

Dissertation

**Investigating the Impact of Cognitive Style and Visual Behavior on  
Phishing Attack Susceptibility**

**Taoufik Sousak**

**University of Cyprus**



**DEPARTMENT OF COMPUTER SCIENCE**

**May 2023**

**UNIVERSITY OF CYPRUS**  
**DEPARTMENT OF COMPUTER SCIENCE**

**Investigating the Impact of Cognitive Style and Visual Behavior on Phishing Attack  
Susceptibility**

**Taoufik Sousak**

Supervisor

Prof. Andreas Pitsillides

Co-Supervisor

Dr. Marios Belk

The individual thesis submitted for partial fulfillment of the requirements for obtaining the  
degree of Computer Science, Department of Computer Science, University of Cyprus

May 2023

## Acknowledgments

I want to take this opportunity to express my gratitude to numerous individuals who have been instrumental not only in my academic journey but also in shaping the path I have chosen.

This thesis is dedicated to my mother, Lamis, who started this journey with me on day 1 of elementary school and tirelessly struggled to provide me with the resources I needed to thrive ever since. I am also grateful to my sister Aya, and my father, Muhannad, who have always been there for me, offering their unwavering support and encouragement.

I am thankful to Dr. Andreas Pitsillides for his invaluable guidance throughout the completion of my thesis. I also extend my heartfelt thanks to Dr. Marios Belk, who has placed his trust in me since my second year of university and has continued to support me to this day.

Lastly, I would like to express my appreciation to my friends Loukas Papalazarou for being a reliable academic and career partner; Stylianos Sofokleous, for the countless hours of assistance we have exchanged; and Andreas Papadopoulos, for sharing with me both the most challenging study nights and the most joyful moments during my studies.

*“Never regret your effort.”*

## Abstract

One of the biggest online security problems of our era is phishing attacks. In this study, we aimed to further our understanding of the factors that make users more susceptible to these phishing attacks in an effort to guide them into making safer decisions. In particular, **we wanted to examine the role that cognitive style differences, specifically field dependence, and eye gaze behavior play in people's susceptibility to phishing attacks.**

To examine this, we conducted an experiment with 50 volunteers (N=50), who took a GEFT test to determine their cognitive style (field-dependent or independent) and then we asked them to participate in a phishing attack experiment, during the experiment their eye gaze behavior was measured and think aloud protocol was applied. Our aim was to understand whether cognitive style differences and visual behavior could affect the ability of users to recognize email phishing attacks and whether certain patterns of eye gaze or cognitive style characteristics might be associated with a greater ability to avoid these attacks.

# Table of Contents

Acknowledgments .....	3
Abstract .....	4
1. Introduction and background .....	8
1.1 The problem with passwords .....	8
1.2 The danger of phishing attacks .....	9
1.3 Cognitive differences .....	10
2. Motivation and innovation.....	11
2.1. Related Works .....	11
2.2 Motivation .....	13
2.3 Innovation .....	13
2.4 Research Questions .....	14
3. Methodology .....	14
3.1 Tools and materials used.....	14
3.1.1 Group Embedded Figures Test (GEFT).....	14
3.1.2 Gazepoint Eye Tracking and Gaze Analysis .....	16
3.1.3 The Think Aloud Protocol .....	17
3.1.4 Technology stack .....	18
3.1.5 GPT - Large Language Models .....	18
3.2 Studying the effect of cognitive differences and eye gaze behavior on the vulnerability of phishing attacks.....	19
3.2.1 Recruitment.....	19
3.2.2 Preparation .....	20
3.2.3 Process .....	23

3.2.4 Sampling .....	24
3.2.5 Demographics .....	25
3.2.6 Eye Gaze - Areas of Interest .....	27
4. Relationship between cognitive style and susceptibility .....	32
4.1 Raw findings .....	32
4.2 Possible interpretation.....	35
5. What users look at when examining an email .....	36
5.1 Fixation Count Heatmaps and statistics .....	36
5.2 Fixation Count Statistics by Field Dependency Group.....	39
5.3 Fixation Count Statistics by Sex .....	40
6. Eye gaze behavior and susceptibility.....	42
6.1 T-Test and Revisit count distribution between tricked individuals and individuals that recognized the attack.....	42
6.2 The Mann-Whitney U test and area of interest revisits.....	45
6.3 Histograms and explanations .....	47
7. Think Aloud Protocol and Explanations.....	49
7.1 Relationship between Field Dependency and Sentiment. ....	51
7.2 Relationship between sentiment and actions taken.....	52
7.2.1 Sentiment for Spoofed email .....	53
7.2.2 Sentiment for Kaspersky email.....	53
7.3 Other factors.....	54
8. Conclusions.....	57
8.1 Summary of Findings.....	57
8.2 Implications.....	59
8.3 Limitations .....	60

8.4 Future Research Directions:.....	61
8.5 Final Thoughts .....	62
Bibliography .....	63
Appendix 1 – Python scripts for statistical analysis .....	67
Script 1: Formatting Data.....	67
Script 2: T-Test and histogram per area of interest.....	67
Script 3: Normality test and histogram script .....	68
Script 4: Mann-Whitney U test script for group (tricked/sex/etc.) and area of interest. ....	69
Script 5: Heatmap plotting of fixation count.....	69
Script 6: Mann-Whitney U test to determine whether total average difference is statistically significant. ....	70
Script 7: Chi-square test for independence .....	71
Appendix 2: Additional tests and graphs .....	74
Test 1: Relationship between shapes and spoofed email recognition. ....	74
Appendix 3: Consent Form.....	76

# 1. Introduction and background

## 1.1 The problem with passwords

The most common way to authenticate yourself, or in other words prove to a system that you are indeed who you're claiming to be online today is passwords. Passwords rely on secrecy to keep the user safe; this means that when a user chooses a password, they're expected to be the only person that knows said password. However, passwords have several shortcomings by design, that hinder their ability to be reliable and secure [1]. **The fact that passwords rely on secrecy is their worst vulnerability**, someone could come to know your password, either because they guessed it, or because you failed to keep it secure. Phishing attacks, for instance, are designed to do just that, trick the user into giving up their password by impersonating someone else. Other than that, passwords could be broken into using brute force attacks, or any other technique that does not need the user to voluntarily give up their credentials [2].

The fact that passwords need to be remembered, makes people find workarounds to make their life easier. **Users tend to reuse their passwords** for many services, and often for all the services that require a password. This allows an attacker who was able to obtain a password for one account, to obtain access to all the user's accounts at the same time.

Passwords can also be stolen through **data breaches** where the user has no role to play in the protection of their data. This means that even if you've chosen a very strong password and made an excellent effort at keeping it safe and secret, you may still be the victim of a data breach that has targeted a company in possession of your credentials.

For the aforementioned reasons, we can conclude that **passwords are not an effective means of authentication** in this day and age. It is still important for users to choose strong, unique passwords and keep them secret in order to keep their accounts safe, however, if possible, people should opt for alternative authentication methods, and ideally, two-factor authentication.



## 1.2 The danger of phishing attacks

As we mentioned before, **phishing attacks are one of the most common threats to credentials**. They usually work by impersonating a company or individual, through fraudulent emails or websites, and then ask the person to provide their credentials. Some phishing attacks can also happen through instant messaging apps, SMS messages, or even pop up notifications. Phishing attacks often rely on urgency and emotion to limit the time a person takes to process the information they've been provided.

According to the 2021 Verizon Data Breach Investigations Report [3], **32% of data breaches involved the use of phishing attacks**. The goal of these attacks is often to grant the attacker access to the victim's account and any sensitive information that is stored there, as a first step to a larger attack. The attacker can then use the information they have accessed to blackmail a company or proceed to hack other accounts within the same organization.

Phishing attacks can sometimes be very hard to recognize, this makes the exceptionally effective against some users. **Attackers usually impersonate a trusted entity**, such as a government organization (like the famous IRS scam in the USA), a bank, a courier service asking for a payment, a social network security team, or any other legitimate entity. They use official-looking logos, branding, and similar URLs to the impersonated entity, to seem more legitimate [4] [5].

Phishing attacks can also be very deceiving by using known **tailored information to a specific person or organization**. This is known as a **spear phishing attack**, and they are **one of the most effective** phishing attacks out there. By using the subject's name, interests, job role, or any other information that makes it seem like the entity is indeed whom they are claiming to be, they tend to convince subjects to give up their credentials more easily [5] [6].

Overall, one of the best ways for users **to be safe online is to be informed** about these kinds of attacks, and practice caution whenever they are asked to click a link or enter their

credentials, especially when the credentials are requested by a communication method such as emails or messages.

### 1.3 Cognitive differences

**Cognitive style refers to the way an individual prefers to traverse and process information**, and the way they tend to solve problems. It's an important factor to consider when trying to understand how people interact with their environment, and react to stimuli, as well as how people make decisions and solve problems. Recently there has been a growing interest in the scientific community to further understand how cognitive style can impact a person's perceptions, thoughts, and actions [19].

**Two of the categories in which people are often divided based on their cognitive style are field-dependent (FD) and field-independent (FI) individuals.** Field-dependent individuals tend to rely on the context and background information when processing information, while field-independent individuals tend to focus on the information itself and are less influenced by the context [19] [20].

One other way a person's cognitive style can affect their ability to perceive and process information is the following: **field-dependent individuals are usually more susceptible to the influence that their environment can have** and find it difficult to separate the relevant information from the less relevant information. Meanwhile, **field-independent individuals tend to be more objective and analytical** when it comes to processing information [21].

In addition to its impact on information processing, **cognitive style can also affect problem-solving abilities.** Field-dependent individuals may struggle with abstract problems, while field-independent individuals tend to excel in these types of situations [22] [23].

**Cognitive style can also affect decision-making**, with field-dependent individuals relying more on the opinions and input of others, while field-independent individuals tend to make decisions based on their own analysis and evaluation of information [21] [23].

Overall, cognitive style is an important factor that can have an impact on how an individual perceives, thinks, and acts. It's a relatively complex concept with many factors that should be considered when evaluated. **Researching cognitive style** and its effects on information processing, problem-solving, and decision-making, among other things, **can really help us deepen our understanding on how individuals see the world and interact with it** [19] [23] [40].

