" becomes larger, the worse a participant performed. The "Ranked Score +/-", although alleviating the flaw of the "Ranked Score Distance" fails to capture the magnitude of how poorly a participant performed as every wrong answer is only penalized with a deduction of 1. For example, if participant A placed a password with an estimated Zxcvbn strength of 2 into the "Very Weak" bucket and placed another password with an estimated Zxcvbn strength of 3 into the "Weak" bucket, they will receive a "Ranked Score Distance" of 4 (2 - 0 + 3 - 1). If participant B placed a password with an estimated Zxcvbn strength of 2 into the "Weak" bucket and placed another password with an estimated Zxcvbn strength of 3 into the "Normal" bucket, they will receive a "Ranked Score Distance" of 2 (2 - 1 + 3 - 2), which depicts a poorer performance than participant B. In contrast, both participants would receive an identical "Ranked Score +/-" of -2 for getting both passwords wrong.

In regards to selecting an appropriate metric for the "Ranked Score", mean square error (MSE) was also a viable candidate, but did not add any additional benefits compared to "Ranked Score Distance". MSE also do not capture the plus-minus scale, which was desired as explained above.

We looked into the utilization of both of the aforementioned metrics ("Ranked Score Distance" and "Ranked Score +/-"), but neither resulted in a significant correlation. A Pearson correlation was calculated on the two metrics and resulted in an r of 0.933 and $p < 0.0001$, indicating both metrics are almost identical.

Password Selection.

Our experiment included a password selection task right before password creation (i.e., a bank password selection before the bank password creation and the email password selection before the email password creation), where participants were greeted with 5 varying levels of passwords to select from, each of which have different strengths of security as

ranked by Zxcvbn. We acknowledge that this might have had a priming effect in participants. Nevertheless, such priming would have been consistent across all of the participants, making us believe that it would not have altered the correlations between password characteristics and personality traits. Further studies are needed to confirm our belief. Jeske et al. [36] concludes that "nudges can effectively and significantly change behavior" and although this was not confirmed in our experiment, we expected the varying levels of secure passwords in the selection phases would not have influenced the created password in the following creation phases. Though not exactly the same as, Ur et al. [71] reported that a combination of visual and text feedback was the most effective intervention in the design of password strength meters. No textual or visual feedback was given to participants after the selection activities. Furthermore, Jeske et al. [36] found that when Wi-Fi networks had the same colored font, while not being ordered by security, no influences on participant selection was observed. Passwords given to participants in the password selection activities were presented in a similar fashion (i.e., white font, grey background, random ordering).

Findings: Password Strength. Considering the anticipated analysis on password characteristics, the lack of inclusion of password policies for the password creation activities (bank and email) was of importance. Not having password policies telling participants to have a minimum number of characters, which included at least one letter, one digit, and one symbol for example allowed participants to freely create a password. Although almost all email providers and online bank sites implement some sort of password policy, the intent behind not having one was to treat the situation as generically as possible, since password policies differ significantly between sites. This also allowed us to perform the password characteristics analysis without much restriction.

To ensure an inadvertent cross-correlation was not observed in our tests, a point-biserial correlation analysis was run on the data between "Account Hijacking" and password strength score on all 4 scenarios (email creation, email selection, bank creation and bank selection).

It was found that no significant correlations between "Account Hijacking" and password score was present; see Table 5.1.

Activity	r (corr.)	p (sig.)
Bank Selection	-0.056	0.208
Bank Creation	-0.068	0.126
Email Selection	0.036	0.421
Email Selection	0.015	0.733

Table 5.1: Account Hijacking VS. Password Strength

Correlation analysis between password strength and "Account Hijacking" done with a point-biserial consideration.

Findings: Password Characteristics. Perhaps one of the most interesting portions of this research, the several correlations between *Extroversion*, *Openness* and *Conscientiousness* with password characteristics, have proposed very intriguing conclusions. We note that this portion of the research was only conducted on four (length, number of letters, number of digits and number of symbols) very low-level attributes that contribute to the make up of a password. This cannot be correlated with the strength of said passwords; i.e., it cannot be said that just because two participants include 5 digits and 5 symbols in their passwords, it can be assumed that they are of equal strength. For example, if participant A creates a password $1h3b2n7k5l$ and participant B creates a password $abcde12345$, their password strengths are vastly different. Using Zxcvbn as a benchmark for calculating the strength of these passwords, although both passwords are constructed in a very similar structure by means of our password characteristics analysis, participant A's password would receive a strength of 4 and participant B's would receive a strength of 0.

Findings discussed in Section 4.2.2 show that Openness correlated negatively with length and number of letters in created passwords. All participants in this correlation

were ones who responded “No” to “Account Hijacking”, implying they have never been required to change an online account password in the past due to a security breach. It could be worth discussing that those participants may have used their higher level of outer experience and knowledge to create a more diverse password (letters, numbers and symbols), but without paying much attention to the length of the password. Openness also correlates positively with creativity and originality, perhaps further justifying the creation a diverse, creative password as opposed to a longer one. No correlation was found between password length and number of letters used when participants answered “Yes” to “Account Hijacking”, which implies that participants who have a higher level of Openness seem to learn that a longer password with more letters also contributes to a stronger password.

Although solid evidence to several correlations were obtained, more substantial research in this field would be beneficial. Conducting this experiment yielded promising results in correspondence with the initial hypotheses. However, a couple points need to be taken into consideration such as multiple comparisons and false discovery rate.

To account for multiple comparisons and false discovery rate, the Bonferroni correction was run through the primary results, and all of them passed. However, the exploratory results were not strong enough to pass the Bonferroni correction – more work is necessary in this area to solidify confidence. It is also worth noting that these correlations are not indications of predictor variables. Although there is a strong correlation in the observed findings, it does not say that certain personality types are predictors of password strength, rather just a correlate.

