A Study of Correlations Between Trait Affect and Phishing Susceptibility

Samantha Emily Smith

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

Master of Science

University of Washington

2020

Committee:

Marc Dupuis

Johnny Lin

David LeBlanc

Program Authorized to Offer Degree:
Computing and Software Systems

University of Washington

Abstract

A Study of Correlations Between Trait Affect and Phishing Susceptibility

Samantha Emily Smith

Chair of the Supervisory Committee:
Dr. Marc Dupuis
Computing and Software Systems

Although phishing emails have been in use for decades, these social engineering attacks are still prevalent because they keep working; in fact, they are a leading cause of data breaches. In this research, I attempt to discern how an individual's trait affect levels are related to their susceptibility to clicking on links in phishing emails, with particular attention on how this relationship may vary based on the type of phishing email employed. Trait Affect is a term from psychology that references a subset of one's disposition and tendency towards certain moods and emotions. Trait Affect is further broken down into positive affect and negative affect, which are largely independent. Positive Affect reflects one's tendency to act, while Negative Affect reflects one's tendency towards experiencing negative emotions. Trait Affect has been shown to influence user's behaviors and risk perception. Additionally, it is generally stable over an individual's lifetime, making it a useful metric with which to model behavior. Being able to model an individual's behavior in response to phishing is important to lowering the rates of phishing. While the creation of such models is outside the scope of this paper, the relationships examined will prove useful in future attempts to model such behaviors. To obtain data as close to a real-world scenario as possible, phishing susceptibility was measured on a click-through basis of emails sent to participant's personal emails. This process caused some difficulty in managing to make emails that were both compelling and capable of passing automated email filtering. The process was further complicated by legal

concerns surrounding the real-world approach to phishing. It is important to note, however, that no user data was taken – all measurements were based around a user clicking on a link rather than entering any information.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

AFFECT: A term frequently used in psychology to represent the combination of mood, emotion, and trait affect. Affect influences decision making processes. It is generally comprised of the two higher order dimensions of Positive Affect and Negative Affect. (Russell, 2003)

TRAIT AFFECT: A form of affect representing a stable, lifelong affinity.

POSITIVE AFFECT: One of two high order dimensions of trait affect. Positive affect is the extent to which an individual has a tendency towards action. High levels associated with descriptors such as "preppy" or "excited".

NEGATIVE AFFECT: One of two high order dimensions of trait affect. Negative affect is the extent to which an individual has a tendency towards experiencing negative emotions. High levels are associated with descriptors such as "nervous" or "jittery".

PHISHING: Fraudulent attempts to gain information about a person, such as financial information or passwords, through social engineering. typically, this involves impersonating a legitimate business. In this study, the scope is limited to phishing emails.

ACKNOWLEDGMENTS

The author wishes to express sincere appreciation to Marc Dupuis for his assistance as my advisor and to the University of Washington.

DEDICATION

This thesis is dedicated to my wonderful parents, Flint Smith and Emily Smith. Without your unfailing support and faith in me, I wouldn't be who I am today.

Chapter 1

INTRODUCTION

1.1 Background

There is no shortage of malicious actors in the current environment. Every year, companies and home users suffer from attackers stealing data or personal information for profit. Some attacks are based in technical oversights, exploiting vulnerabilities in software to achieve their ends. These can be vicious, but are also often quickly patched, and following basic security practices can very effectively limit the potential for disaster. However, other attacks do not rely on technical vulnerabilities. Attackers can use social engineering to trick users into voluntarily giving up their data. Social Engineering frequently involves emotional appeals, assumed identities, or simply acting as if one has a right to the information and therefore receiving it. A common form of social engineering is phishing, which uses email to connect with people and gain information. In 2018, 62% of businesses experienced at least one phishing attack (Milkovich, 2019). While it may sound less dangerous than technical vulnerabilities, $17,700 is lost every minute due to phishing attacks (Fruhlinger, 2020). The Anti-Phishing Working Group tracks phishing reports received by its members, and has data to support that over 90,000 unique phishing campaigns are launched each month (APWG, 2017). The prevalence of these phishing campaigns is rooted in the fact that people still fall victim to these attacks. Thus, determining how and why individuals continue to click on links or otherwise engage with phishing emails helps to effectively train users away from such behaviors. One lens through which user behavior can be measured is affect.

Affect is a psychological umbrella term used to describe a person's mood, emotions, and disposition. Trait affect, sometimes called dispositional affect, is a form of affect that is generally stable across one's life. It is measured along two orthogonal axes: positive and

negative affect. Positive affect is linked to one's propensity towards action, with high levels of positive affect using descriptors such as"enthusiastic" or "peppy" and low levels associated with "drowsy" or "dull". Negative Affect is a measure of one's tendency towards negative emotion. High levels of negative affect are described as "jittery" or "scornful", while low levels are described as "at rest" or "placid" (Watson and Tellegen, 1985). Several studies have researched the role of trait affect in various IT behaviors, such as one's cognitive absorption (Agarwal and Karahanna, 2000), the perceived usefulness of a system (Davis, 1989), and many others.

Trait affect has been shown to affect an individual's risk perception and behaviors both when alone (Johnson and Tversky, 1983) and in groups of people (Tarditi et al., 2020). Risk perception is influenced differently depending on the type of risk. Large, abstract risks, such as disease or natural disasters are strongly swayed by optimistic bias. Those with high positive affect tend to underestimate risks, while those with high negative affect moderates the optimistic bias of participants, particularly in regards to threat severity (Borkenau and Mauer, 2006; Helweg-Larsen and Shepperd, 2001; Taylor and Shepperd, 1998). The other way in which affect can modify risk perception pertains to smaller scale risks, such as the choice to gamble. With small, immediate risk, individuals with high levels of positive affect tend to gauge risk more accurately, with high engagement on small risk and low engagement with higher risk options (Isen et al., 1988).

1.2 Propositions

Hypothesis 1: An individual's susceptibility to phishing emails will vary based on the topic of phishing email.

Hypothesis 2: Trait affect will effect how an individual responds to phishing emails.

Hypothesis 2A: High levels of positive affect will be related to higher levels of susceptibility to phishing emails.

Hypothesis 2B: High levels of negative affect will be related to lower levels susceptibility to phishing emails.

1.3 Contributions

This thesis extends the current understanding of phishing susceptibility by looking at a previously unexplored factor, trait affect. This is important as there is not yet a definitive understanding about what makes users susceptible to phishing, and pinpointing why a person falls for such social engineering attacks is vital in preventing them.

Similarly, it extends existing research on the role of trait affect on cybersecurity practices exhibited by individuals. Increasing the body of work focusing on trait affect will help enhance the basis for future studies exploring the intersection between trait affect and security. Additionally, relationships found in this paper will be useful for future attempts at modeling the causes of phishing susceptibility.

It also includes a novel way to examine phishing susceptibility in an environment that more closely resembles the real world, by having participants unaware of the phishing aspect of the study. Evaluating the effectiveness of this approach can help future studies looking to examine phishing response in a setting outside a known phishing study.

Finally, this paper explores phishing susceptibility not as a black and white division between "susceptible to phishing" and "resistant to phishing", but focuses on susceptibility to types of phishing by type as well as in aggregate.

1.4 Methodology

In this study, participants were surveyed as part of an unrelated study. Participants completed a series of surveys for this unrelated cover study which included measures on trait affect and various demographics. After completing the cover study and receiving compensation, participants were sent crafted phishing emails. In order to preserve as much realism as possible, participants were not told to expect any phishing attacks nor to monitor their emails more closely. Participants were paid for their participation in the study.

Chapter 2
LITERATURE REVIEW

2.1 Introduction

This chapter first covers trait affect and its role in related fields: risk perception, impulse, job searching, and general security behaviors. Next I discuss current knowledge surrounding phishing email susceptibility.

2.2 Trait Affect

2.2.1 What is Trait Affect?

The term affect has been used in a variety of contexts, and is used to express the combination of emotion, mood, and disposition (Russell, 2003). Affect has been shown to influence a variety of decision making processes, including risk perception (Johnson and Tversky, 1983), impulse shopping (Thompson and Prendergast, 2015; Habib and Qayyum, 2017; Darrat et al., 2016), job searches (Turban et al., 2013), and security behaviors (Dupuis, 2014).