## **2. Motivation and innovation**

### **2.1. Related Works**

In the study by Van Der Heijden and Allodi [14], the authors wanted to see if they could use cognitive assessment tasks as a predictor of whether someone would fall victim to a phishing attack. They used data found in a large financial organization to find out how effective it would be to use automated cognitive assessment tests to detect susceptibility to phishing attacks. They found that they could indeed benefit from these tests and that they were able to **successfully predict whether a phishing attack would be successful**. Their work and findings have been used to make training on phishing attacks and campaigns more effective.

A study was conducted by Musuva, Gato, and Chepken where they used data from 192 cases to study the effects that cognitive processing has on the susceptibility to phishing attacks [15]. The researchers used various methods such as hypothesis testing, mediation analysis, and multi-group moderation to analyze the data that they had collected and found that **threat detection was indeed very effective at protecting users** from phishing attacks. They suggest that organizations should invest in training their staff so they could recognize the clues of a phishing attack and be better able to identify them as this was the most effective strategy.

In the paper by Wang et al [16], the authors conducted a study examining how individuals process and decide whether to respond to a phishing email. They used a real phishing email

as a stimulus and surveyed 321 members of a public university community who were intended victims of a spear phishing attack. The authors **found that attention to visceral triggers, such as stressing the urgency to respond, increases the likelihood to respond to a phishing email, while attention to phishing deception indicators, such as grammar errors and the sender's address, decreases the likelihood to respond**. Additionally, they found that knowledge of email-based scams increases attention to phishing deception indicators and directly decreases the likelihood of responding. The authors suggest that organizations should focus on increasing awareness of phishing threats and providing training to help individuals effectively use detection cues to identify phishing attacks.

The authors of *“The Phishing Email Suspicion Test (PEST) a lab-based task for evaluating the cognitive mechanisms of phishing detection.”* [17] wanted to study the factors that make individuals more susceptible to phishing attacks. They asked 227 people to take a test that asked about cognitive processing, phishing knowledge, and susceptibility to phishing attacks. The authors found that **cognitive processing and phishing knowledge had a direct effect on susceptibility to phishing attacks**. They recommended that institutions should try to educate people on phishing in order to decrease their chance of being scammed.

In *“Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model”* [18] Vishwanath et al. present an integrated model of phishing susceptibility grounded in prior research on information processing and interpersonal deception. They validate the model using a sample of intended victims of an actual phishing attack and find that it explains close to 50% of the variance in individual phishing susceptibility. The results suggest that **most phishing emails are peripherally processed** and that individuals make decisions based on simple cues embedded in the email. Additionally, the study finds that habitual patterns of media use and high levels of email load have a strong influence on an individual's likelihood to be phished and that computer self-efficacy influences elaboration but is diminished by domain-specific knowledge. This research contributes to understanding how individuals process and respond to phishing emails and the factors that influence their susceptibility to such attacks.

## 2.2 Motivation

Phishing attacks have become a significant issue in recent years, with individuals, companies, and institutions falling victim to these attacks on a regular basis. [24] As we've seen above, the research heavily supports that **understanding these attacks** and educating people about them and their dangers **are the best countermeasures we have**. [25] In light of this, my thesis aims to investigate the cognitive and behavioral factors that influence an individual's vulnerability to phishing attacks. [26]

Previous research in this area has provided valuable insights into the mechanisms underlying phishing susceptibility. For example, studies have shown that attention to **"visual triggers"** and **"phishing deception indicators" in emails influences decision-making** processes and ultimately the likelihood of responding to a phishing email [16]. Additionally, research has shown that **threat detection, cognitive processing, and phishing knowledge all play a role** in reducing an individual's susceptibility to phishing attacks [15][17].

## 2.3 Innovation

However, there is still much to be understood about the underlying mechanisms of phishing susceptibility and how to effectively reduce it. For example, research in this area has primarily focused on the role of cognitive factors, such as attention and elaboration, in phishing susceptibility, with **no found research on the effect of field dependency**. Additionally, there is a lack of research on the specific strategies that organizations can use to reduce the vulnerability of their employees to phishing attacks, other than education and training.

**My thesis seeks to address these gaps in the literature by investigating the impact of field dependency and eye gaze behavior on phishing susceptibility.**

To achieve these goals, I will conduct an experiment in which participants will receive phishing emails and their behavior and cognitive processes will be monitored using eye-tracking technology and a think-aloud protocol. By examining the impact of field dependency on phishing susceptibility and investigating the effectiveness of different

strategies for reducing vulnerability, **my thesis aims to contribute to a deeper understanding of the mechanisms underlying phishing susceptibility** and provide practical recommendations for organizations looking to protect themselves against these types of attacks. [27] We used email as the method of delivery of the attacks due to its practicality regarding eye tracking and the fact that it's historically the most common type of attack.

## 2.4 Research Questions

The aim of this study is to attempt and answer the following questions with as much accuracy as possible, by avoiding bias and applying appropriate experiment techniques and statistical analysis methodologies.

- RQ1. How do cognitive style differences between field-dependent and field-independent individuals affect their susceptibility to phishing attacks?
- RQ2. What do people examine when evaluating an email?
- RQ3. How does eye gaze behavior impact users' susceptibility to phishing attacks?
- RQ4. How can this information be used to guide users to be safer online?

## 3. Methodology

### 3.1 Tools and materials used

#### 3.1.1 Group Embedded Figures Test (GEFT)

The group embedded figures test is an accredited test that **aims to measure an individual's cognitive style**, specifically their field-dependence or field-independence. Field-dependence refers to a cognitive style in which an individual tends to rely on the global context or overall structure of a task or situation, while field-independence refers to a cognitive style in which an individual tends to focus on the individual parts or details of a task or situation [7].

The GEFT, or Group Embedded Figures Test, is a tool that researchers often use to explore the link between someone's thinking style and their behavior. **It is first administered to**

**determine the person's cognitive style**, and then other test are administered or the person is observed in various situations to see how this cognitive style effects how they act [7].

For instance, they might look at how a person's thinking style influences their ability to learn new things, solve problems, or even navigate space. In the field of education, the GEFT can be used to see how students' thinking styles affect their learning and how they understand directions. Additionally, the test is also used in the workplace to examine the connection between an employee's thinking style and their job performance [8].

**The GEFT is considered to be a reliable and valid measure of cognitive style**, with several studies finding strong correlations between scores on the GEFT and other measures of cognitive style [9].

The GEFT consists of a series of figures that are embedded within larger figures, and the participant is asked to **identify the smaller figures as quickly as possible**. The time it takes for the participant to identify the smaller figures is used to calculate their score, with a longer time indicating a more field-dependent cognitive style and a shorter time indicating a more field-independent cognitive style. A simple example used in the instructions of the test can be seen in Figure 3.1 while one of the shapes that's actually used in the test can be seen in Figure 3.2.

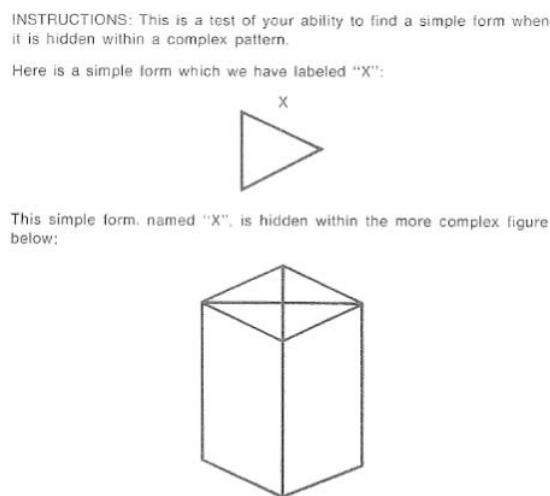


Figure 3.1 – GEFT instructions

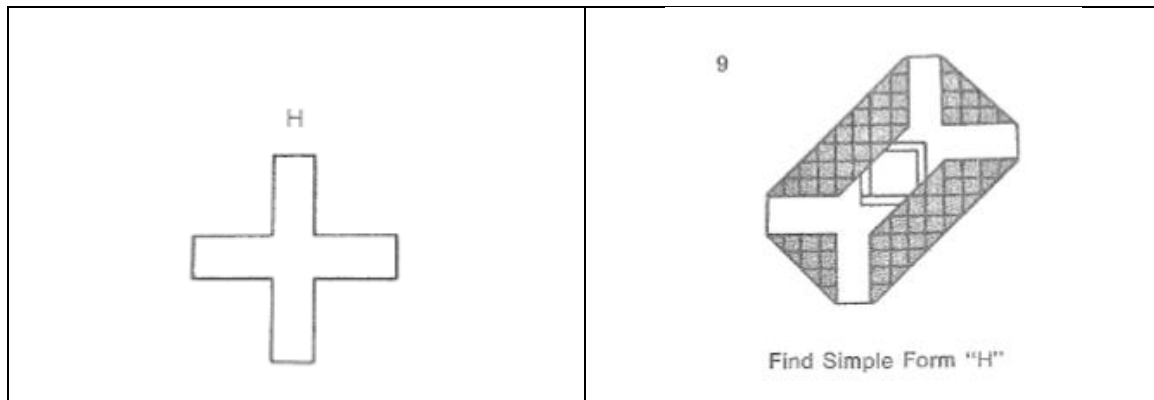


Figure 3.2 – An example from the GEFT

### 3.1.2 Gazepoint Eye Tracking and Gaze Analysis

The eye tracking device used in the study was the **Gazepoint GP3** which uses infrared technology to measure the gaze of the user. The device has a small camera and an infrared light source that can either be placed in front of a monitor or mounted on it using special equipment. **The camera then tracks the eye movements** of the user and stores their gaze data which can be analyzed afterwards using special software provided by the same company [28] [29].

The mentioned Gazepoint GP3's **software is used to analyze and visualize the gaze data** collected by the eye tracking device. The software allows the researcher to view and analyze the data in a variety of ways, such as by displaying gaze plots, heat maps, and other types of visualizations. It also provides tools for filtering and organizing the data, and for performing statistical analyses.

Gazepoint GP3's software is often used in research and other applied settings **to study eye gaze behavior and understand how it relates to various factors**, such as cognitive processes, attention, and decision-making. It is also used in fields such as usability testing and market research to understand how users interact with websites, software, and other digital products [29] [30].



This software also allows researchers to set **dynamic areas of interest** in the recorded video, in order to track the fixation count, revisits, time spent, and time to first view of these areas of interest. This has been extremely useful in the analysis of the eye tracking data collected in this study [30].