In the study done on 117 undergraduate students by Landers et al. [46], it was mentioned that participants who displayed a lower level of Extroversion were strongly correlated with spending more time on the internet. This then begs the question: is it that introverts use weaker passwords for practicality, given that they seem to spend more time on the internet? This observation should be taken into consideration for future work.

Privacy Concerns. The nature of this research may give rise to a number of privacy concerns. With any sort of data collection, comes potential security risks and concerns. If companies choose to observe personalities of employees to help company security, that personality-based data would have to be considered as sensitive employee information and thus, handled correctly with adherence to appropriate laws and regulations. If online companies plan on collecting user personalities, it could create another barrier for users whereby they might find the sign-up process tedious and unnecessary. Some may even find it invasive in terms of required personal data. Concerns as viable as these must be addressed transparently and its intent should be made completely aware to users upon sign-up.

Future Applications. With further research, the results found in this research can help introduce not only proactive security measures, but reactive ones as well. A proactive application could be the augmentation of password policies and password meters to have a more personalized feel. The tendency to satisfy a password policy for the sake of creating an acceptable password often causes fatigue in users. User frustration can also spawn from not achieving a desirable score on a password meter with the reluctance to adopt a suggested password. When creating passwords, a web browser which is aware of a user's personality may be able to make more intelligent suggestions such as, "Try creating a password with your dog's name, the year it was born, followed by the first letter of 4 of your most loved ones." This could in turn, be more memorable, secure and acceptable than using "P@$$w0rd" instead of "password".

On top of companies and organizations employing a larger focus on employee personality types to further understand relative security behaviours, understanding password security tendencies within people could help the movement toward reactive measures to administer appropriate security training and awareness. Although Neuroticism and Agreeableness can very well play a role in digital authentication, it seems as though Extroversion, Conscientiousness and Openness are more immediate correlates, perhaps opening the door

to further, more specific research into these domains.

Observing that there are positive and negative correlations between Extroversion and password strength, it is very interesting that a personality trait that has been shown to be influential mostly in social settings has an effect on password authentication. We believe there is also the potential for more research in this field.

5.2.1 Limitations

This section is to discuss some apparent limitations in this research. All the limitations listed here are considered for future work and should be observed carefully for further research.

Experiment Activities.

Throughout the experiment, when participants were greeted with the creation activities, they were told to create a password that was different from the one created in the previous creation activity. The experiment was designed this way to determine whether certain personalities would treat their online accounts differently in terms of security. Although this assumes participants use different passwords for all their online accounts, which is not always the case. We acknowledge this as a limitation of our work. However, for the requirements of our research, analyzing two different passwords per participant provided more useful insights rather than not.

Another caveat of the second password creation scenario was requesting participants to "create the most secure password [they] feel comfortable using and [they'll] be able to remember." This then introduces the question, can users be trusted that they will really create a memorable password? Is it appropriate to assume participants are aware of what makes a memorable password? The motivation behind the statement in the activity was to have participants create a password they feel comfortable using, to replicate more of a real-life scenario. The experiment was not conducted with a follow-up experiment in

mind, which does not help with actually verifying participants remembered the password they created. Additionally, password memorability was not a key piece of our work. This is another limitation of our work, whereby participants creating passwords in this experiment may have only created a password for the sake of the experiment, rather than creating one they may use in a real life situation.

Our research also heavily depends upon the "Account Hijacking" question where participants were asked if they have ever had to change an online account password as a result of a security breach. This is under the assumption that a user would have remembered the incident had it ever occurred. If it is the case where a participant had been a part of a security breach, but just did not know it, that does not affect the outcome of the research; we are interested in participants who are aware of a security breach if any and as a result, are more password aware. Knowing whether participants are password aware is of particular interest in this study as it will help determine, to some degree, the level of exposure one may have had with password strength. This also does not cover the edge case where participants have had to change a password in the past due to forgetfulness or other natural causes. If one answered "Yes" to "Account Hijacking" when never being involved in a security breach, but rather because they changed a password in the past due to forgetfulness, it could have a very minor influence on results, however very unlikely.

5.3 Conclusion and Future Work

Personality traits have shown consistent relationships with a wide variety of human attitudes and behaviours, including humour preference, goal-setting and motivation. Moreover, these correlations span across a variety of technology-relevant behaviours as well, such as internet usage and social media usage. Our research bridges the gap between the Big-five personality traits and the strength of one's chosen password. We also conclude that

there is a distinct relationship between one's personality trait and the way they construct their password, with room for future expansion.

Our findings identified several important discoveries. Extroversion correlated positively with password strength, but only when the participant had not previously been required to change an online account password before. This confirms our first hypothesis, and suggests that there may be an important relationship between one's level of Extroversion, and their reaction to previous security breaches. Although not significant, our second hypothesis also received weaker support, as Extroversion related somewhat negatively to password strength when the participant had previously been required to change an online account password. This could suggest an important relationship between introverts and being more password aware when being exposed to a security incident in the past.

An extension of our work, briefly mentioned in Section 2, would be the similarities between personality types and password policies. Password policies have been shown to induce annoyance and/or increase the difficulty of password creation. A common trait within people who demonstrate higher levels of Extroversion and Openness is creativity, while Conscientiousness and Openness are linked to achievement. Creativity and achievement are both factors needed to conform to password policies and successfully create a password. It would be within reason to hypothesize that these three dimensions would have an effect on user's interactions with password policies. It also stands to reason that Conscientiousness would be highly correlated with responding positively to password policies and finding them useful as Conscientious individuals tend to seek out a sense of security. Creating a similar experiment to the one we have conducted with the addition of varying password policies per password creation activity would be a good fit for future work to test the previously mentioned.

Additional exploratory analyses indicated that Openness was positively correlated with the ability to distinguish between stronger and weaker passwords, when a user has been

required to change an online account password in the past (i.e., answering "Yes" to "Account Hijacking"). It was also observed that when a participant has to change an online account password for the first time, Extroversion is directly correlated with creating and selecting shorter passwords, Openness is directly correlated with creating passwords with fewer letters and more numbers and symbols, and Conscientiousness is directly correlated with creating passwords with fewer symbols.