In this context, affect is broken into two higher order dimensions, positive affect and negative affect. Positive affect is related to an individual's proclivity towards energetic action. High positive affect is associated with descriptors such as active or excited, while low positive affect is described as sluggish or dull. negative affect is related to unpleasant arousal and unhappiness. High negative affect is described as being distressed, fearful, or hostile, while low negative affect is characterized by being calm or relaxed. The two dimensions are independent (Watson and Tellegen, 1985; Diener and Emmons, 1984). Early examinations of affect work with a framework based on positive and negative affect as two sides of a bipolar continuum (Johnson and Tversky, 1983). However, working with a framework of two distinct orthogonal valences representing positive and negative affect offers increased

discriminant and convergent validity (Watson and Clark, 1997).

The Positive and Negative Affect Schedule (PANAS) or its expanded form (PANAS-X), provide low inter-correlations between Positive and Negative Affect, and good reliability. These are self-reported scales by which to measure trait affect (Watson et al., 1988; Watson and Clark, 1999).

2.2.2 Consistency

Affect is generally split between State Affect and Trait Affect. State Affect is a form of affect that reflects a current state of being, such as one's emotions or mood (Grös et al., 2007). Trait Affect, on the other hand, is a reflection of a more stable and lifelong tendency of an individual. Trait Affect is generally persistent across several years (Watson and Clark, 1999; Watson and Walker, 1996; Dupuis, 2014). Several studies have demonstrated the stability of trait affect over 6 months or less, finding high levels of stability (Diener and Larsen, 1984; Watson et al., 1988; Izard et al., 1993). Watson ans Walker (1996) evaluated the stability of trait affect on the self-reported PANAS scale with a retest interval of 56 to 99 months (4.67 to 8.25 years), finding Positive and Negative Affect to be relatively stable, with correlations in the range or .40 to .45. The study targeted undergraduates who graduated throughout the retest interval. Notably, Negative Affect had a lower long-term stability than Positive Affect.

One assumed cause of the difference in stability is the decline in stress during the years after graduation. This assumption was supported by studies showing higher levels of neuroticism in college students than general adult samples (Costa and Mccrae, 1992; Watson and Clark, 1999). In a series of studies testing the stability of Negative Affect over various time periods, it was found that stability over time periods less than six moths ranges from 0.80 to 0.86, stability over six months to 13 years ranged from .54 to .65, and that even over 30 years there remained a reliability of 0.40 (Watson and Clark, 1984). However, the number of studies for intervals past two years is far smaller than the shorter intervals, giving less weight to the recorded stability.

Cross-situational consistency varies across studies. Early research found high consistency (Diener and Larsen, 1984), and later research determined that situational changes greatly alters reported levels of affect (Oishi et al., 2004). The key difference between these studies is the form of cross-situational stability studied. The original study focused on inter-individual cross-situational consistency. Individuals who reported higher than average levels of positive/negative affect in one situation tended to report similarly higher levels of positive/negative affect in other situations (Diener and Larsen, 1984). The later study verified the inter-individual aspect. However, when looking at intra-individual cross-situational affect, there was large variance (Oishi et al., 2004).

Cultural differences have been shown to have little bearing on affect overall, but can alter the situational differences for levels of affect on an intra-individual perspective. (Oishi et al., 2004)

2.2.3 Mapping Trait Affect to Related Psychological Traits

As trait affect is a measure of one's tendency towards certain states of being, many researchers have looked into where trait affect and personality traits intersect. There are strong links between trait affect and personality dimensions. Generally, personality traits are separated into 5 aspects: Extraversion, Neuroticism, Openness, Agreeableness, and Conscientiousness. Of these five traits, Neuroticism and Extraversion have the closest ties to Affect. Extraversion is linked to positive affect, and Neuroticism is linked to negative Affect (Izard et al., 1993; Watson and Clark, 1992). Particularly, Extraversion is related to the duration of positive emotions, and neuroticism is related to the frequency of negative emotions (Verduyn and Brans, 2012).

Although there are strong correlations between personality and trait affect, as well as between trait affect and state affect, there are only mild correlations between personality and state affect (Meyer and Shack, 1989).

2.2.4 Risk Perception and Decision Making

Negative mood increases perceived risk, while positive mood decreases perceived risks. (Johnson and Tversky, 1983). Within the scope of Johnson's study, affect was treated as a bipolar scale between positive and negative states, rather than the later model of using positive and negative affect as different dimensions of affect. Additionally, the affect considered here is state affect rather than trait affect, as the positive and negative moods were induced. However, as trait affect and state affect are correlated, this could still suggest an influence of trait affect (Watson et al., 1988).

Isen investigated risk perception in a scenario involving gambling with varied win probabilities. Across several studies, participants with induced positive affect would consistently choose low risk scenarios over high risk ones (Isen et al., 1988). This finding is an example of hot trait affect alters risk perception through mood congruency. Mood congruency is the theory that individuals in a positive state will act in a way to preserve that state.

Isen's findings for positive affect are partially contradictory with the findings in Johnson's study (Johnson and Tversky, 1983). This is likely due to a difference in the meaning of "risk". In Johnson's study, risk pertains to the likelihood and effects of large scale, generally hypothetical risks, such as natural disasters or disease. The risks of Isen's group, on the other hand, are much smaller in scale. The risks are a loss of some money, and it also comes with the potential for gains.

The classification of phishing as a risk in comparison to Isen and Johnson's risk studies is somewhere between the two. On the one hand, loss from phishing emails is hypothetically large, and getting phished feels like a rather hypothetical situation, similar to Johnson's risks. On the other hand, many phishing emails offer rewards for user participation, placing them closer to Isen's line of risks.

Several studies concerning trait affect and risk perception have found a relationship between trait affect and optimistic bias. Optimistic Bias is the tendency for an individual to perceive themselves at less risk for a myriad of unpleasant events than their actual likelihood

of such an event. These findings are similar to those of Johnson's study. Additionally, the risks in theses studies tend to be larger scale risks, such as flooding, earthquakes, or illness.

Optimistic Bias is not altered by the participant's state affect, but is correlated with trait affect (Borkenau and Mauer, 2006). Negative affect is primarily a moderator to personal optimistic bias, though it does also weakly lower estimates of risk to others (Helweg-Larsen and Shepperd, 2001). Helweg-Larsen's research drew in part upon the prior discovery that threat severity only seemed to change personal risk estimates, and that controlling for differences in Negative Affect among the participants removed the correlation between severity and risk estimates (Taylor and Shepperd, 1998).

Given the influence of trait affect on risk perception and related behaviors, it is likely that it will influence ones' susceptibility to phishing.

2.2.5 Security and Computer Usage behaviors

Due to the stable, lifelong nature of trait affect and its use in predicting risk perception, several studies have explored the links between trait affect and various security behaviors.

Previous works have shown that trait affect influences users' security behaviors on personal devices. Participants with high levels of positive affect were associated with lower perceived threat vulnerability in regards to security risks that would affect a computer's performance (Dupuis, 2014).

Ormond investigates the link between trait and state affect and security policy compliance (Ormond et al., 2019). The study focuses on both negative affective absorption, a combination of negative trait affect and the disposition to become deeply involved, and negative affective flow, a state of absorption at which nothing else matters. The paper supports the stated hypothesis that negative affective absorption is positively related to negative affective flow, which in turn is negatively related to information security policy compliance. This suggests that negative trait affect, from which negative affective absorption is derived, is indirectly negatively related to security policy compliance.

2.2.6 Job Searching

Trait affect has been shown to affect an individual's habits when job searching.

Côté et al. investigate the correlations between trait affect and job search clarity and intensity. Job search clarity is the extent of an individual's clear, well-defined search objectives; within the bounds of Cote's work the definition was expanded to include the way in which one searched for a job (Côté et al., 2006). In the context of receiving unsolicited phishing emails, job search clarity would likely be a moderator on one's likeliness to click on a link – having a well-defined objective lessens the temptation offered by other unsolicited opportunities. Job search intensity, on the other hand, would likely increase the chance of an individual interacting with a phishing email. The study found that high levels of positive affect correlated positively with higher levels of job clarity and with higher levels of job search intensity; there were no significant correlations found with negative affect (Côté et al., 2006).