### 3.1.3 The Think Aloud Protocol

The think-aloud protocol is a method used in order to further understand the reasons behind an individual's decision-making in experiments and usability testing. **The user is requested to articulate their thoughts** and actions when performing a task, their thoughts can either be recorded, transcribed, or noted by researchers. This allows researchers to understand how participants are thinking about the task and what strategies they are using to solve problems or make decisions [10].

To use the think aloud protocol in an experiment, participants are typically asked to complete a task or perform some other activity while speaking aloud and describing their thoughts and actions as they go. **Researchers may also ask participants to provide additional verbal feedback or clarification if needed** [10] [11].

One of the most recognized benefits to using this protocol, is that **it allows the researchers to observe the cognitive processes of the participant** and understand it better. This method allows you to have insights into how people think about problems, and how they rationalize their methodology of solving them [10] [11] [12].

The think aloud protocol is **very easy to implement**, and it can be done without any special equipment. It can be followed either by hardware or by having a researcher aid the documentation of thoughts.

Unfortunately, this protocol also comes with some shortcomings, for instance, it can be very **time consuming** to have a researcher record what the participant is saying every time the study is run. Other than that, **the participant's ability to verbally express themselves or**

**willingness to share their thoughts may have an impact on the results.** Additionally, some research has suggested that the act of verbalizing thoughts may alter the cognitive processes of participants, which could potentially impact the validity of the results [12].

#### 3.1.4 Technology stack

**HTML (Hypertext Markup Language)** is the standard markup language used to create web pages. It provides the structure and layout of a webpage, including headings, paragraphs, images, and links [31].

**CSS (Cascading Style Sheets)** is a language used to describe the presentation of a webpage, including its layout, colors, and fonts. It allows for the separation of the presentation of a webpage from its structure and content, making it easier to maintain and update [32].

**Python** is an interpreted programming language. It is used a lot for web development, data analysis, scientific computing, artificial intelligence, and more. It is known for its simple, easy-to-read syntax, making it a popular choice for beginners [33].

#### 3.1.5 GPT - Large Language Models

A large language model is an advanced artificial intelligence system designed to understand and generate human-like text based on the input it receives. Built upon machine learning algorithms and vast amounts of data, these models are trained to recognize patterns and relationships within the text, enabling them to mimic human language effectively. One of the most prominent examples of a large language model is OpenAI's GPT series, which has demonstrated remarkable capabilities in a variety of tasks. Such models can be used in numerous applications, including natural language processing, machine translation, sentiment analysis, content generation, summarization, and question-answering systems.

**The GPT series has been used in this work to provide sentiment analysis** on the think aloud protocol comments [39].

## **3.2 Studying the effect of cognitive differences and eye gaze behavior on the vulnerability of phishing attacks.**

### **3.2.1 Recruitment**

For our experiment, **we set out to recruit 52 volunteers**. As explained in the sampling section, our methodology for recruitment was to reach out to individuals that we personally know who fit the criteria for balancing our demographics and informing them about the study. We contacted these individuals either by instant messaging or by phone call. For those who were interested to hear about the study, we provided a consent form (Appendix 3) that explained the purpose of the study, the information we were collecting, and how we were ensuring their privacy. We then booked appointments with the interested individuals at a time that suits them best.

Getting people to agree to spend 2 hours of their time to come to the premises of the lab and do an experiment without knowing details that would ruin the results did pose a challenge. **In total 108 people were asked to participate in the study, of which 52 ended up agreeing and showing up to their appointment.**

The **consent form explained that the purpose of the study was to understand how people interact with interactive systems and** that their participation would help improve research in usability and security. It also explained that their participation in the study was voluntary and that they could take a break or leave at any time.

**The consent form outlined the information we wanted to collect**, including interaction data, eye gaze behavior, responses to a questionnaire and a paper and pencil test, and notes on comments and actions. It also explained that we may publish research papers and reports that may include their comments and actions, but that their data would be anonymous, and their name and identity would not be linked to anything they said or did.

Finally, **the consent form outlined our policy for withdrawing consent and destroying personal data** and provided the names and contact information of the researchers who could be contacted if a participant wanted to withdraw their consent or request that their personal data be destroyed.

No further information was given about the study in order to maintain a level of authenticity and normal behavior by the participants.

### 3.2.2 Preparation

To prepare for our experiment, **we created two phishing emails and two corresponding websites** that were designed to mimic real-world phishing attacks.

**The first email was a replica of an email that Facebook sends when someone logs into your account** from an unexpected location. This email along with the relevant areas of interest is shown in Figure 3.3.

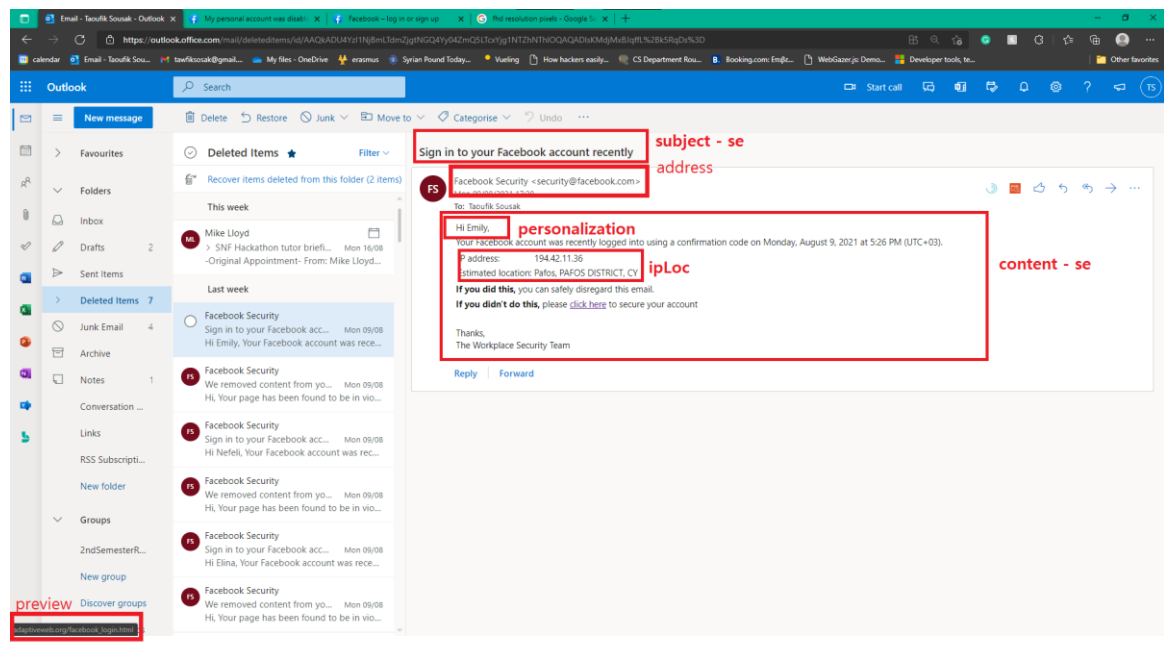


Figure 3.3 – Spoofed email of unexpected login (with AOIs)

**This email will be referred to as “Spoofed email” from here on out.**

Clicking the link in this email would lead the participants to a **spoofed replica of Facebook's login page**. This page along with the relevant areas of interest is shown in Figure 3.4.

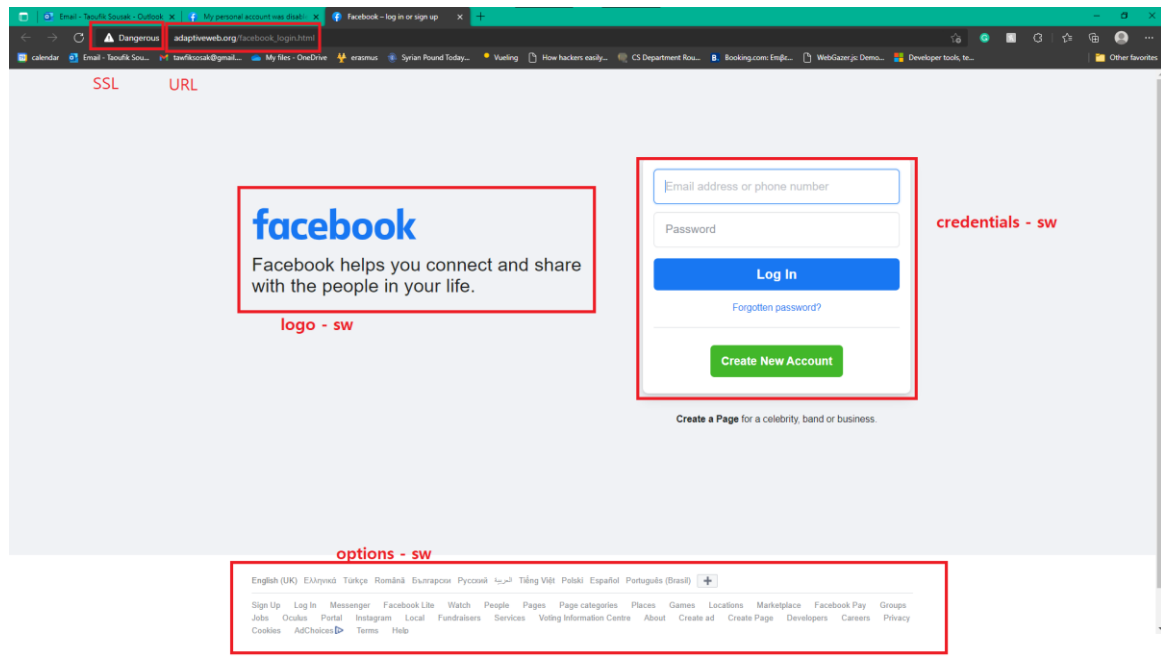


Figure 3.4 – Replica of Facebook's login page (with AOIs)

**This website will be referred to as “Spoofed website” from here on out.**

The second email was a replica of a popular phishing attack that was being investigated by Kaspersky [13]. This email along with its areas of interest is shown in Figure 3.5.