The size of the current study was substantial at 510 participants; moreover, we tested password strength in several diverse ways. Nonetheless, future research could gain additional reliability by making use of a larger sample, or within real-world systems such as, web browsers which can collect a user's personality type, save it, then suggest passwords as required. Interesting future work in this category can include passively collecting data about a user's password habits when not in a test environment; observing how they construct passwords in a variety of ways such as, time it takes to create a password, how often they modify their password before confirming it and how many letters, digits and symbols they use when provided with a password policy versus without. This will help shed light on how certain personality traits react to passwords on different levels such as time, decisiveness, and conformity to password strength standards.

Another interesting addition to this work would be the investigation of more abstract password characteristics and their correlation to individual personality types; for example, whether users with a certain personality type are more likely to use a string of digits or a string of characters as their password. Our research looked into password characteristics from a low level only taking into consideration the number of characters, digits, letters, and symbols as opposed to a higher level observation such as the inclusion of whole words within the password. This would be a good indication as to how users create, select, and respond to passwords, and may help facilitate more secure and memorable passwords. Weir et al. [76] found varying types of strings and digits used within the password study they con-

ducted. VanDam et al. [72] from LastPass also reported findings denoting a large amount of participants using initials, friends, family names, significant dates, numbers, pet names, birthdays, their hometown and even their school name or mascot in created passwords. Using this information, a similar study could be conducted to observe created passwords and whether there is a tendency for a specific personality trait to create a password with significant constructs.

From a predictive standpoint, using all the data collected, it is of interest to observe whether or not some sort of predictive algorithm can be put in place to identify one's password strength profile based on their personality and previous experiences with passwords. Being able to determine how weak or strong one may create a password can help warn that person before they create a password on a site if needed, as a preventative measure.

Bibliography

[1] ADAMS, A., AND SASSE, M. A. Users Are Not the Enemy. *Commun. ACM 42*, 12 (Dec. 1999), 40–46.

[2] BARRICK, M. R., AND MOUNT, M. K. The Big Five Personality Dimensions and Job Performance: A Meta-Analysis. *Personnel Psychology 44*, 1 (Mar. 1991), 1–26.

[3] BARRICK, M. R., AND MOUNT, M. K. Autonomy as a moderator of the relationships between the Big Five personality dimensions and job performance. *Journal of Applied Psychology 78*, 1 (1993), 111–118.

[4] BONNEAU, J. The Science of Guessing: Analyzing an Anonymized Corpus of 70 Million Passwords. In *2012 IEEE Symposium on Security and Privacy* (May 2012), pp. 538–552.

[5] BONNEAU, J., HERLEY, C., OORSCHOT, P. C. V., AND STAJANO, F. The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes. In *Proceedings of the 2012 IEEE Symposium on Security and Privacy* (Washington, DC, USA, 2012), SP '12, IEEE Computer Society, pp. 553–567.

[6] BONNEAU, J., AND PREIBUSCH, S. The password thicket: Technical and market failures in human authentication on the web. In *WEIS* (2010).

[7] BORGATTA, E. F. The structure of personality characteristics. *Behavioral Science 9*, 1 (Jan. 1964), 8–17.

[8] BRAWLEY, A. M., AND PURY, C. L. Work experiences on mturk: Job satisfaction, turnover, and information sharing. *Computers in Human Behavior 54* (2016), 531–546.

[9] BROSTOFF, S., AND SASSE, M. A. "Ten strikes and you're out": Increasing the number of login attempts can improve password usability. In *Presented at: CHI 2003 Workshop on Human-Computer Interaction and Security Systems, Fort Lauderdale, Florida. (2003)* (Fort Lauderdale, Florida, Apr. 2003).

[10] BUHRMESTER, M., KWANG, T., AND GOSLING, S. D. Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data? *Perspectives on psychological science 6*, 1 (2011), 3–5.

[11] BURR, W. E., DODSON, D. F., AND POLK, W. T. *Electronic authentication guideline*. US Department of Commerce, Technology Administration, National Institute of Standards and Technology, 2004.

[12] CATTELL, R. B. The description of personality: basic traits resolved into clusters. *The Journal of Abnormal and Social Psychology 38*, 4 (1943), 476–506.

[13] CATTELL, R. B. *Description and measurement of personality.* Description and measurement of personality. World Book Company, Oxford, England, 1946.

[14] CATTELL, R. B. Confirmation and clarification of primary personality factors. *Psychometrika 12*, 3 (Sept. 1947), 197–220.

[15] CATTELL, R. B. Concepts and methods in the measurement of group syntality. *Psychological Review 55*, 1 (1948), 48–63.

[16] CHITTARANJAN, G., BLOM, J., AND GATICA-PEREZ, D. Who's Who with Big-Five: Analyzing and Classifying Personality Traits with Smartphones. In *2011 15th Annual International Symposium on Wearable Computers* (June 2011), pp. 29–36.

[17] COSTA, P. T., AND MCCRAE, R. R. The neo personality inventory.

[18] COSTA, P. T., AND MCCRAE, R. R. The revised neo personality inventory (neo-pi-r). *The SAGE handbook of personality theory and assessment 2*, 2 (2008), 179–198.

[19] COSTA JR., P. T., AND WIDIGER, T. A. *Personality disorders and the five-factor model of personality.* 1994.

[20] DE ALVARE, A. M. How Crackers Crack Passwords or What Passwords to Avoid. Tech. Rep. UCID-21515, Lawrence Livermore National Lab., CA (USA), Sept. 1988.

[21] DELL'AMICO, M., MICHIARDI, P., AND ROUDIER, Y. Password strength: An empirical analysis. In *INFOCOM, 2010 Proceedings IEEE* (2010), IEEE, pp. 1–9.