These findings are corroborated by a later study which researched the influence of affect on procrastination and intensity. Positive affect is negatively correlated to procrastination, which in turn is negatively correlated to job search intensity. Again, negative affect was not significantly correlated to either procrastination or intensity (Turban et al., 2013).

Given that positive affect positively correlates with factors assumed to decrease and increase phishing susceptibility, the way in which individuals response will be determined by which factor is more dominant. Job clarity and search intensity primarily affect individuals currently actively looking for a job. However, there is also the potential for users to be spurred into action regardless of theirjob searching status. As such, I predict that overall, response to a phishing job offer would be increased by high positive affect, which implies a tendency towards action rather than inaction.

2.3 Phishing

2.3.1 What is Phishing, and why does it matter?

Phishing is a form of social engineering that makes use of widely distributed messages, often emails, to request personal information from victims. Usually, these messages will make use of names belonging to legitimate entities such as government branches or well-known businesses to disguise their identity. Although phishing emails have been in use for decades, these social engineering attacks are still prevalent, because they keep working. According to Verizon's 2019 Data Breach Investigation Report, phishing remains a leading cause of data breaches (Verizon, 2019).

2.3.2 Susceptibility by Individual

When measuring phishing susceptibility, it is important to pay attention to false positives and false negatives. If an individual considers all emails to be phishing, they will not become victims of a phishing campaign. However, by ignoring all legitimate emails, they will effectively not have an email account. Measures of phishing susceptibility can be tainted in a lab environment due to a form of the Hawthorne effect; as participants are aware that they are being given phishing emails, they are likely more attentive to the possibility of malicious content(Pattinson et al., 2012).

Confidence in one's ability to detect phishing emails is positively correlated with phishing email detection and the accurate assessment of degree of maliciousness. Neither computer knowledge nor awareness of phishing have been shown to impact one's ability to successfully identify a phishing email (Kleitman, 2018). This will be useful within the scope of this study, which focuses on trait affect rather than specific background.

There is a positive correlation between neuroticism and being phished, and a positive correlation between pessimism and higher estimated risk (Halevi et al., 2013). As neuroticism is strongly correlated with negative affect, it is likely that negative affect will be correlated with phishing susceptibility. I earlier hypothesized that, based on the role of negative affect

on optimistic bias, high levels of negative affect will correlate negatively with susceptibility to phishing emails. The connection with pessimism lends support to this hypothesis, though the correlation with neuroticism contradicts it (Hypothesis 2B).

2.3.3 Susceptibility by Message Components

Phishing emails will often contain visual cues that can betray their nature as phishing. These cues include impersonal greetings, grammatical errors, hyperlink mismatches, time limits, or falsified sender fields (Downs et al., 2006). However, many visual cues come with misconceptions, such as the HTTPS lock icon that is shown on a "secure" website. Though many know the general advice that the lock means a website is secure, fewer are aware that anyone can obtain an SSL certificate for their site, thus displaying a lock icon. Attention diverted to visceral triggers decreases a user's cognitive effort dedicated to processing the email, increasing phishing susceptibility (Wang et al., 2012).

Messages with content considered to have a strong and succinct narrative are less susceptible to perception changes caused by minor aesthetic adjustments than that of messages with little meaningful content. The study suggested that some future phishing attempts might include messages with minimal narratives but strong design, less immediate action requirements, or use high profile news to produce messages with greater meaning (Tsow and Jakobsson, 2007). There is little difference in phishing susceptibility between threat based and reward based messages (Harrison et al., 2016).

In a role-play scenario, research indicates that genuine emails are sorted correctly more frequently than phishing emails (Pattinson et al., 2012). This is contradictory to the findings of other research, which showed a similar rate of false positives and false negatives (Kleitman, 2018). A different study asked a wider variety of questions from users as well as differentiating how questions were asked to different groups. This approach found that participants weighted false positives and false negatives as equally costly. Additionally, the study found that the examined behaviors depended on the way a participant was asked to classify the email. When asked if an email was phishing, participants tended to minimize false alarms, but when given

a behavioral question would instead minimize false negatives (Canfield et al., 2016).

2.3.4 Phishing Research Methods

Previous studies have used a variety of experiments to test user susceptibility. These experiments have varied between controlled lab conditions, wherein participants were fully aware of the nature of the study, to unaware participants receiving phishing emails and being notified and surveyed after the fact. The first type of study frequently suffers from the Hawthorne effect, as being aware that a study is based on phishing alters the way in which participants would otherwise behave.

Measuring Phishing with Lab Conditions

Tsow and Jakobsson (2007) asked users to evaluate a series of screenshots of emails on a five point scale from "Certainly Phishing" to "Certainly not phishing". The study used phishing screenshots where the changes to links were visible, rather than requiring participants to mouse over text to check a link's contents. The study ensured there was a roughly even split between legitimate and modified emails to prevent trivial rating. By using this method, they were able to craft phishing email in pairs, where the legitimate and phishing emails only differed by specific changes, allowing the study to evaluate how each type of change affected participants' perception. Additionally, the study featured similar tests related to detecting phishing websites rather than emails (Tsow and Jakobsson, 2007).

A similar study a decade later investigated phishing susceptibility by offering survey participants a series of 40 emails, and asking them to determine whether they were legitimate or phishing, and the "degree of maliciousness" (Kleitman, 2018). This method differs from the previous study in two major respects. The first is that the emails were not designed in pairs that would assess how the inclusion of specific features would sway the risk perception, instead using a larger sample size and drawing similarities between the samples. The second was the inclusion of asking participants the suggested action to take for each email, with the options to "keep it, trash it or seek further information".

Canfield et al. (2016) took the previously used research method and added a mechanism to ensure the participants remained engaged throughout the study. Similar to previous studies, participants were shown a series of emails and asked to perform analysis across several dimensions. The unique contribution was that the study also included attention checks. Participants were posed questions at the beginning about the scenario, and two of the emails directly told which answer a participant should select. The majority of participants answered incorrectly on the attention check informing participants to mark the email as legitimate, and thus the attention check was removed from the study's analysis. The remaining three checks were then used to create a binary variable called attention. The study also tracked responses that were not self consistent between what to do and whether the message was phishing, as well as participants who spent less than 10 seconds on multiple emails (Canfield et al., 2016). These attention checks add an additional element to similar studies by ensuring that participants are focused on the task throughout the procedure and do not begin selecting answers at random. This reduces noise within the data, and limits the count false positives and negatives.

Other studies have attempted to gain realism by cloning real world phishing emails for participants to evaluate. Wang uses a web-based study using images of phishing emails that had been previously carried out against the university. Participants were asked about the likelihood they would interact with the email, without any mentions regarding phishing. Unlike previous studies, only a single email was being evaluated, rather than a series of stimuli (Wang et al., 2012).

Other than simply analyzing emails, some research studies have made use of role-play scenarios to keep participant attention and create realism in their study. Although these studies do shift the basic premise of analyzing a series of emails by adding context, they do not fundamentally change what a participant is being asked. An individual is still presented with a series of emails or images and forced to analyze them on a set of criteria. One of the key observed differences is that the number of participants are typically reduced, as they often include a think aloud protocol.

A version of this type of study was performed with fewer participants (17), but expanded upon the study by asking participants to verbally share their thought process as they went. Their screen actions and voice were recorded. As a follow up, participants were asked about their confidence in their answers, which features lent credibility, and which features were suspicious. One of the participant statements highlights a flaw when asking participants to judge a series of emails, stating "Probably any of these emails I got I would have just deleted ... I wouldn't read anything that looks not important to me." (Jakobsson et al. 2007, pp. 3).

Downs et.al (2006) used a role-play scenario in which users were given false credentials and asked to react to a series of eight emails as if they had received them. Responses were recorded by video cameras over the shoulder of the participants. Three of the emails were "real" emails, while the other 5 were either phishing or spam. During the role-play, interviewers would ask questions regarding the participant's choices and reasoning over how to react. When selecting participants, the study skewed towards security naïve members, only accepting those who had never altered the security preferences on their computers. While the study did not explicitly mention phishing to the participants, no attempts were made to disguise the nature of the study (Downs et al., 2006).