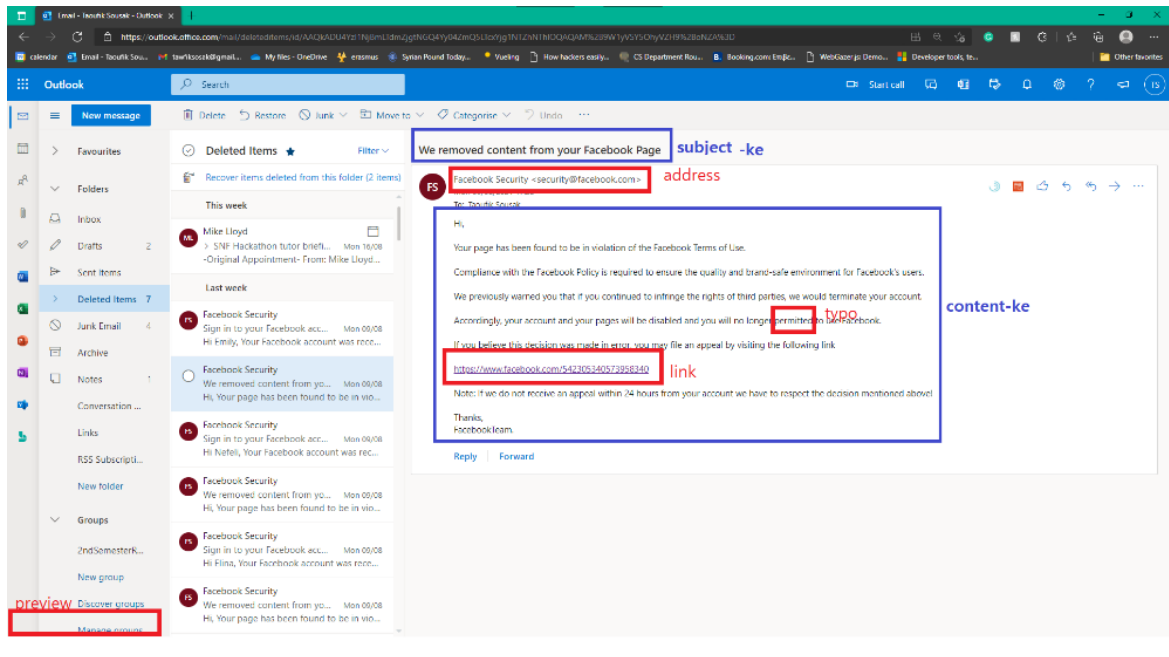


Figure 3.5 – Phishing email | Kaspersky (with AOIs)

**This email will be referred to as “Kaspersky email” from here on out.**

Clicking the link in this email would lead the participants to a **replica of the page used in the actual phishing attack in circulation**. This website and its areas of interest can be shown in Figure 3.6.

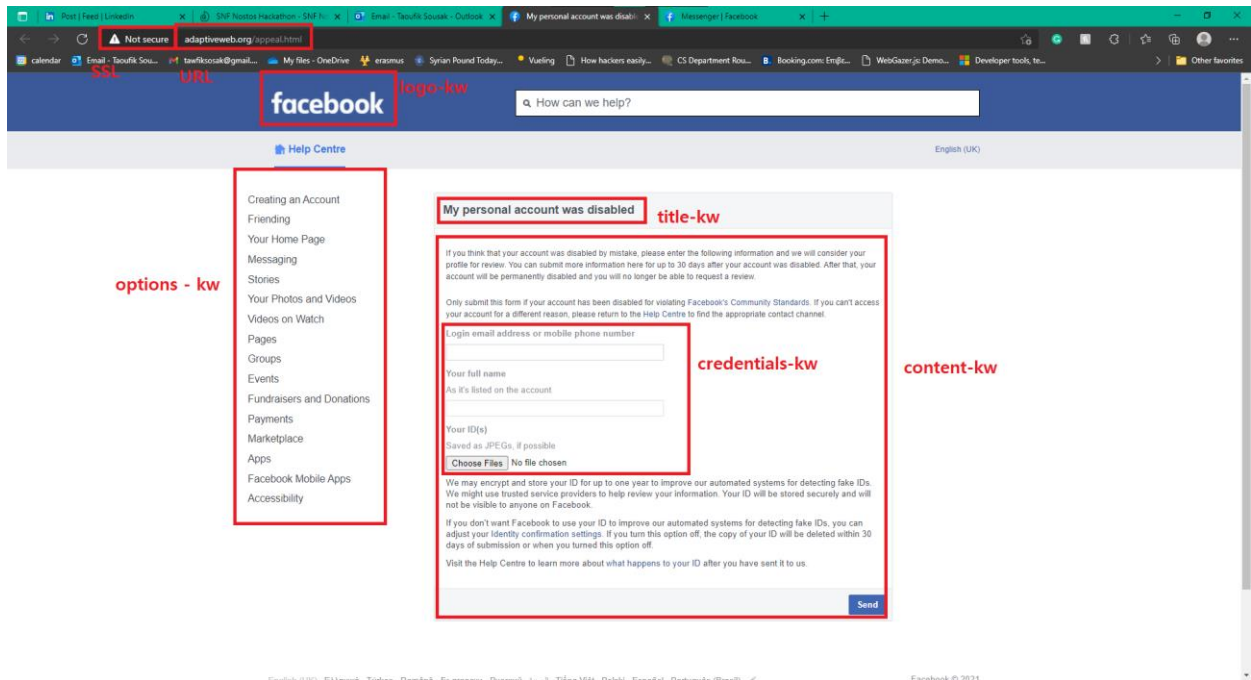


Figure 3.6 – Phishing website | Kaspersky (with AOIs)

**This website will be referred to as “Kaspersky website” from here on out.**

These emails were sent by the correct email that Facebook security would use to contact its users in these scenarios ([security@facebook.com](mailto:security@facebook.com)) and led to a replica of the website that this attack led to.

In creating these phishing emails and websites, we took care to ensure that they were as realistic as possible.

While choosing our phishing emails, we wanted to represent emails that are realistic, recent and being circulated out there, so using the one that Kaspersky had identified was a suitable option. Other than that, we wanted to have a base email that looks as much like a real email as possible so we borrowed one of Facebook’s real emails and spoofed it. **We took great care to ensure that the phishing emails and websites we created were realistic, well-designed, and compliant with relevant laws and regulations.**

### 3.2.3 Process

To conduct our experiment, we followed the following procedure:

1. We **administered the Group Embedded Figures Test (GEFT)** to all participants in order to classify them as either field-dependent (FD) or field-independent (FI). The GEFT test was taken in groups of 1 to 4 people, and the instructions were explained by the researcher according to the guidelines.
2. We **familiarized participants with the eye tracking device** and the experimental environment. The device was **calibrated for each user** before the experiment began.
3. Participants were asked to log into their personal email accounts and wait for two emails from "Facebook security." **They were instructed to act as if they were at home** with no limitations on what they could do.
4. Each participant **received two phishing emails** leading to two spoofed websites.

During the experiment, **we monitored whether the participant opened the email, whether they clicked the link, and whether they filled in their login information.**

At each of these stages, **the eye-tracking device recorded what the participant examined** in order to make their decision. The time taken between each stage was also measured.

**The think aloud protocol was followed** throughout the experiment.

### 3.2.4 Sampling

For the sampling of this experiment, **we used a hybrid of the proportional stratified sampling technique with the convenience sampling technique.**

The stratified sampling guarantees the desired representation of relevant subgroups within the sample. **It is often used when the objective of the study is to compare the behavior of subgroups**, in our case we wanted to compare the behavior of field dependent and field independent individuals. While we couldn't know the field dependency of participants before recruitment, we tried to balance the gender and background of individuals as much as possible [35].



However, due to the lack of time, manpower, and resources, **it was impossible for the researchers to strictly follow this technique** for choosing their sample. This is where convenience sampling comes in [36] [37].

The age range of the participants was 18-25, mainly because of their availability and accessibility to the researcher. Other than that, the field of study of participants was as close to real-world analogies as possible. In other words, **we followed a best-effort approach to gather our participants** while complementing our sample size with whoever was available and willing to participate.

### 3.2.5 Demographics

To get a feel of the participants' demographics, take a look at the following tables. The breakdown for the participants' sex can be seen in Figure 3.7.

SEX	
F	28
M	22
<b>Grand Total</b>	<b>50</b>

Figure 3.7 – breakdown of sex of participants.

The breakdown of participants by their general field of studies or work can be seen in Figure 3.8.

FIELD OF STUDY OR WORK	COUNT
SOCIAL AND EDUCATIONAL SCIENCES	13
INFORMATION TECHNOLOGY	12
ECONOMICS AND ADMINISTRATION	6
SCIENCE AND ENGINEERING	6

ARTS	4
HUMANITIES	3
AVIATION AND MARINE TRANSPORTATION	2
SPORTS - PHYSICAL THERAPY	2
MEDICINE	2
GRAND TOTAL	<b>50</b>

Figure 3.8 – Breakdown of the general field of study of participants

The average age of participants was 20.32, with an age range of 18-25 as seen in Figure 3.9.

<b>AGE</b>	
AVERAGE AGE:	20.32
AGE RANGE:	18-25

Figure 3.9 – age of participants.

Fortunately, the breakdown of **field dependency** was also **balanced**, even within subcategories, this can be seen in the following demographics tables. The field dependency breakdown can be seen in Figure 3.10.

<b>Field Dependency</b>	
Field Independent	24
Field Dependent	26
<b>Grand Total</b>	<b>50</b>

Figure 3.10 – Field Dependency breakdown.

The breakdown of field dependency by field can also be seen in Figure 3.11. Although it is fairly balanced, it is worth noting that Social and Educational studies did have a significantly more field-dependent share, also there were no field-independent participants in the humanities section.

<b>Field Dependency by Field</b>			
Row Labels	FD	FI	Grand Total
ARTS	1	3	4

AVIATION AND MARINE TRANSPORTATION	1	1	2
ECONOMICS AND ADMINISTRATION	2	4	6
HUMANITIES	3		3
INFORMATION TECHNOLOGY	5	7	12
MEDICINE	1	1	2
SCIENCE AND ENGINEERING	2	4	6
SOCIAL AND EDUCATIONAL SCIENCES	10	3	13
SPORTS - PHYSICAL THERAPY	1	1	2
<b>Grand Total</b>	<b>26</b>	<b>24</b>	<b>50</b>

Figure 3.11 - Field Dependency by Field

The breakdown of field dependency by sex can also be seen in Figure 3.12.