[22] DONNELLAN, M. B., OSWALD, F. L., BAIRD, B. M., AND LUCAS, R. E. The Mini-IPIP Scales: Tiny-yet-effective measures of the Big Five Factors of Personality. *Psychological Assessment 18*, 2 (2006), 192–203.

[23] DOUGLAS, H. E., BORE, M., AND MUNRO, D. Coping with university education: The relationships of time management behaviour and work engagement with the five factor model aspects. *Learning and Individual Differences 45* (2016), 268–274.

[24] FISKE, D. W. Consistency of the factorial structures of personality ratings from different sources. *The Journal of Abnormal and Social Psychology 44*, 3 (1949), 329–344.

[25] FLORENCIO, D., AND HERLEY, C. A Large-scale Study of Web Password Habits. In *Proceedings of the 16th International Conference on World Wide Web* (New York, NY, USA, 2007), WWW '07, ACM, pp. 657–666.

[26] GABRIELE PAOLACCI, AND JESSE CHANDLER. Inside the Turk: Understanding Mechanical Turk as a Participant Pool. *Current Directions in Psychological Science 23*, 3 (June 2014), 184–188.

[27] GAW, S., AND FELTEN, E. W. Password Management Strategies for Online Accounts. In *Proceedings of the Second Symposium on Usable Privacy and Security* (New York, NY, USA, 2006), SOUPS '06, ACM, pp. 44–55.

[28] GENE M. SMITH. Usefulness of Peer Ratings of Personality in Educational Research. *Educational and Psychological Measurement 27*, 4 (Dec. 1967), 967–984.

[29] GOLDBERG, L. R. The structure of phenotypic personality traits. *American psychologist 48*, 1 (1993), 26.

[30] GOLDBERG, L. R. A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality psychology in Europe 7*, 1 (1999), 7–28.

[31] GOSLING, S. D., VAZIRE, S., SRIVASTAVA, S., AND JOHN, O. P. Should we trust web-based studies? a comparative analysis of six preconceptions about internet questionnaires. *American psychologist 59*, 2 (2004), 93.

[32] GRATIAN, M., BANDI, S., CUKIER, M., DYKSTRA, J., AND GINTHER, A. Correlating human traits and cyber security behavior intentions. *Computers & Security 73* (Mar. 2018), 345–358.

[33] HAKEL, M. D. Normative Personality Factors Recovered from Ratings of Personality Descriptors: The Beholder's Eye. *Personnel Psychology 27*, 3 (Sept. 1974), 409–421.

[34] HALEVI, T., LEWIS, J., AND MEMON, N. A Pilot Study of Cyber Security and Privacy Related Behavior and Personality Traits. In *Proceedings of the 22Nd International Conference on World Wide Web* (New York, NY, USA, 2013), WWW '13 Companion, ACM, pp. 737–744.

[35] HALEVI, T., MEMON, N., LEWIS, J., KUMARAGURU, P., ARORA, S., DAGAR, N., ALOUL, F., AND CHEN, J. Cultural and Psychological Factors in Cyber-security. In *Proceedings of the 18th International Conference on Information Integration and Web-based Applications and Services* (New York, NY, USA, 2016), iiWAS '16, ACM, pp. 318–324.

[36] JESKE, D., COVENTRY, L., BRIGGS, P., AND VAN MOORSEL, A. Nudging whom how: Nudging whom how: It proficiency, impulse control and secure behaviour.

[37] JUDGE, T. A., EREZ, A., BONO, J. E., AND THORESEN, C. J. Are measures of self-esteem, neuroticism, locus of control, and generalized self-efficacy indicators of a common core construct? *Journal of personality and social psychology 83*, 3 (2002), 693.

[38] JUDGE, T. A., HIGGINS, C. A., THORESEN, C. J., AND BARRICK, M. R. The big five personality traits, general mental ability, and career success across the life span. *Personnel psychology 52*, 3 (1999), 621–652.

[39] JUDGE, T. A., AND ILIES, R. Relationship of personality to performance motivation: A meta-analytic review. *Journal of Applied Psychology 87*, 4 (2002), 797–807.

[40] KAMP, P.-H. LinkedIn Password Leak: Salt Their Hide. *acmqueue 10*, 6 (2012).

[41] KELLEY, P. G., KOMANDURI, S., MAZUREK, M. L., SHAY, R., VIDAS, T., BAUER, L., CHRISTIN, N., CRANOR, L. F., AND LOPEZ, J. Guess Again (and Again and Again): Measuring Password Strength by Simulating Password-Cracking Algorithms. In *2012 IEEE Symposium on Security and Privacy* (May 2012), pp. 523–537.

[42] KOMANDURI, S., SHAY, R., KELLEY, P. G., MAZUREK, M. L., BAUER, L., CHRISTIN, N., CRANOR, L. F., AND EGELMAN, S. Of passwords and people: measuring the effect of password-composition policies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2011), ACM, pp. 2595–2604.

[43] KOTSIANTIS, S., KANELLOPOULOS, D., AND PINTELAS, P. Data preprocessing for supervised leaning. *International Journal of Computer Science 1*, 2 (2006), 111–117.

[44] KRÄMER, N. C., AND WINTER, S. Impression Management 2.0. *Journal of Media Psychology 20*, 3 (Jan. 2008), 106–116.

[45] KUHN, B. T., AND GARRISON, C. A survey of passwords from 2007 to 2009. In *2009 Information Security Curriculum Development Conference* (2009), ACM, pp. 91–94.

[46] LANDERS, R. N., AND LOUNSBURY, J. W. An investigation of Big Five and narrow personality traits in relation to Internet usage. *Computers in Human Behavior 22*, 2 (Mar. 2006), 283–293.

[47] LEBOWITZ, S. Big 5 personality traits predict who will become a leader - business insider. `http://www.businessinsider.com/big-five-personality-traits-predict-leadership-2016-12`, December 2016. (Accessed on 02/25/2018).

[48] MCCRAE, R. R. Universal Features of Personality Traits From the Observer's Perspective: Data From 50 Cultures. *Journal of Personality and Social Psychology 88*, 3 (2005), 547–561.