Pattinson et al. (2012) focused on how informing users of the purpose of a study may compromise the results. A group of 117 participants performed a phishing role-play, with a control group that had been informed that this was a phishing exercise, and a test group that had not been informed of the study's purpose. Both groups were asked to assess 50 email images, and instructed to advise the recipient on how to respond to each of the emails. The study found that participants who were warned that they were participating in a phishing study performed significantly better than the uninformed group for phishing emails, and similarly for the genuine emails (Pattinson et al., 2012). This study shows a clear instance of the Hawthorne effect, where knowledge of the study's purpose affected the validity of the study's results. Though not explicitly proven in the study, it lends credence to the idea that assessing phishing from within the context of analyzing a series of emails provides results tainted by the participants' knowledge. Being asked explicitly to analyze emails informs

participants that whatever it is they are looking for, something about the emails is being tested, keeping them alert where they may not otherwise be.

Simulated Phishing and Real-World Data

A few attempts have been made at acquiring phishing responses from participants outside the setup of a survey of laboratory trial, emailing participants directly. Descriptions of these methods are added chronologically.

One of the earlier attempts at such a study launched spear phishing attacks on university students after crawling social media to discover their acquaintances. The study sent emails with the content "Hey, check this out!" followed by a link and the sender's first name. In a control group, the emails were sent out from a fictitious email with the university domain. Targeted members were sent the email from one of their acquaintances, found in the social media crawl. 16% of the 94 participants in the control group fell for the attack, compared to 72% of the of 487 participants in the targeted group. In a debriefing after the study, 1.7% of the participants complained about the study. Participants expressed anger for having been included, despite no sensitive information being retained or stored. Additionally, several participants with email accounts that were spoofed were under the impression that the researchers had hacked their email accounts (Jagatic et al., 2007). One of the factors about this study that likely caused the outrage is that it appears that participants did not sign up for the study, beyond enrolling in the school, nor were they compensated for their participation. This differs from current study, as participants actively signed up to participate in a research study, and were compensated for their efforts.

The next study did not send out fake phishing email to participants, but instead leveraged a series of two real-world phishing attacks that occurred within close succession. The researchers contacted undergraduates at their university and offered a survey concerning email use. Then participants were presented with the two phishing emails, with the order randomized for each student. They were asked if they remembered receiving the emails and how they would respond to them (Vishwanath et al., 2011).

Halevi et al. (2013) conducted a study which did include sending out phishing emails, focusing on the intersection between Facebook activity, personality, and phishing susceptibility. Participants were sent an email promising Apple products, from a Yahoo address claiming to be "CSAW". The email consisted of text only, with several spelling errors, and a hyperlink with a URL different to that shown. Those who clicked the link in the email then proceeded to log in were considered phished. 17% of the 100 participants were phished (Halevi et al., 2013). The sampling in this study was vastly uneven, with 83 male participants and only 17 females.

Harrison et al. (2016) separated 194 participants into two groups, and sent each group a phishing email. One group received an email offering a refund after an overcharge and the other group received an email warning of an account error. Emails were sent to the students from a GMail address with a domain belonging to the university ("XXITinfo@XX.edu"). After sending the emails, participants were asked why they did or did not interact with the phishing message. 47% of participants clicked on the included link within five hours of receiving the email (Harrison et al., 2016).

2.4 Conclusion

Within this section, I discussed trait affect, and its role within various human behaviors, as well as phishing research focused on both the discoveries and methods.

Trait affect is broken down into two independent higher order dimensions: positive and negative affect. These are further broken down into lower order dimensions, which are used when assessing an individual's levels of affect. Within the context of modifications on human behavior, research has shown that trait affect alters risk perception both through optimistic bias and through mood preservation. Additionally, positive and negative affect play a large role in the separation between impulsive and compulsive buying behavior, and in the way an individual approaches a job search. Trait affect's influence on risk perception is of interest for its potential to impact generalized phishing susceptibility, while the influence on other behaviors are of interest for their potential to change phishing susceptibility relative to

different types of emails.

Previous research into phishing has shown that several visual cues, such as misspelling, mismatches between the sender and the email contents, or mismatches in the displayed URL and actual link, are flags for identifying illegitimate email. Additionally, these visual cues lose effectiveness as the narrative in the message grows stronger. For emails without much actual substance, the visual cues are often a determining factor in how users react. However, emails with compelling messages inherently seem more real, and the visual cues are often ignored. This is true for moth phishing messages and legitimate email: having a message that actually says something important brings the focus on the message rather than the visual appeal.

The phishing research studies frequently assess phishing related problems by performing laboratory studies. A series of "legitimate" and "phishing" emails are evaluated for their legitimacy, often alongside other aspects. This method provides data with stability from the stimuli, but is prone to being modified by the Hawthorne effect. Some attempts have been made to gather phishing data from real-world scenarios, either by sending out phishing emails or leveraging pre-existing attacks by third-party actors. Such studies often only send one or two emails out to participants, and the level of consent is not always made clear. In one case, participants did not volunteer to be part of the study (nor any other study), and researchers suffered backlash from unwilling subjects. In the current study, care was taken to ensure that all participant had been registered in the cover study and had been compensated for their participation. Additionally, no credentials were captured, stored, or viewed at any part of the process.

Chapter 3

METHODS

3.1 Introduction

This chapter covers the methods employed in this study. First, I discuss the ethical concerns involved in the subterfuge surrounding the research study. Next, I lay out the process of participant selection, and the overall structure of the research study, as well as the layout of the research being used as the cover for the phishing aspect.

Next, I discuss the process of creating phishing emails, and the mechanism through which they are deployed. I briefly comment on lessons learned and changes made. Further commentary on the lessons learned from this process are provided within the Chapter 5: Discussion.

3.2 Ethical Considerations

Institutional Review Board (IRB) approval was obtained prior to engaging in the phishing experiment. Informed consent for this experiment was not required nor obtained given the realism being sought in this study. Participants did consent to take part in the study being used as a cover, and were paid for their participation. The phishing email used was consistent with similar phishing emails individuals receive on a regular basis as a part of their daily life. However, I did take precautions to protect participants' privacy and data.

Due to the subterfuge involved in collecting data in a real-world scenario, I did not attempt to gain a user's information in any way. Links in phishing emails lead to redirects to the proper site, or end without any location for user input. Other studies that take user data as part of the phishing campaign do so in a laboratory setting, where users are aware they are being phished. As the participants are presented with the emails outside of

a controlled research environment, I did not want to include any data collection. All that I was tracking was whether a participant clicked on the links, when the participant received the email, and when they clicked on the link.

3.3 Participants

Initially, participants were contacted using flyers posted around UW Seattle and UW Bothell campus. In an effort to increase the participant pool, individuals were recruited through Facebook groups, such as those for local cites, neighborhoods, and for the university. Following the initial responses, participants were given a qualifying survey.

The participant pool contained 497 members. Of these, 189 (38%) were male, 303 (61%) were female, and 5 identified as other (1%).

3.4 Structure of Research Study

Participant information used in this experiment came from an unrelated study, termed the "cover study", for which they were recruited.

After participants finished the second survey of the cover study and were paid for their participation, they were considered eligible to receive the "phishing" emails. For the first phishing email, sent through GoPhish, participants who provided only a Microsoft email (i.e. @outlook.com, @live.com, @hotmail.com) were omitted from the participant pool due to the inability to pass phishing through the Microsoft automatic filtering (see Section 5.6 Crafting Phishing Emails). Subsequent phishing emails were sent through a different mechanism, which managed to circumvent Microsoft filters. Logs were kept of which participants had clicked on the link of a phishing email (see Section 5.7 Sending Phishing Emails), then analyzed relative to data obtained in the surveys.

3.5 Crafting Phishing Emails

Due to the limited time frame in which this research was performed, I attempted to ensure the emails had several controlled aspects that limited the differences between emails. I wanted

to see which types of emails would attract different demographics of the participants, so endeavored to make the key differences be the email offers, rather than the formatting. The aspects are detailed in Table 3.1.