SEX	FD	FI	Grand Total
F	19	9	28
M	7	15	22
<b>GRAND TOTAL</b>	<b>26</b>	<b>24</b>	<b>50</b>

Figure 3.12 – Field Dependency by Sex

Though we could not have intentionally balanced the field dependency, due to the fact that the assessment of field dependency was part of the experiment and took place in the lab after recruitment, **we were lucky enough to have group sizes that are very close to ideal.**

### 3.2.6 Eye Gaze - Areas of Interest

After the study, we had to go through the screen recordings of all the participants and **create areas of interest** that were outlined in section 4.2 Preparation for each video. The areas were chosen after careful examination of relevant papers and articles. In general we followed the “better safe than sorry” approach and we tried to cover all the possible sections that could have significance in the results [34].

For the Spoofed email, the chosen areas of interest and what they stand for can be seen below:

- Subjectse: The subject of spoofed email.
- Addresse : The email address for the spoofed email.

- Pers: The name of the participant mentioned in the email.
- Iploc: The ip address and location that the email claims the login came from.
- Contentse: The content of the email in general.

For the Spoofed website, the chosen areas of interest and what they stand for can be seen below:

- Sslsw: The ssl certificate of the spoofed website.
- Urlsw: The url of the spoofed website
- Logosw: The facebook logo and moto in the spoofed website
- Credentialssw: The credentials box in the spoofed website

For the Kaspersky email, the chosen areas of interest and what they stand for can be seen below:

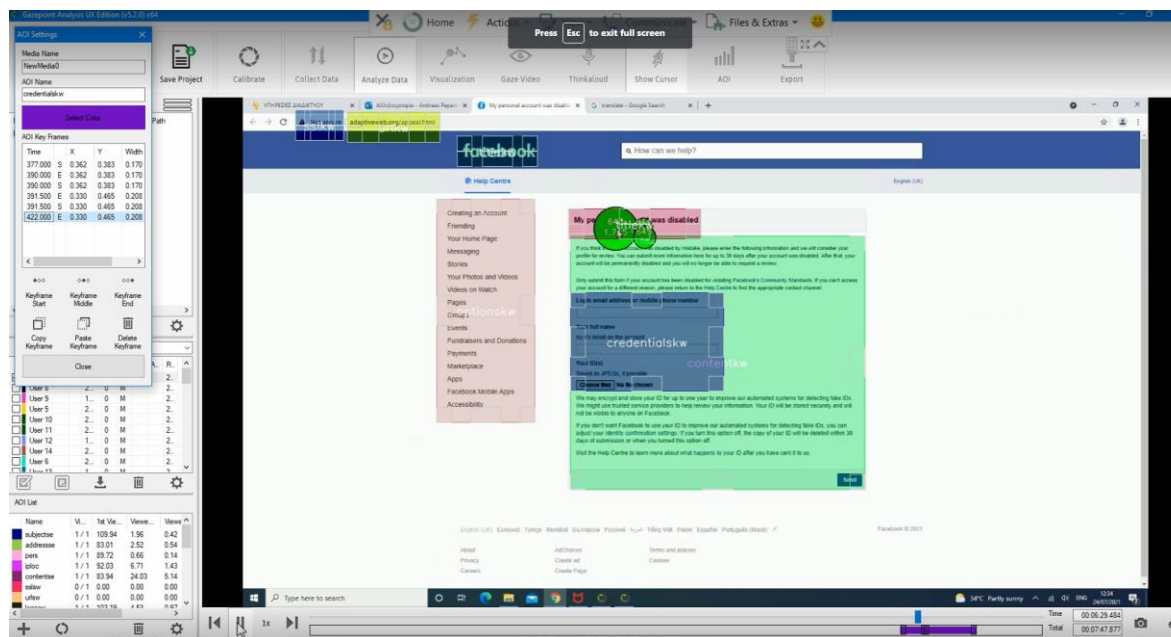
- Subjectke: The subject of the Kaspersky email
- Addresske: The email address for the Kaspersky website
- Typo: A grammatical error that was present in the email
- Link: The link that them email wanted the participants to click
- Contentke: The content of the email in general.

For the Kaspersky website, the chosen areas of interest and what they stand for can be seen below:

- Sslkw: The ssl certificate of the Kaspersky website.
- Urlkw: The url of the Kaspersky website
- Logokw: The facebook logo and moto in the Kaspersky website
- Credentialskw: The credentials box in the Kaspersky website
- Optionskw: The options sidebar in the Kaspersky website
- Titlekw: The heading of the website's main section
- Contentkw: The contents of the website in general.

One challenge that we faced is that due to the lack of restrictions on what the participants were allowed to do, they tended to scroll, zoom in and out, and change tabs frequently. This meant that **we had to create dynamic areas of interest** that followed the exact alignment of the screen of each and every video.

One example of a user zooming in can be seen in Figure 3.13 where the areas of interest changed in size and location in order to remain accurate.



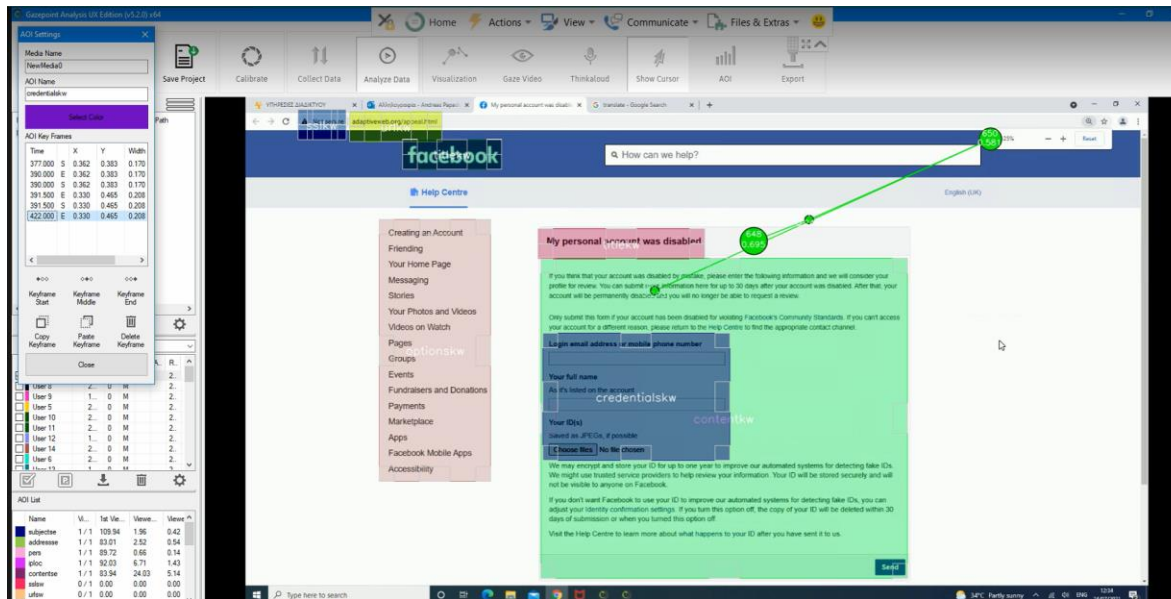


Figure 3.13 – Areas of interest adjusting to zoom in

The areas of interest were selected to be of different size and coordinates as shown in Figure 3.14 in order to simulate these dynamic areas of interest.

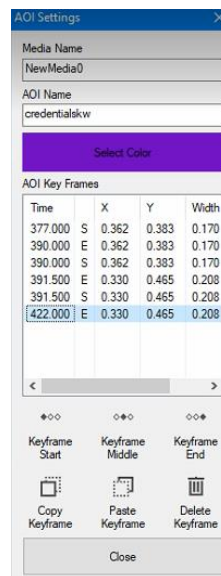


Figure 3.14 – Settings of credentialskw in order to simulate dynamic AOIs.

Sometimes, the area of interest completely changes in shape as well, however we still wanted it to be measured, for example take a look at Figure 3.15 where the address area of interest changes throughout the experiment.

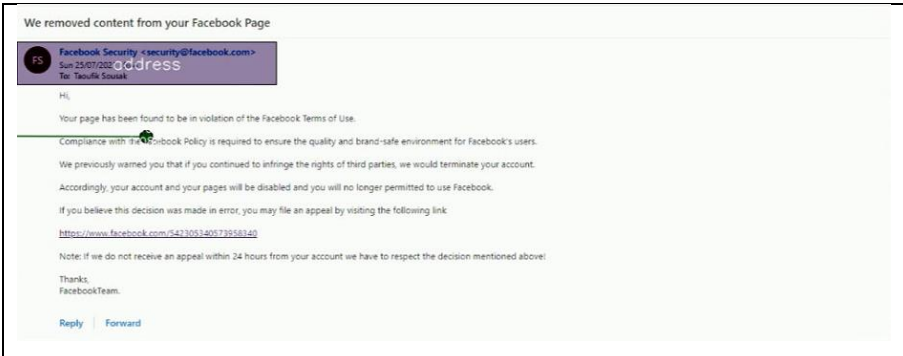
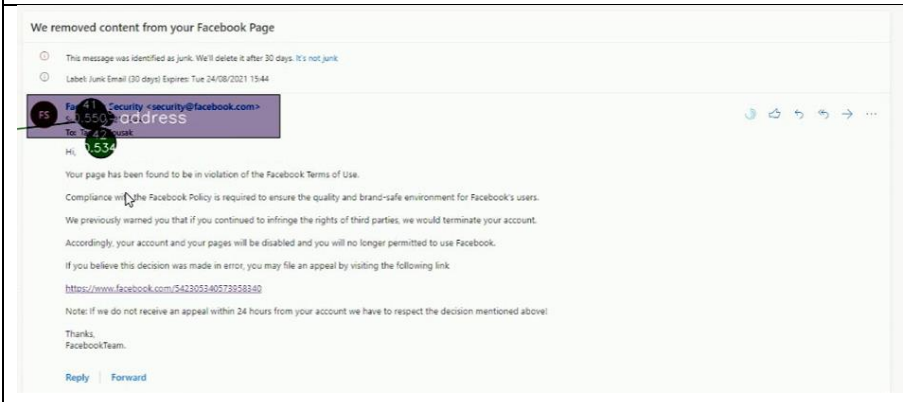
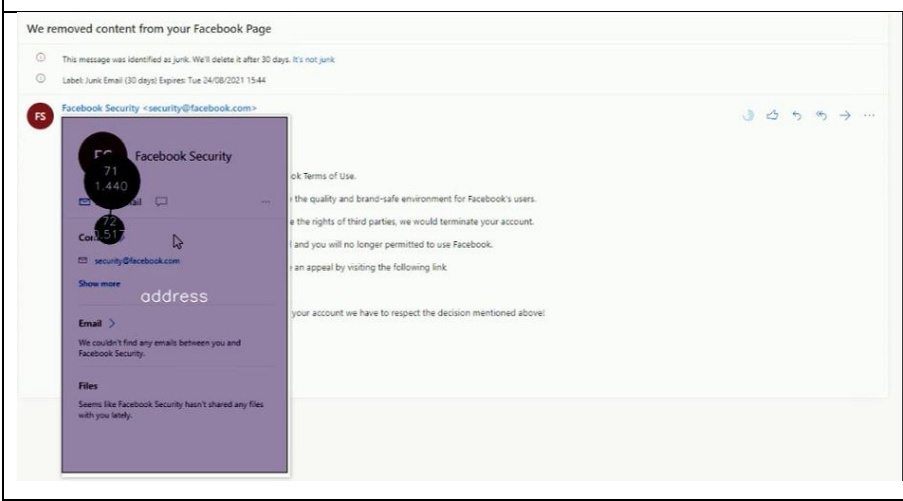
	Initial position of Address area of interest.
	Area of interest position moved down due to notification pushing the address further down.
	Address area of interest adjusting in size and shape to accommodate the popup that the user is examining.