[49] MCCRAE, R. R., AND JOHN, O. P. An Introduction to the Five-Factor Model and Its Applications. *Journal of Personality 60*, 2 (June 1992), 175–215.

[50] MORRIS, R., AND THOMPSON, K. Password security: A case history. *Communications of the ACM 22*, 11 (1979), 594–597.

[51] NEUPANE, A., SAXENA, N., KURUVILLA, K., GEORGESCU, M., AND KANA, R. K. Neural signatures of user-centered security: An fmri study of phishing, and malware warnings. In *NDSS* (2014).

[52] NORMAN, W. T. Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *The Journal of Abnormal and Social Psychology 66*, 6 (1963), 574–583.

[53] ONES, D. S., VISWESVARAN, C., AND REISS, A. D. Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology 81*, 6 (1996), 660.

[54] OZER, D. J., AND BENET-MARTINEZ, V. Personality and the prediction of consequential outcomes. *Annu. Rev. Psychol. 57* (2006), 401–421.

[55] PAOLACCI, G., CHANDLER, J., AND IPEIROTIS, P. Running Experiments on Amazon Mechanical Turk. SSRN Scholarly Paper ID 1626226, Social Science Research Network, Rochester, NY, June 2010.

[56] ROCCAS, S., SAGIV, L., SCHWARTZ, S. H., AND KNAFO, A. The big five personality factors and personal values. *Personality and social psychology bulletin 28*, 6 (2002), 789–801.

[57] ROSEN, P. A., AND KLUEMPER, D. H. The Impact of the Big Five Personality Traits on the Acceptance of Social Networking Website.

[58] SASSE, M. A., BROSTOFF, S., AND WEIRICH, D. Transforming the 'Weakest Link' - a Human/Computer Interaction Approach to Usable and Effective Security. *BT Technology Journal 19*, 3 (July 2001), 122–131.

[59] SAUNDERS, F. W. *Katherine and Isabel: Mother's Light, Daughter's Journey*, first edition ed. Nicholas Brealey Publishing, Palo Alto, Calif, Jan. 1995.

[60] SCHAEFER, P. S., WILLIAMS, C. C., GOODIE, A. S., AND CAMPBELL, W. K. Overconfidence and the big five. *Journal of research in Personality 38*, 5 (2004), 473–480.

[61] SCHMIDT, F. L., LE, H., AND ILIES, R. Beyond alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual-differences constructs. *Psychological Methods 8*, 2 (2003), 206.

[62] SCHRETLEN, D. J., VAN DER HULST, E.-J., PEARLSON, G. D., AND GORDON, B. A neuropsychological study of personality: Trait openness in relation to intelligence, fluency, and executive functioning. *Journal of clinical and experimental neuropsychology 32*, 10 (2010), 1068–1073.

[63] SHANNON, C. E. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review 5*, 1 (2001), 3–55.

[64] SOLDZ, S., AND VAILLANT, G. E. The big five personality traits and the life course: A 45-year longitudinal study. *Journal of Research in Personality 33*, 2 (1999), 208–232.

[65] SPSS, I., ET AL. Ibm spss statistics for windows, version 20.0. *New York: IBM Corp* (2011).

[66] TANESKI, V., HERICKO, M., AND BRUMEN, B. Password security #x2014; No change in 35 years? In *2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (May 2014), pp. 1360–1365.

[67] THORPE, J., AL-BADAWI, M., MACRAE, B., AND SALEHI-ABARI, A. The presentation effect on graphical passwords. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (2014), ACM, pp. 2947–2950.

[68] TUPES, E. *Personality traits related to effectiveness of junior and senior Air Force officers*, usaf personnel training research ed. 1957.

[69] TUPES, E. C., AND CHRISTAL, R. E. Recurrent Personality Factors Based on Trait Ratings. *Journal of Personality 60*, 2 (1961), 225–251.

[70] UR, B., BEES, J., SEGRETI, S. M., BAUER, L., CHRISTIN, N., AND CRANOR, L. F. Do users' perceptions of password security match reality? In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (2016), ACM, pp. 3748–3760.

[71] UR, B., KELLEY, P. G., AND KOMANDURI, S. How Does Your Password Measure Up? The Effect of Strength Meters on Password Creation | USENIX.

[72] V, L. Introducing The Psychology of Passwords, Sept. 2016.

[73] VERAS, R., THORPE, J., AND COLLINS, C. Visualizing Semantics in Passwords: The Role of Dates. In *Proceedings of the Ninth International Symposium on Visualization for Cyber Security* (New York, NY, USA, 2012), VizSec '12, ACM, pp. 88–95.

[74] VERDUYN, P., AND BRANS, K. The relationship between extraversion, neuroticism and aspects of trait affect. *Personality and Individual Differences 52*, 6 (2012), 664–669.

[75] VERNON, P. A., MARTIN, R. A., SCHERMER, J. A., AND MACKIE, A. A behavioral genetic investigation of humor styles and their correlations with the Big-5 personality dimensions. *Personality and Individual Differences 44*, 5 (Apr. 2008), 1116–1125.

[76] WEIR, M., AGGARWAL, S., COLLINS, M., AND STERN, H. Testing Metrics for Password Creation Policies by Attacking Large Sets of Revealed Passwords. In *Proceedings of the 17th ACM Conference on Computer and Communications Security* (New York, NY, USA, 2010), CCS '10, ACM, pp. 162–175.

[77] WHEELER, D. L. zxcvbn: Low-Budget Password Strength Estimation | USENIX, Aug. 2016.

[78] WHITTY, M., DOODSON, J., CREESE, S., AND HODGES, D. Individual Differences in Cyber Security Behaviors: An Examination of Who Is Sharing Passwords. *Cyberpsychology, Behavior and Social Networking 18*, 1 (Jan. 2015), 3–7.

[79] WILKES, M. V. Time-sharing computer systems.