Aspect	**Varied or Controlled**
Phishing Subject	Varied: Finance, Security, Job Offer
Tone	Varied:Opportunity or Warning
Sender	Controlled Format[1]
Spelling Mistakes	Controlled: Few to None
Time Pressure	Controlled: No explicit time.
Ratio of Text to Images	Controlled: Text and Logo

Table 3.1: Features of Phishing Emails

One of the key constraints when crafting emails was that participants would receive them through their normal email carriers. This meant that emails needed to be able to pass the email filtering used by common email providers. To perform this, I first sent emails to my personal and dummy accounts on GMail, Yahoo, and Outlook, sending them through GoPhish as I would the finished emails. If they did not pass the spam filter, I would modify them until they did. This stage caused me to reject several more complicated emails that used legitimate messages as a base and simply modified the message content and the URLs. While these emails looked more realistic, they were caught by filters and categorized as spam. Using simple text with logos managed to pass the email filtering steps more consistently, though many emails still failed.

However, as an additional complication, the adaptive nature of modern filters meant that those accounts used for testing quickly learned to be more permissive of the emails, as I was

[1]Each email is sent from a different sender, however the format of the sender's name and email are consistent. Each email is sent from [Division Name] <[Division Name]@[Sub-domain Matching Links]>

receiving and accepting most of them. To ensure that the emails would pass the filters on regular email, I created additional accounts on Yahoo, Outlook, and GMail which had never contacted any of the other accounts. Then I sent each email as a test to those accounts, as a final delivery validation.

Unfortunately, while phishing emails could pass the filtering on GMail and Yahoo, most emails were rejected by Outlook and sent directly to the Junk folder. The Yahoo account accepted any and all emails. The GMail accounts would filter most emails out as either Spam or as Phishing, both of which caused the emails to skip the inbox and go directly to the Spam folder. For emails that are deemed suspicious, but are sent to a GMail account from an established email account, links will show a popup warning that the link seems suspicious, before allowing the recipient to navigate to the destination.

3.6 Sending Phishing Emails

Initially, emails were sent though GoPhish, a software application designed to send out emails to a group of users, append tracking information to links automatically, and to spoof email names and addresses. The first email in this study, the Stimulus Check email, was sent with this method.

However, sending emails through GoPhish invoked stricter filtering on the receivers' side, which made several phishing emails impossible to deliver with this mechanism. In an attempt to address this, I created a python script to send emails and append tracking ids, linked to an actual email account instead of spoofing the sender. However, these messages were similarly rebuffed, so was not used in any campaigns.

Additionally, during the first round of emails with GoPhish, I found that after the 150th participant, no recipient opened the email or clicked on the links. As this was roughly half the group, all participants after the 150th were discounted as having received the first email, with the assumption that something prevented the delivery of the remaining emails. As GoPhish was not used in the remaining email tests, the causes and consistency of this phenomenon were not explored further.

The remaining emails were sent manually from a laptop using the Thunderbird email application and an add-on called MailMerge. The second set of emails was sent out with a 20 second delay between each email. With this set, a similar event to the first emails occurred, in which later emails were not clicked on by participants. In the second campaign, test accounts were added to the list in the middle and end, as well as in the beginning. This revealed the cause of the abrupt cut-off in responses. Later email messages were classified by the email providers as spam and directed away from participants' inboxes. Looking into this, it seems to be similar to the solution presented in previous research (Verma and Dhar, 2014).

3.7 Conclusion

The phishing emails used to obtain the data for this thesis were sent as a hidden aspect of a larger study. In the first phase, participants of the cover study were given surveys, and paid for their participation. Amongst other metrics, the surveys assessed individuals for their levels of positive and negative trait affect. Compensation was then distributed to the study subjects.

In the second phase, participants were sent phishing emails and were considered to be "phished" if they clicked on the link provided. The links redirected to a legitimate site or to an empty destination. Initially, emails were sent through GoPhish, then were later sent manually. Throughout this process, no participant credentials were collected.

Each link was assigned a parameter ID that identified which participant had been sent the email. The phished participants were tabulated by checking the RAW logs of the registered domains.

Chapter 4

RESULTS

4.1 Introduction

Within this chapter, I discuss the way in which the raw data was processed. I then present the results of the phishing campaigns, and the observed correlations with the trait affect of the participants. The implications of the results are discussed within Chapter 5: Discussion.

4.2 Results

4.2.1 Processing of Data

After acquiring the list of users who clicked on the data, I analyzed the correlations by looking at the Pearson Correlation and Spearman Correlations between the users who were sent the phishing emails and the values of trait affect observed in the cover study. Each Correlation was associated with a two-tailed significance p-value. The processing of the data was done using Statistical Package for the Social Sciences (SPSS).

Although referred to as Pearson Correlation above and in later sections, it is important to note that this is more specifically the mathematically equivalent Point-Biserial Correlation, due to the fact that the variable concerning a participant's state of following a hyperlink in an email is a binary choice, rather than a continuous scale.

4.2.2 Data

For the first campaign, a message pretending to be from the Internal Revenue Service (IRS) regarding Stimulus checks, 147 participants were sent the email. Of the 147 participants who received the email, 55 identified as male, 85 as female, and 2 as other. Within that

campaign, 16 participants (10.88% of all participants) clicked on the phishing link: 11 male (20.00% of males) and 5 female (5.88% of females).

In the second campaign, a generic email regarding an account sign-in, 221 participants received the email. The campaign targeted 84 male, 123 female, and four other participants. Of the targets, 16 participants (7.58% of all participants) clicked the phishing link, 6 male (7.14% of males) and 10 female (12.30% of females). Between both campaigns, 84 participants received both emails.

In the third campaign, a faux job opportunity, 278 participants received the email. The gender demographics comprised of 105 male, 171 female, and 2 other participants. Of the targets, 25 participants clicked the phishing link.

When compared against the other phishing campaigns and the higher-order dimensions of positive and negative trait affect, no significant correlations were found, see Table 4.1. However, when looking solely at the female population, there was a correlation between the Stimulus Check email and positive trait affect.

Trait	**Measure**	**Stimulus (N=142)**	**Log-In (N=211)**	**Job Opp. (N=290)**
Negative	Rho	-.024	.003	-.003
Affect	p-value	.777	.969	.963
Positive	Rho	-.099	-.010	.040
Affect	p-value	.243	.890	.497

Table 4.1: Correlations Between Phishing Engagement and Higher-Order Trait Affect

In addition to examining the correlations between the phishing studies and higher order positive and negative trait affect, I also examined the correlations between the phishing campaigns to the lower orders of trait affect, as seen in Table 4.2. Similar to the higher orders, few correlations were found. The sole correlation within the data is a moderate negative

correlation between trait joviality and a tendency to be phished in the first campaign. In this case, clicking on the phishing email was encoded as the "1" state. A negative coefficient here indicates that higher trait joviality is related to lowered phishing susceptibility in the faux-stimulus check campaign ($p \leq 0.01$).

Figures 4.1 and 4.2 display the mapping of a participant's positive and negative affect respectively against phishing click rate. The phishing click rate was determined by taking the number of campaigns in which the user was "phished" divided by the number of campaigns in which they were included.

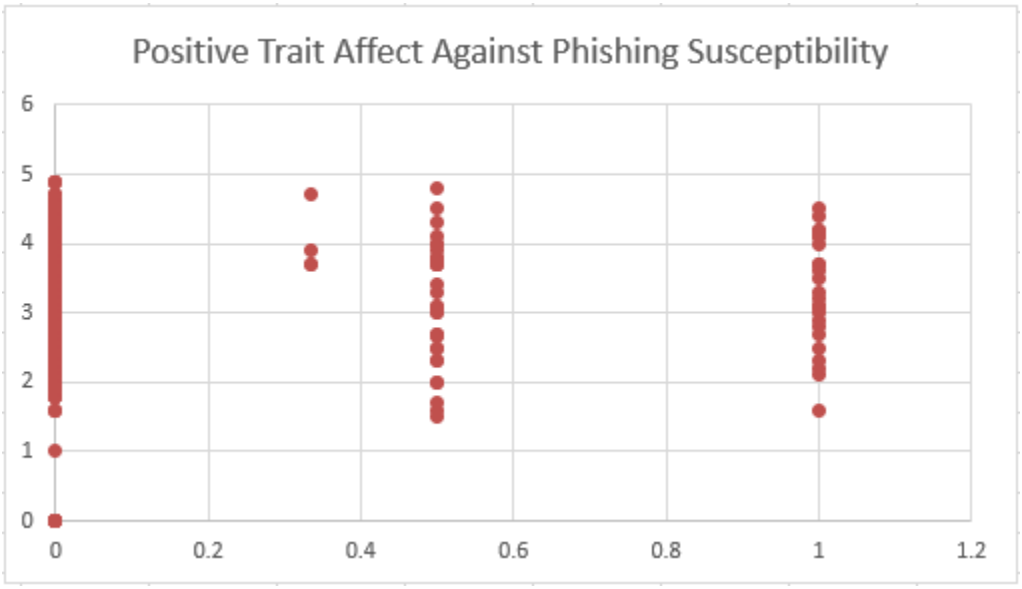

Figure 4.1: A scatter plot of positive trait affect and phishing rate.