Figure 3.15 – Address area of interest adjusting to changes.

## 4. Relationship between cognitive style and susceptibility

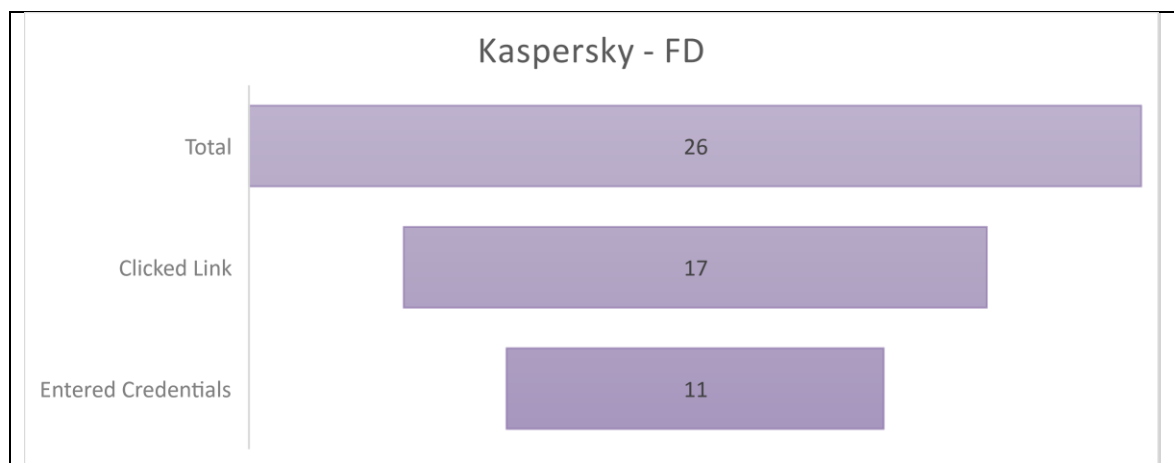
### 4.1 Raw findings

Our initial results suggest that **cognitive style may have an impact on how individuals interact with phishing emails**. In our sample of 50 participants, we found that **field-dependent individuals were more likely to click on the malicious link** in the phishing email both in Kaspersky email (17 out of 26 participants) and spoofed email (16 out of 26 participants) compared to field-independent individuals (11 out of 24 participants in both cases). The specific breakdown of actions taken by participants can be seen in Figure 4.1 and Figure 4.2 for the Kaspersky email and in Figure 4.3 and Figure 4.4 for the spoofed email.

Interestingly, in the spoofed email, field-independent individuals were less likely to enter their credentials after clicking on the link (7 out of 24 participants) compared to field-dependent individuals (13 out of 26 participants), however, in the Kaspersky email, the difference is negligible with between field-dependent (11 out of 26 participants) and independent (9 out of 24 participants) individuals.

Row Labels	FD	FI	ALL
Recognized attack	9	13	22
Clicked link	17	11	28
Entered credentials	11	9	20

Figure 4.1 – Actions taken in Kaspersky email.





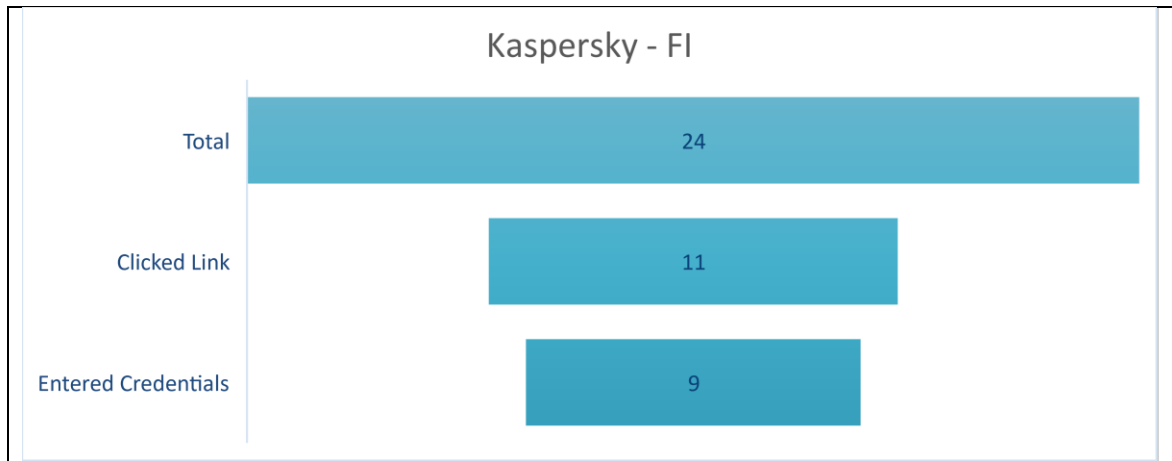
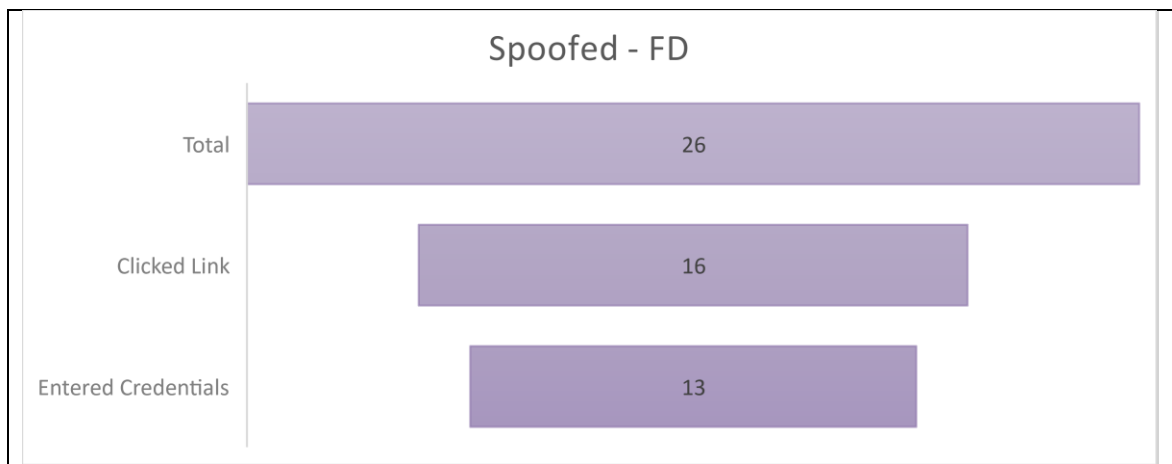


Figure 4.2 - Actions taken in Kaspersky email.

Row Labels	FD	FI	All
Recognized attack	10	13	23
Clicked link	16	11	27
Entered credentials	13	7	20

Figure 4.3- Actions taken in Spoofed email.



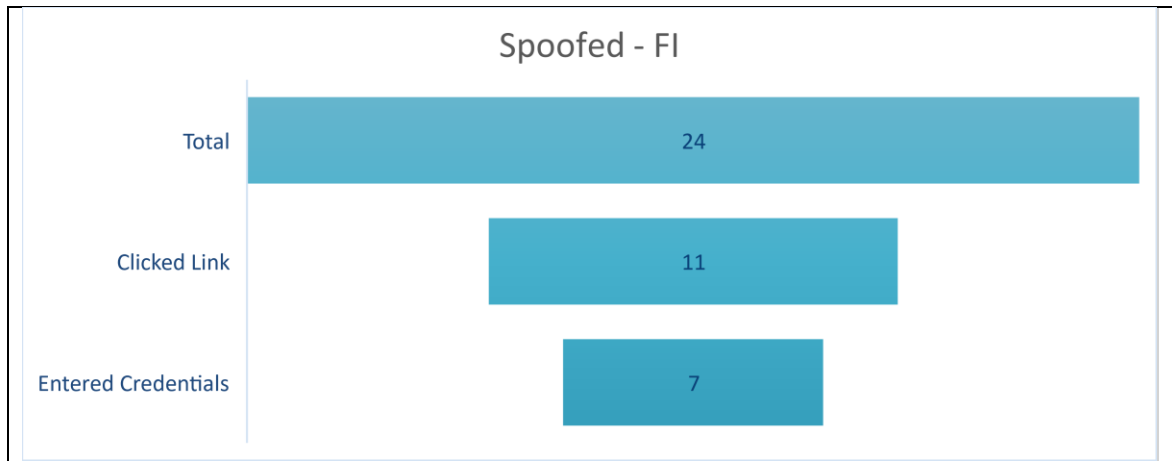


Figure 4.4- Actions taken in Spoofed email.