[80] YAN, J., BLACKWELL, A., ANDERSON, R., AND GRANT, A. Password Memorability and Security: Empirical Results. *IEEE Security and Privacy 2*, 5 (Sept. 2004), 25–31.

Appendix

A.1 More Exploratory Analysis

The following tables include all of the findings within the exploration portion of the research. All the findings displayed here are not presented in the paper as they either do not pass the Bonferroni correction test, are not hypothesized, or are just not significant enough to add.

A.1.1 Statistically Significant before Bonferroni

The significant correlations here became non-significant after the Bonferroni correction.

Table A.1: Created Email Password Strength

	r	p	N
Neuroticism	-0.019	0.767	252
Openness	-0.031	0.628	252
Conscientiousness	0.004	0.945	252
Agreeableness	-0.134*	0.033	252
Extroversion	-0.043	0.494	252

Correlations between the password created during the email password creation activity for participants who responded "no" to "Password Awareness".

Table A.2: Created Email Password Strength

	r	p	N
Neuroticism	-0.024	0.648	358
Openness	-0.033	0.535	358
Conscientiousness	0.052	0.331	358
Agreeableness	-0.12*	0.023	358
Extroversion	-0.007	0.891	358

Correlations between the password created during the email password creation activity for participants who responded "no" to "Security Training".

Table A.3: Selected Email Password Strength

	r	p	N
Neuroticism	-0.029	0.586	358
Openness	0.014	0.786	358
Conscientiousness	0.081	0.125	358
Agreeableness	0	0.999	358
Extroversion	0.104*	0.05	358

Correlations between the password selected during the email password selection activity for participants who responded "no" to "Security Training".

Table A.4: Created Bank Password Strength

	r	p	N
Neuroticism	-0.081	0.319	154
Openness	-0.018	0.822	154
Conscientiousness	0.035	0.671	154
Agreeableness	-0.029	0.725	154
Extroversion	0.151	0.061	154

Correlations between the password selected during the email password creation activity for participants who responded "no" to "Account Hijacking".

Table A.5: Created Bank Password Strength

	r	p	N
Neuroticism	0.072	0.175	356
Openness	-0.057	0.284	356
Conscientiousness	-0.037	0.486	356
Agreeableness	0.003	0.957	356
Extroversion	-0.079	0.136	356

Correlations between the password selected during the email password creation activity for participants who responded "yes" to "Account Hijacking".

Table A.6: "Very Weak" passwords in Password Ranking

	r	p	N
Neuroticism	-0.031	0.243	510
Openness	-0.083*	0.03	510
Conscientiousness	-0.049	0.132	510
Agreeableness	-0.033	0.229	510
Extroversion	0.024	0.292	510

Correlations between the true Zxcvbn estimated password strength and the implied strength of passwords placed in the first bucket within the Password Ranking activity. Negative correlations indicate a better performance (i.e., less of a distance between the Zxcvbn estimated strength and implied strength).

Table A.7: “Ranked Score Distance” in Password Ranking

	r	p	N
Neuroticism	-0.041	0.288	185
Openness	-0.157*	0.017	185
Conscientiousness	-0.088	0.117	185
Agreeableness	-0.005	0.474	185
Extroversion	0.112	0.065	185

Correlations between the “Ranked Score Distance” metric during the password ranking activity for participants who responded “yes” to ‘Password Awareness” and “Account Hijacking”. Negative correlations indicate a better performance as the lower the “Ranked Score Distance”, the better a participant performed on the activity.

A.1.2 Password Ranking Activity

The correlations in this section depict performance amongst individual buckets within the Password Ranking activity (i.e., how each personality trait performed on the “Very Weak”, “Weak”, “Normal”, “Strong”, and “Very Strong” buckets).

Table A.8: “Very Weak” Bucket

	r	p	N
Openness	-0.023	0.601	510
Neuroticism	-0.029	0.519	510
Conscientiousness	-0.104*	0.019	510
Agreeableness	0.054	0.226	510
Extroversion	-0.004	0.936	510

Correlation between password placed in “Very Weak” bucket in the password ranking activity for all participants

Table A.9: “Weak” Bucket

	r	p	N
Openness	-0.035	0.427	510
Neuroticism	-0.031	0.489	510
Conscientiousness	0.092*	0.037	510
Agreeableness	-0.021	0.634	510
Extroversion	-0.030	0.495	510

Correlation between password placed in “Weak” bucket in the password ranking activity for all participants

Table A.10: “Normal” Bucket

	r	p	N
Openness	0.031	0.489	510
Neuroticism	-0.016	0.724	510
Conscientiousness	-0.017	0.695	510
Agreeableness	-0.027	0.541	510
Extroversion	0.007	0.873	510

Correlation between password placed in “Normal” bucket in the password ranking activity for all participants

Table A.11: “Strong” Bucket

	r	p	N
Openness	0.039	0.381	510
Neuroticism	-0.016	0.711	510
Conscientiousness	-0.002	0.972	510
Agreeableness	0.013	0.763	510
Extroversion	0.026	0.557	510

Correlation between password placed in “Strong” bucket in the password ranking activity for all participants

Table A.12: “Very Strong” Bucket

	r	p	N
Openness	-0.002	0.961	510
Neuroticism	0.035	0.423	510
Conscientiousness	0.039	0.382	510
Agreeableness	-0.023	0.609	510
Extroversion	0.010	0.828	510

Correlation between password placed in “Very Strong” bucket in the password ranking activity for all participants

A.1.3 Passwords

The following tables include all the passwords used within the experiment.