Interestingly, when looking at the raw data, no participants fell for more than one campaign. While intriguing, this is not enough to support Hypothesis 1, as the limited number of emails does not allow for firm confirmation for or against the lack of intersection.

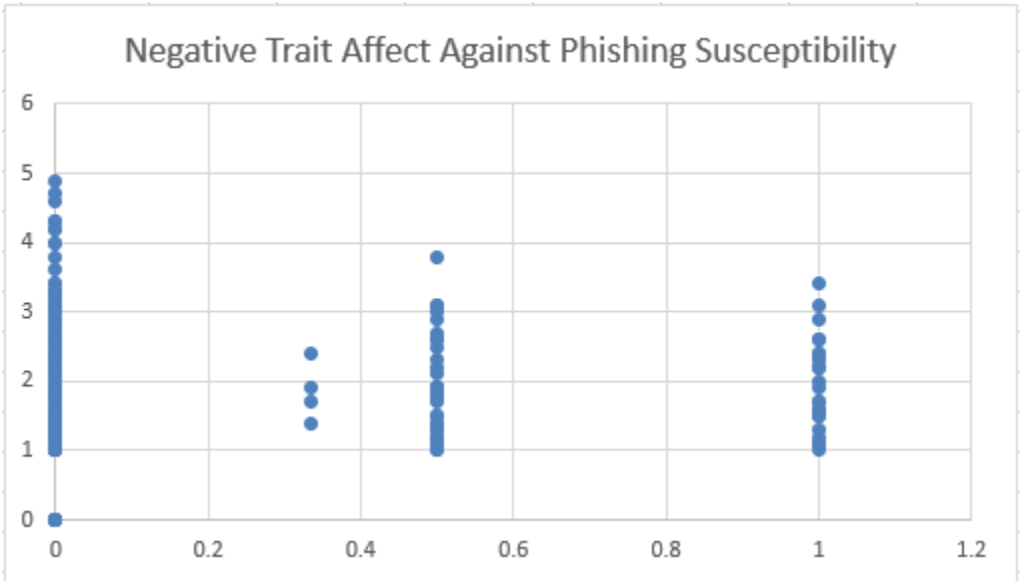

Figure 4.2: A scatter plot of negative trait affect and phishing rate.

4.3 Conclusion

After analyzing the response to the phishing campaigns, no significant correlations were found between either phishing campaigns or the higher orders of trait affect. When compared with lower-order dimensions of trait affect, one correlation was found between the first phishing campaign and trait joviality. Trait joviality is moderately negatively correlated with susceptibility to the first phishing campaign ($p \leq 0.01$).

When analyzing only the female participants, there was correlation between positive trait affect and phishing susceptibility to the Stimulus phishing email.

Trait	**Measure**	**Stimulus (N=142)**	**Log-In (N=211)**	**Job Opp. (N=290)**
Fear	Rho	-.087	.005	.029
	p-value	.305	.937	.628
Hostility	Rho	-.003	-.048	-.017
	p-value	.969	.491	.767
Guilt	Rho	.001	.048	-.042
	p-value	.995	.489	.481
Sadness	Rho	.041	-.032	-.039
	p-value	.632	.641	.508
Joviality	Rho	**-.223**	.094	-.019
	p-value	**.008**	.174	.747
Self Assurance	Rho	-.020	-.057	.005
	p-value	.817	.411	.929
Attentiveness	Rho	-.014	.004	.028
	p-value	.872	.959	.636
Shyness	Rho	-.052	.034	-.019
	p-value	.535	.627	.752
Fatigue	Rho	-.051	.032	.009
	p-value	.545	.649	.880
Serenity	Rho	-.008	.088	.004
	p-value	.925	.202	.939
Surprise	Rho	-.066	.074	.023
	p-value	.438	.283	.693

Table 4.2: Correlations Between Phishing Engagement and Lower-Order Trait Affect

Chapter 5

DISCUSSION

5.1 Introduction

Within this chapter, I discuss the limitations of the study, as well as obstacles posed when conducting the research. I also discuss the results and how they align with the hypotheses.

Several aspects of the study faced legal and technical challenges. Ultimately, the final counts of participants who were assessed was drastically lower than anticipated from the participants in the study. The results found little support for the proposed hypotheses, though the limited effective sample size in the study may be a contributing factor in the lack of observable correlations.

5.2 Limitations

The phishing emails used in this study did not request the user's information, to protect participants' data. Metrics are based solely on which users clicked the included "malicious" links.

During the period of this study, the COVID-19 pandemic created an unprecedented disruption in the daily life of participants, and several months of sustained societal anxiety bracketed the study. This environment may have increased the negative affect expressed in the results. Previous research studies suggest that the effect of the situation will not alter the inter-individual results (Diener and Larsen, 1984). Participants who would normally be above or below average should remain non-average in this situation (Oishi et al., 2004).

While most participants were undergraduate students, others were of an older demographic. This in particular affects the job-searching aspect of the study, as older demographics are less likely to be interested in job offers aimed towards undergraduates. However, due

to the life-long, stable nature of trait affect, age is not expected to sway results for other phishing subject types.

The spam filters protecting email services such as GMail and Outlook are both adaptive and aggressive. Many attempts to create emails that mimicked legitimate emails, such as banking correspondence or Amazon advertisements with information changed, would not make it past the phishing protection on the inboxes. While this is promising news for anti-phishing protection, this limited the capacity for phishing emails with more than text and limited images.

5.3 Lessons Learned

Over the course of this research, there were many opportunities to learn.

Legal Complications

One of the primary lessons learned within this study was the risk posed to the researcher in sending emails that impersonate entities. The first phishing campaign sought to impersonate a federal agancy by adopting the IRS branding and imagery, which is illegal.

Within a few hours of sending the first email, the domain host was contacted by a member of the Internal Revenue Service's Online Fraud Detection and Prevention team. Despite possessing ethics board approval for this research, remediation required an immediate cessation of the campaign, and a follow up email sent to all recipients that the previous email had not been legitimate communication from the IRS. The original research email facetiously adopted the IRS brand, in violation of 31 USC 333. The particular legal branch involved was the use of the IRS brand, not the sending of emails nor tracking links.

This is, perhaps, the largest takeaway when attempting to perform phishing research by sending email to participants. Though the ethical portion of this research was approved, it is far too easy to accidentally breach laws. Moving forward in the study, all email campaigns avoided impersonating any legitimate organizations. Campaigns instead either offered fictitious organization information or none at all.

Mistakes in the Sending Mechanism

A mistake in the first email campaign launched was that the sub-domain was stripped from the link during sending. This caused users to be sent to a page that was very obviously illegitimate. Users were shown the web page directory structure (Fig. 5.1) instead of being redirected to the official IRS page detailing the steps taken in regards to COVID-19.