Simply looking at the percentage of participants that recognized the email as malicious, we can see a clear difference between the two cognitive styles under examination. If we combine both emails and sum the instances of participants recognizing the danger of engaging with the link, versus clicking it, we can see that **Field Independent people are 18% more likely to recognize the phishing attack**. This can be seen in Figures 4.5 and 4.6 below showing the percentage of tricked Field Independent and tricked Field Dependent individuals respectively.

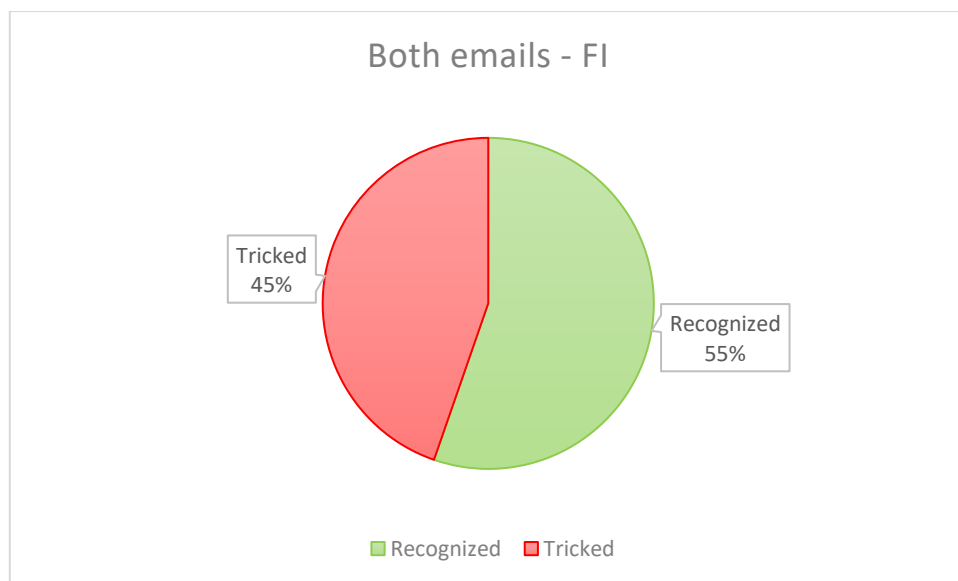


Figure 4.5 – Percentage of FI people recognizing the phishing attack or getting tricked.

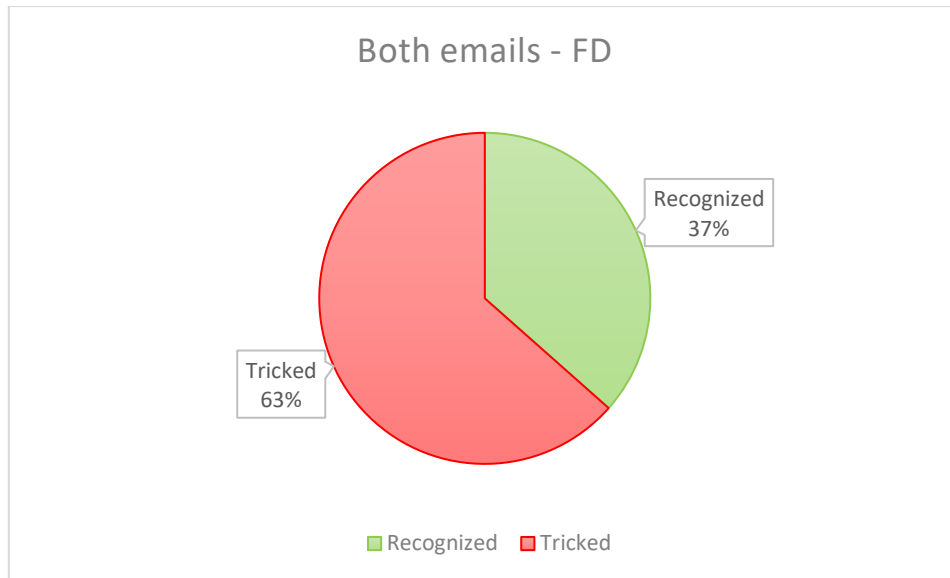


Figure 4.6 – Percentage of FD people recognizing the phishing attack or getting tricked.

These findings suggest that **field-dependent individuals may be more susceptible to phishing attacks**, as they are more likely to engage with malicious links and provide their credentials.

However, it's worth noting that **the sample size is relatively small and further analysis is needed** to draw any definitive conclusions. Additionally, it's worth noting that the type of phishing email can also affect the results.

## 4.2 Possible interpretation

The theory of cognitive differences suggests that individuals who are field-dependent tend to rely on external cues and context when processing information, while individuals who are field-independent tend to rely on their own internal frame of reference.

In the context of phishing emails, **field-dependent individuals may be more likely to click on the malicious link in the email and enter their credentials** because they are more likely to rely on external cues such as the appearance of the email, the branding, and the language used in the email.

They may also be more susceptible to the sense of urgency or other emotional triggers used in the email. On the other hand, **field-independent individuals may be more likely to rely on their own internal frame of reference and critically evaluate the email**, using their own knowledge and experience to determine whether the email is legitimate or not. This could explain why field-independent individuals are less likely to click on the malicious link and enter their credentials in the spoofed email.

## **5. What users look at when examining an email**

To **examine what areas users are drawn to** when examining an email, we looked at the average fixation counts, and heatmaps for the areas of interest. The first step we took was to draw a fixation count heatmap of all the areas of interest.

Although we have collected eye tracking data for the entirety of the test, we made a strategic decision to **focus on the eye tracking data collected from the spoofed email only**, due to several reasons. For instance, we chose to allocate our limited time and resources to examining data from one source in more depth to ensure the statistical power of our analysis. The data we had from the rest of the screens had some inconsistencies, or missing values, this is why we chose the spoofed email recordings.

For the rest of this document, we will only refer to the data collected from the spoofed email eye tracking.

### **5.1 Fixation Count Heatmaps and statistics**

To determine the frequency of which users look at certain areas of interest, **we plotted a heatmap** of the fixation counts (Appendix 1 – Script 5) on all the areas of interest as you can see in Figure 5.1

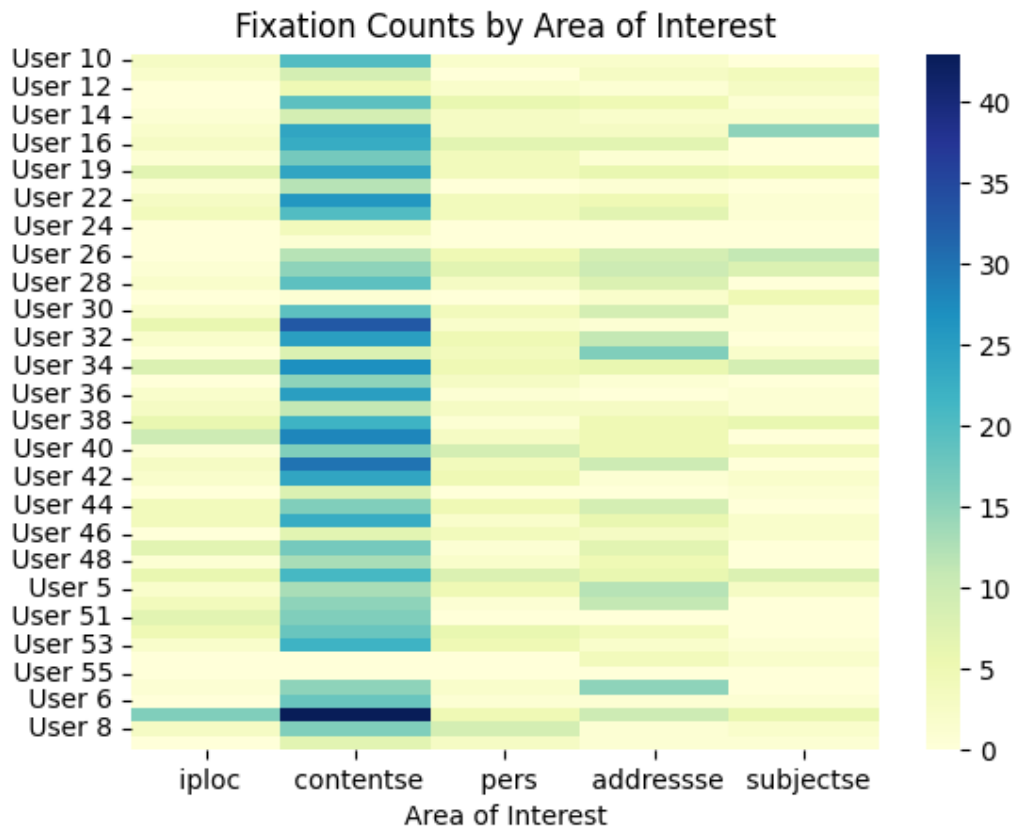


Figure 5.1. – Area of interest heatmap for all users

However, due to the fact that the contentse area of interest is big and contains most of the areas of interest within it, as seen in Figure 6.2., **we decided to run it again without the contentse area** to give ourselves a chance to observe the differences between the rest of the areas of interest.

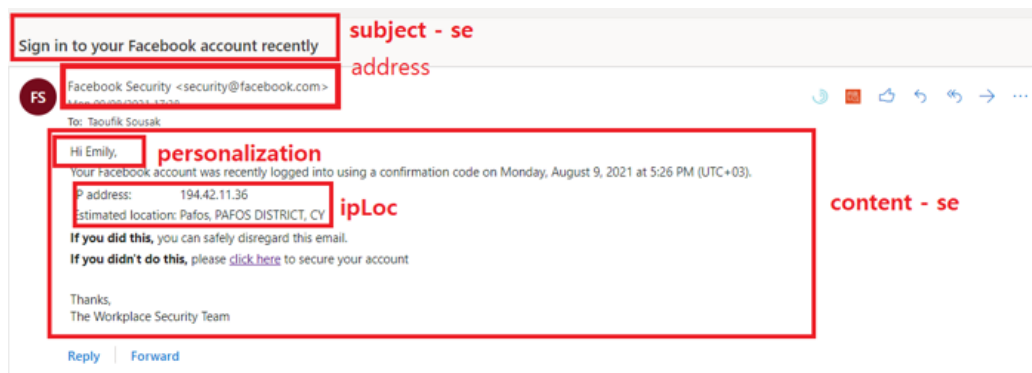


Figure 5.2. – All areas of interest for the spoofed email.