Table A.13: Passwords used of Zxcvbn-strength 0

november	angelica	spiderman	cristian
september	courtney	christopher	internet
december	precious	slipknot	garfield
kimberly	veronica	rockstar	qwertyuiop
pictures	sebastian	christine	jordan23
midnight	nicholas		

Table A.14: Passwords used of Zxcvbn-strength 1

sexylady	catarina	playboy69	honeypie
drpepper	superman1	mongoose	felicidade
simpleplan	happiness	losangeles	fantasia
lovehurts	godisgood	princess01	tigger12
peterpan	beautiful1	ladybird	ihateyou1
love4ever	kittykat	heavenly	timberland
thuglife	madison1	fernandes	quiksilver
aquarius	kayleigh	isabella1	babycakes
soccer17	candy123	softball12	cardinals
slamdunk	squirrel	revolution	mississippi

Table A.15: Passwords used of Zxcvbn-strength 2

ilovedave	lilmomma	babygirl7	beachbabe
iloveandy	bootylicious	scotland1	punkista
spoiled1	highschoolmusical	loverboy1	pinkstar
fashionista	cowgirl1	maryjane1	

Table A.16: Passwords used of Zxcvbn-strength 3

2fast2furious	promo2006	johnterry26	im2cute4u
lilwayne2	myhusband1	jiggaman1	ipodnano1
boomboom1	kamikaze1		