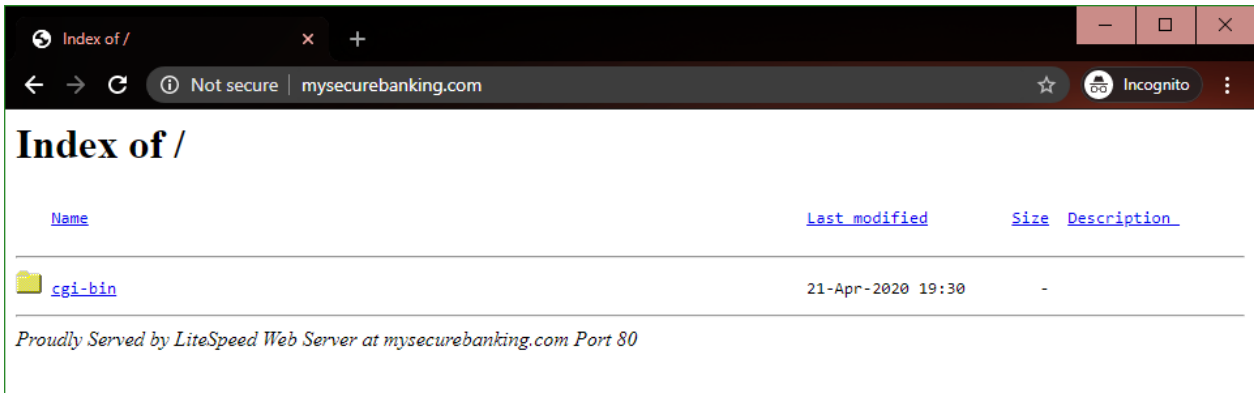

Figure 5.1: The hyperlink from the first campaign, showing the incorrect destination.

Additionally, when examining the results of the campaign, all events in which a user either opened an email or clicked on the link occurred within the first 150 users. The email had been sent to 296 participants, making it highly improbable that all events occurred strictly within the first half of participants. As such, all events after the first 150 were discounted, with the assumption that there was an error in sending the emails.

In the second email campaign, a similar cut-off in the emails was observed. Having noticed the decline in the previous campaign, the follow up added test accounts interspersed throughout the list of participants, to see if something was changing on the receiving end. After a number of emails had been sent to a specific carrier, emails were instead delivered to the Spam inbox. This was observed in both GMail and Yahoo. Interestingly, the spam notification on the GMail emails (fig. 5.3) was different to the standard spam notification (fig. 5.2). Where the default message identifies the email as similar to other emails, the

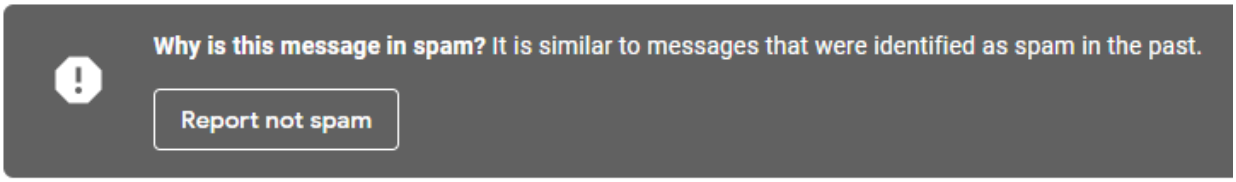

Figure 5.2: The generic spam notification from GMail.

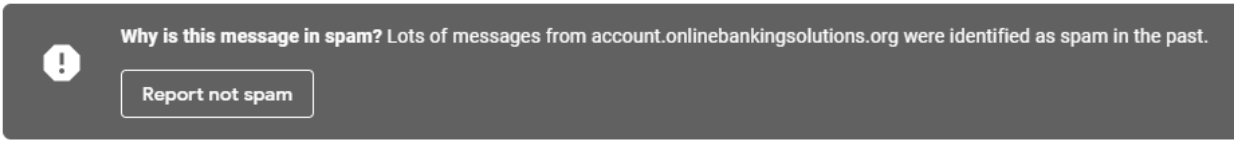

Figure 5.3: The tailored spam notification from GMail.

specific email identified the sender as having sent spam in the past.

Tracking Images

There were two intended statistics to determine phishing susceptibility, though only one was used reliably. Initially, each email had a clickable URL to test if a user would click a suspicious link and a tracking image that would act as a beacon once the email was opened. However, this email tracking image creates several false positives in the count of opened emails. GMail automatically renders the images in an email seconds after receiving it, without any input from the user. These events occur in the logs with the user agent containing "AppleWebKit/537.36", and obfuscate the count of opened emails. Tracking links with the user agent including "GoogleImageProxy" exhibited behavior suggesting that they were the result of human interaction opening the email. This assumed explanation of the log results is corroborated by findings by email marketing company GMass (Goel, 2019).

Due to the inconsistent results of the tracking images, after the first email was sent, tracking images were not included in the remaining campaigns.

5.4 Results

The results from the study were ultimately disappointing. The hypotheses were unsupported by the evidence. However, it is also noteworthy that for each campaign, a mere 16 participants were considered phished. As a sample size, 16 is remarkably small to establish correlations between what correlations exist. The total number of participants assumed to have received the email campaigns are 104 and 211 respectively, however, the count of those that looked at the email is unknown.

Hypothesis	**Description**	**Supported**
1	Response to phishing emails varies based on type.	Inconclusive
2	Response to phishing emails varies based on trait affect.	No
2A	Higher positive affect will relate to higher susceptibility.	No
2B	Higher negative affect will relate to lower susceptibility.	No

Table 5.1: Summary of Hypotheses

The many complications in the sending mechanism and in the phishing email creation mechanism does not provide enough evidence to support or contest Hypothesis 1.

The single correlation found was between the Stimulus campaign and the trait joviality ($r = -0.223$, $p \leq 0.01$). Joviality is a lower order dimension of positive affect. Higher levels of joviality related negatively to getting phished, which is counter to Hypothesis 2A. However, while there is some correlation between lower-order trait affect and one of the emails, there was no correlation between higher order trait affect and the emails. Ultimately, the data fails to support both Hypotheses 2 and 2B.

Chapter 6

CONCLUSIONS AND IMPLICATIONS

6.1 Concerning the Method

Ultimately, while I did receive data from my attempts, the use of real-world phishing exercises is very difficult. Aside from the potential legal issues, which should not be overlooked, the struggle of bypassing built-in phishing detectors makes this form of research a Sisyphean endeavor. By necessity, each email has to go through several attempts before being sent out, and each attempt weakens the accuracy of the email as a predictor, forcing one to create multiple accounts to escape the learned behavior from GMail.

Beyond the difficulty of sending emails, there is the additional complication that once emails are sent, it is non-trivial to determine what happens to it. The tracking images embedded do give some assurance that the email was received, but as discussed earlier, it is not a concrete value, with both false positives and false negatives clouding the use of it as a metric. Further complications can occur, such as the way in which the latter half of our participants seem not to have received the email, despite the email logs coming through. Though I have lowered the effective participant count based on this occurrence, due to the improbability of the phenomenon occurring without there being a flaw in the delivery, I did not initially know what caused this. Such uncertainties make it difficult to use real emails to test phishing susceptibility.

The same phenomena occurred when sending the second set of emails, which used a different apparatus to send the emails. For the second set of emails, I added test emails at multiple parts throughout the list, instead of only at the beginning. Through this, I found that the emails were being sent to spam after having received a certain number. Both GMail and Yahoo filtered the participants after reaching the later test accounts, though Microsoft

continued to accept the emails until the end.

In addition to the difficulties in sending emails, the large portion of participants that do not open the emails leads to higher required participant counts to get robust assertions about the correlations with phishing susceptibility and desired traits.

One comforting implication of the obstacles faced in the method is that the built in protections offered by email carriers, while not perfect, is good at limiting the spread of such emails. This does not mean that phishing is not a real and present danger, as seen both by the published phishing statistics (Verizon, 2019) and in the participants who were tricked within this study, but it does showcase ways in which the threat is being mitigated through a technological standpoint.

6.2 Concerning the Hypotheses

From the data obtained in this study, there is not enough statistical evidence to support the hypotheses that trait affect influences phishing susceptibility.

It is important to note that the amount of data was limited greatly by the method. For each campaign, only approximately 10% of the recipients who received an email.

6.3 Future Research

Ultimately, the largest takeaway from this research is the evaluation of the method. Attempting to phish participants in a real-world scenario brings a vast complication to the process of obtaining usable data. Participant pools are cut down through email carriers recognizing large movements of email. More research could be done in terms of which configurations maximize useful information gained and emails received by participants. Sending emails in batches small enough to be accepted by the sender's ISP and email carrier, but large and frequent enough to be delivered before participants begin marking the emails as phishing. This does, of course, assume that the transition into being recognized as spam is a result of participant action rather than the email carrier recognizing the large influx of emails.

Another potential improvement in an examination of the method would be to have a

control group, which is tested on the same emails within a laboratory setting. This will help asses how useful the minimization of the Hawthorne effect is in comparison to the complications added by using real-world delivery.

In regards to trait affect, I do believe there is a possibility of correlations between phishing email susceptibility and levels of trait affect. It should be tested using a more standard laboratory study, in which a participant's attention to the emails and decision-making process can be assessed.

Similarly to the trait affect conclusion, I believe the potential for email reception to differ based on the type of email to be worth investigating in a laboratory environment.

BIBLIOGRAPHY

Agarwal, R. and Karahanna, E. (2000). Time Flies When You're Having Fun: Cognitive Absorption and Beliefs about Information Technology Usage. *MIS Quarterly*, 24(4):665–694.

APWG (2017). Phishing Activity Trends Report.

Borkenau, P. and Mauer, N. (2006). Personality, emotionality, and risk prediction. *Journal of Individual Differences*, 27(3):127–135.

Canfield, C. I., Fischhoff, B., and Davis, A. (2016). Quantifying Phishing Susceptibility for Detection and Behavior Decisions. *Human Factors*, 58(8):1158–1172.

Costa, P. and Mccrae, R. (1992). Neo PI-R professional manual. *Psychological Assessment Resources*, 396.

Côté, S., Saks, A. M., and Zikic, J. (2006). Trait affect and job search outcomes. *Journal of Vocational Behavior*, 68(2):233–252.

Darrat, A. A., Darrat, M. A., and Amyx, D. (2016). How impulse buying influences compulsive buying: The central role of consumer anxiety and escapism. *Journal of Retailing and Consumer Services*, 31:103–108.

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3):319–340.

Diener, E. and Emmons, R. A. (1984). The independence of positive and negative affect. *Journal of Personality and Social Psychology*, 47(5):1105–1117.

Diener, E. and Larsen, R. J. (1984). Temporal stability and cross-situational consistency of affective, behavioral, and cognitive responses. *Journal of Personality and Social Psychology*, 47(4):871–883.

Downs, J. S., Holbrook, M. B., and Cranor, L. F. (2006). Decision strategies and susceptibility to phishing. In *Proceedings of the second symposium on Usable privacy and security*, SOUPS '06, pages 79–90, Pittsburgh, Pennsylvania, USA. Association for Computing Machinery.

Dupuis, M. J. (2014). The Role of Trait Affect in the Information Security Behavior of Home Users.

Fruhlinger, J. (2020). Top cybersecurity facts, figures and statistics for 2020.

Goel, A. (2019). Dealing with false opens in Gmail.

Grös, D. F., Antony, M. M., Simms, L. J., and McCabe, R. E. (2007). Psychometric properties of the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA): Comparison to the State-Trait Anxiety Inventory (STAI). *Psychological Assessment*, 19(4):369–381.

Habib, M. D. and Qayyum, A. (2017). A Structural Equation Model of Impulse Buying Behavior in Online Shopping. page 14.

Halevi, T., Lewis, J., and Memon, N. (2013). Phishing, Personality Traits and Facebook.

Harrison, B., Svetieva, E., and Vishwanath, A. (2016). Individual processing of phishing emails. *Online Information Review*, 40:265–281.

Helweg-Larsen, M. and Shepperd, J. A. (2001). Do Moderators of the Optimistic Bias Affect Personal or Target Risk Estimates? A Review of the Literature. *Personality & Social Psychology Review (Lawrence Erlbaum Associates)*, 5(1):74–95.

Isen, A. M., Nygren, T. E., and Ashby, F. G. (1988). Influence of positive affect on the subjective utility of gains and losses: It is just not worth the risk. *Journal of Personality and Social Psychology*, 55(5):710–717.

Izard, C. E., Libero, D. Z., Putnam, P., and Haynes, O. M. (1993). Stability of emotion experiences and their relations to traits of personality. *Journal of Personality and Social Psychology*, 64(5):847–860.

Jagatic, T. N., Johnson, N. A., Jakobsson, M., and Menczer, F. (2007). Social Phishing. *Communications of the ACM*, 50(10):94–100.

Jakobsson, M., Tsow, A., Shah, A., Blevis, E., and Lim, Y.-K. (2007). What Instills Trust? A Qualitative Study of Phishing. In Dietrich, S. and Dhamija, R., editors, *Financial Cryptography and Data Security*, volume 4886, pages 356–361. Springer Berlin Heidelberg, Berlin, Heidelberg.

Johnson, E. and Tversky, A. (1983). Affect, Generalization, and the Perception of Risk. *Journal of Personality and Social Psychology*, 45:20–31.

Kleitman, S. (2018). It's the deceiver and the receiver: Individual differences in phishing susceptibility and false positives with item profiling.

Meyer, G. J. and Shack, J. R. (1989). Structural convergence of mood and personality: Evidence for old and new directions. *Journal of Personality and Social Psychology*, 57(4):691–706.

Milkovich, D. (2019). 15 Alarming Cyber Security Facts and Stats.

Oishi, S., Diener, E., Napa Scollon, C., and Biswas-Diener, R. (2004). Cross-Situational Consistency of Affective Experiences Across Cultures. *Journal of Personality and Social Psychology*, 86(3):460–472.

Ormond, D., Warkentin, M., and Crossler, R. E. (2019). Integrating Cognition with an Affective Lens to Better Understand Information Security Policy Compliance. *Journal of the Association for Information Systems*, 20(12):1794–1843.

Pattinson, M., Jerram, C., Parsons, K., McCormac, A., and Butavicius, M. (2012). Why do some people manage phishing e-mails better than others? *Information Management & Computer Security*, 20:18–28.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, 110(1):145–172.

Tarditi, C., Hahnel, U. J. J., Jeanmonod, N., Sander, D., and Brosch, T. (2020). Affective Dilemmas: The Impact of Trait Affect and State Emotion on Sustainable Consumption Decisions in a Social Dilemma Task. *Environment and Behavior*, 52(1):33–59.

Taylor, K. M. and Shepperd, J. A. (1998). Bracing for the worst: severity, testing and feedback timing as moderators of the optimistic bias. *Personality & Social Psychology Bulletin*, 24(9):915–.

Thompson, E. R. and Prendergast, G. P. (2015). The influence of trait affect and the five-factor personality model on impulse buying. *Personality and Individual Differences*, 76:216–221.

Tsow, A. and Jakobsson, M. (2007). Deceit and Deception: A Large User Study of Phishing.

Turban, D. B., Lee, F. K., Veiga, S. P. d. M., Haggard, D. L., and Wu, S. Y. (2013). Be Happy, don't Wait: The Role of Trait Affect in Job Search. *Personnel Psychology*, 66(2):483–514.

Verduyn, P. and Brans, K. (2012). The relationship between extraversion, neuroticism and aspects of trait affect. *Personality and Individual Differences*, 52(6):664–669.

Verizon (2019). 2019 Data Breach Investigations Report.

Verma, R. and Dhar, J. (2014). Online Spam Filter for Duplicate or Near Duplicate Message Content Detection Scheme. *Journal of Convergence Information Technology*, 9:23–30.

Vishwanath, A., Herath, T., Chen, R., Wang, J., and Rao, H. R. (2011). Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model. *Decision Support Systems*, 51(3):576–586.

Wang, J., Herath, T., Chen, R., Vishwanath, A., and Rao, H. R. (2012). Research Article Phishing Susceptibility: An Investigation Into the Processing of a Targeted Spear Phishing Email. *IEEE Transactions on Professional Communication*, 55(4):345–362.

Watson, D. and Clark, L. A. (1984). Negative affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin*, 96(3):465–490.

Watson, D. and Clark, L. A. (1992). On Traits and Temperament: General and Specific Factors of Emotional Experience and Their Relation to the Five-Factor Model. *Journal of Personality*, 60(2):441–476.

Watson, D. and Clark, L. A. (1997). Measurement and Mismeasurement of Mood: Recurrent and Emergent issues. *Journal of Personality Assessment*, 68(2):267–296.

Watson, D. and Clark, L. A. (1999). The PANAS-X: Manual for the Positive and Negative Affect Schedule - Expanded Form. *Department of Psychological & Brain Sciences Publications*.

Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6):1063–1070.

Watson, D. and Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, 98(2):219–235.

Watson, D. and Walker, L. M. (1996). The long-term stability and predictive validity of trait measures of affect. *Journal of Personality and Social Psychology*, 70(3):567–577.