Table A.17: Passwords used of Zxcvbn-strength 4

chenleixu201	dlanddr84041q	shrdlu2010az
valettab380	xzs5fviopwg6	

A.1.4 Demographics

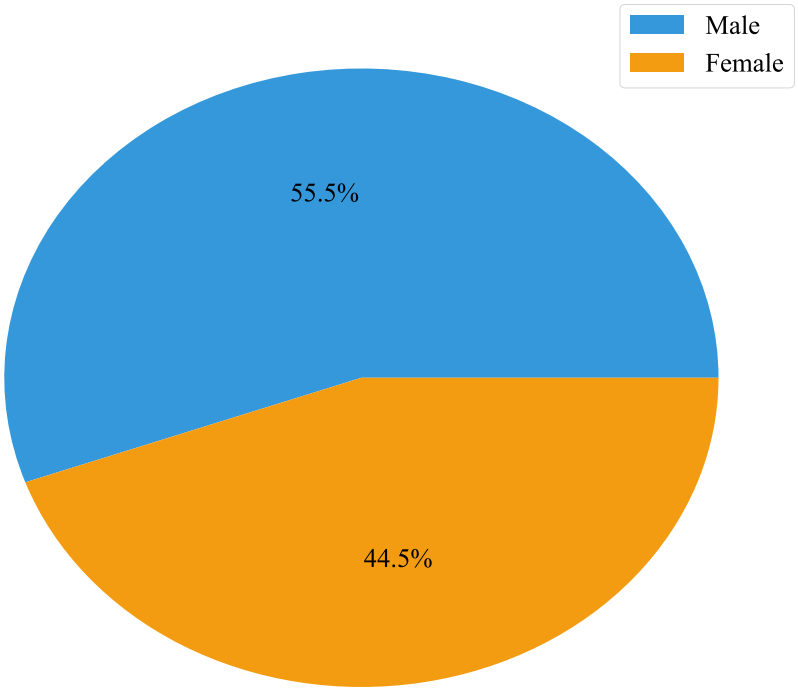

Figure A.1: Gender Count

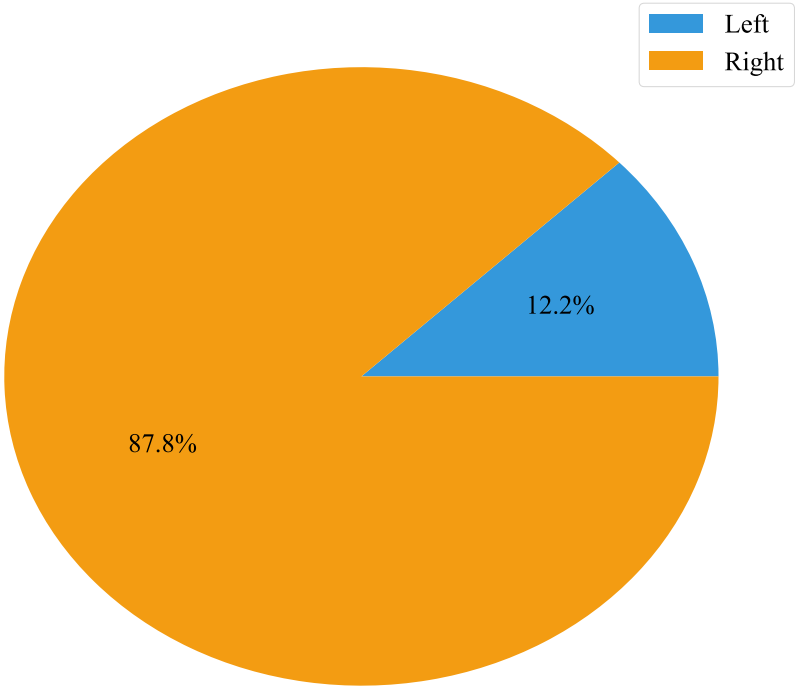

Figure A.2: Handedness Count

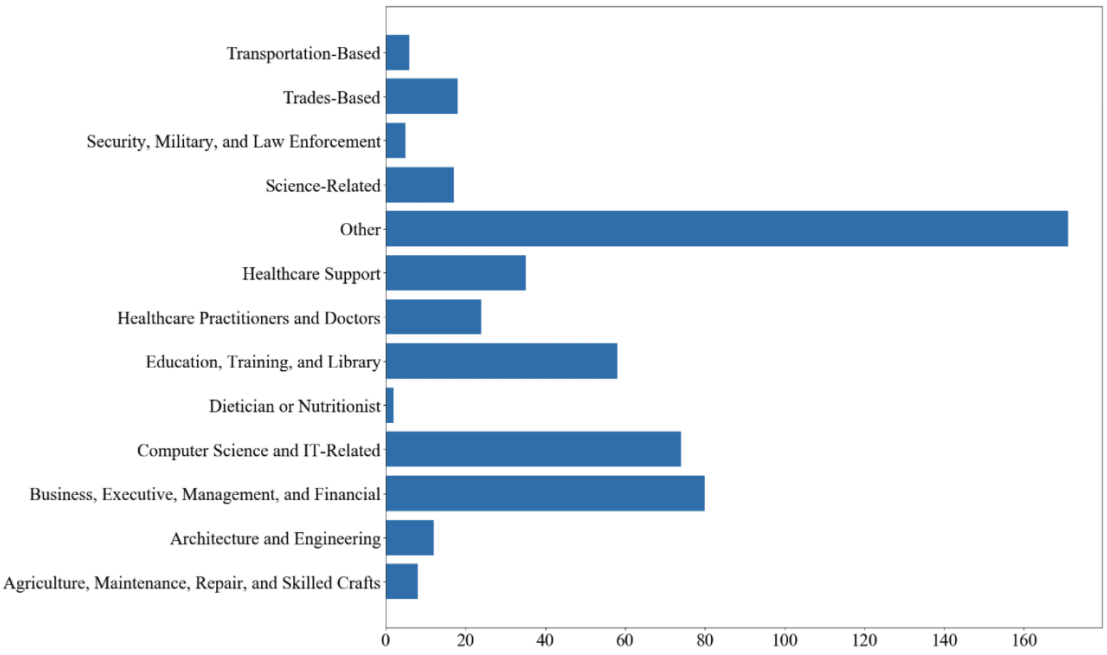

Figure A.3: Occupation Count

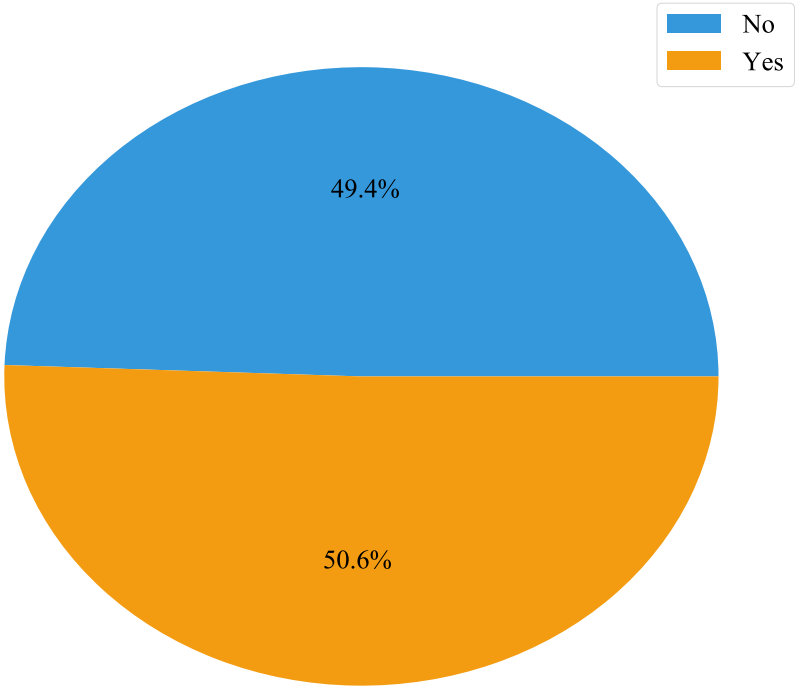

Figure A.4: “Password Awareness” Count

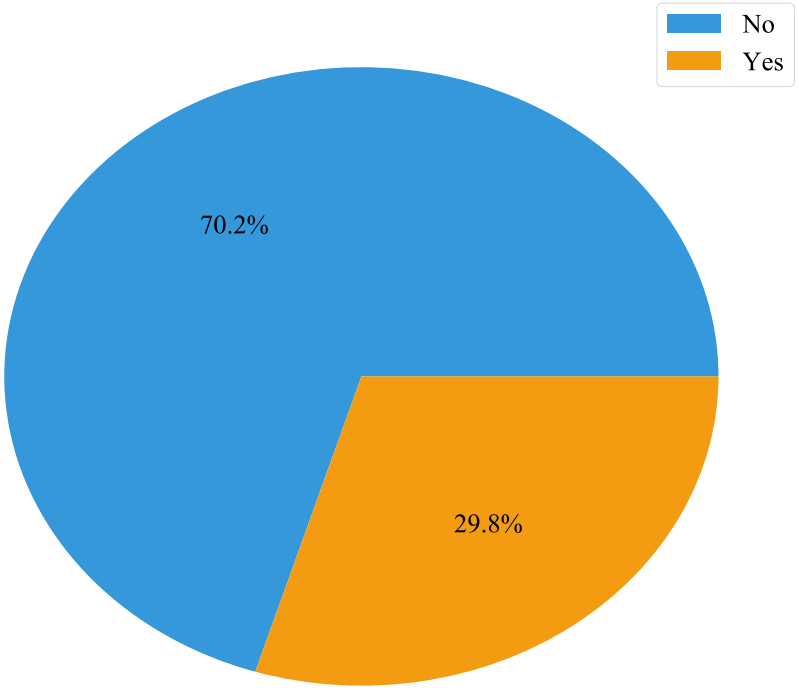

Figure A.5: “Security Training” Count

A.1.5 Mini-IPIP Personality Test

The following questions were used as a part of the administered personality test for participants.

1. “I am the life of the party”,
2. “I sympathize with others’ feelings”,
3. “I get chores done right away”,
4. “I have frequent mood swings”,

5. “I have a vivid imagination”,

6. “I don’t talk a lot”,

7. “I am not interested in other people’s problems”,

8. “I often forget to put things back in their proper place”,

9. “I am relaxed most of the time”,

10. “I am not interested in abstract ideas”,

11. “I talk to a lot of different people at parties”,

12. “I feel others’ emotions”,

13. “I like order”,

14. “I get upset easily”,

15. “I have difficulty understanding abstract ideas”,

16. “I keep in the background”,

17. “I am not really interested in others”,

18. “I make a mess of things”,

19. “I seldom feel blue”,

20. “I do not have a good imagination”.