The new heatmap of the fixation counts per area of interest can be seen in Figure 5.3. where we can see a slight tendency for users to look more times at the address from which the email was received titled addressse.

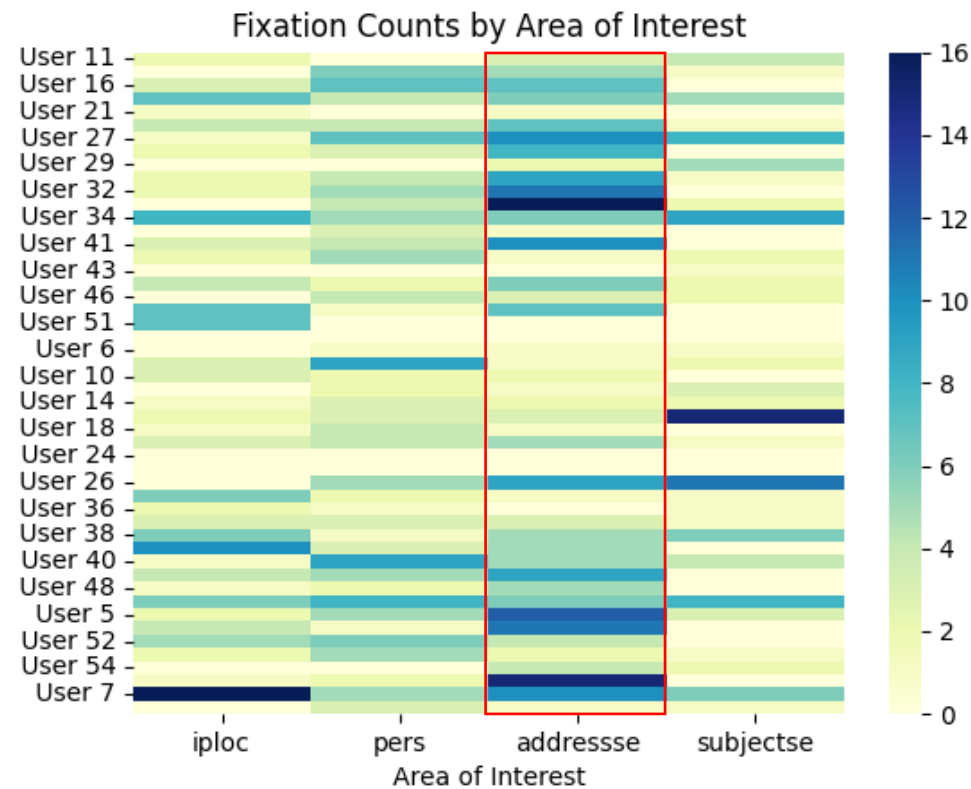


Figure 5.3. – Heatmap for the 4 smaller areas of interest.

To get a clearer view of these results we can look at the table in Figure 5.4. where we can see the average of fixation counts of all users per area of interest. If we ignore the contentse area of interest for the reasons stated above, we can observe that **the addressse area has the highest average of fixation count** of 4.84 fixations per user.

	Iploc	contentse	pers	addressse	subjectse
AVARAGE:	2.74	16.62	3.24	4.84	2.24

Figure 5.4. – Average fixation counts per area of interest

The second most examined area is the personalization area where the recipient’s name is mentioned, then it’s the iploc area with the IP address and location of the supposed breach,

while **the least examined area is the subject line** of the email with an average of 2.24 fixations.

## 5.2 Fixation Count Statistics by Field Dependency Group.

We ran the same tests per field dependency group to find out whether there are any **significant differences** in the areas that users examine based on their cognitive style. You can see the average comparison in the table in Figure 5.5.. You can see that for both groups the address area is the most popular, and that there are no statistically significant differences between cognitive style groups.

For the field-dependent group, 34.5% of fixations were made on the addressse area, while for the field-independent it was 39.9.%.

	lploc	pers	addressse	subjectse	SUM
<b>AVARAGE</b>					
<b>FD:</b>	3.04	3.23	4.65	2.54	13.46
<b>FI:</b>	2.42	3.25	5.04	1.91	12.62
<b>ALL:</b>	2.74	3.24	4.84	2.24	13.06

Figure 5.5. – Average fixation counts per area of interest and cognitive style.

**Notice that there is no significant difference between the total fixation count average either.**

You can see that **there is no significant difference in what the two groups examine** more clearly in the following chart in Figure 5.6.

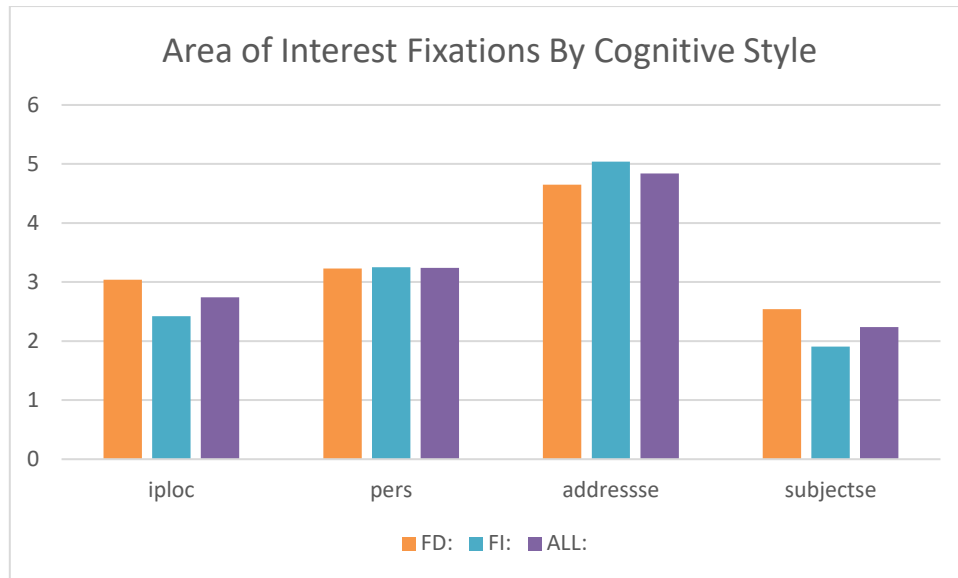


Figure 5.6. – Comparison of area fixation counts between cognitive styles.

We can see the proof of there being **no significant statistical difference in what users examine depending on their cognitive style** when we run a Mann-Whitney U test (Appendix 1 – Script 4). We got the following results in Figure 5.7. indicating no significant statistical differences for any area of interest. [41]

Area of Interest	U-statistic	p-value	Is significant
iploc	334.5	0.66	No
pers	310.0	0.98	No
addressse	293.5	0.73	No
subjectse	319.0	0.90	No

Figure 5.7. – Mann-Whitney U Test

### 5.3 Fixation Count Statistics by Sex

We also ran the same tests after dividing the participants by sex, the average fixation counts per area of interest for each sex can be seen in the table below in Figure 5.8.



SEX/AOI	iploc	pers	addresse	subjectse	SUM
F	2.392857	2.821429	4.428571	2	11.64286
M	3.181818	3.772727	5.363636	2.545455	14.86364

Figure 5.8 – Area of interest fixation average per sex

As we can see in the table above and the graph in Figure 5.9. although **there is no difference between what the sexes examine**, there may be a difference between how many times they fixate on those areas in general.

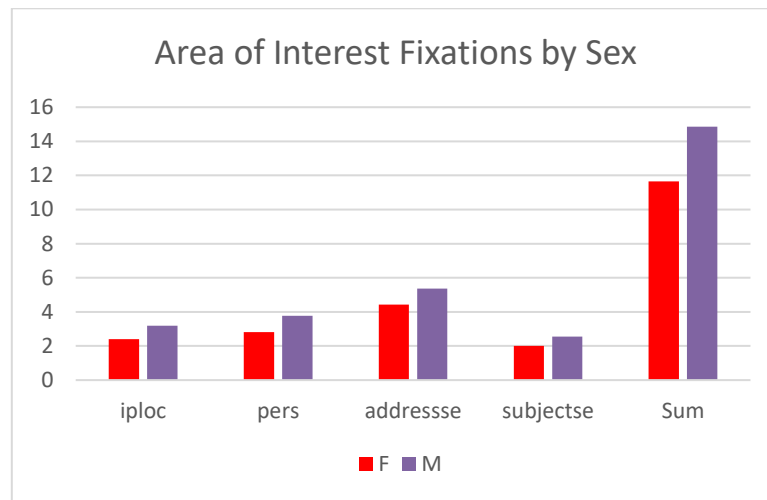


Figure 5.9. Area of interest fixations by sex.

We can prove that there is **no significant statistical difference in what users examine depending on their sex** when we run a Mann-Whitney U test (Appendix 1 – Script 4). We got the following results in Figure 5.10. indicating no significant statistical differences for any area of interest.

Area of Interest	U-statistic	p-value	Is significant
iploc	270.0	0.46	No
pers	230.5	0.13	No
addresse	250.5	0.26	No
subjectse	290.5	0.73	No

Figure 5.10. - Mann-Whitney U Test per AOI

However, as mentioned earlier, **the difference in fixation counts between males and females is statistically significant**, regardless of which areas of interest they examined. The average fixations that a male does is 14.86, while females fixate on those areas of interest an average of 11.64 times.

According to the Mann-Whitney U test that we ran (Appendix 1 - Script 6), this is a statistically significant difference, and the results can be seen below in Figure 5.11.

Area of Interest	U-statistic	p-value	Is significant
All areas	616.0	2.77e-12	Yes

Figure 5.11. - Mann-Whitney U Test all AOIs

**Males obviously did more fixations in general.**

## **6. Eye gaze behavior and susceptibility.**

### **6.1 T-Test and Revisit count distribution between tricked individuals and individuals that recognized the attack.**

The t test is a common statistical test that can be **used to compare the means of two groups**. It's used to determine whether 2 groups of interest differ from each other regarding the tested parameters.

Next, **we wanted to determine whether the areas of interest that an individual revisits is a reliable predictor** to whether they would be susceptible to the phishing attack, and in turn click the provided link or not.

For this we ran a **T-Test using a Python script** (Appendix 1 – script 2) with the purpose of determining if there is a statistically significant difference between the revisits per area of interest, and whether an individual would be tricked.

**The null hypothesis (H0) for this T-Test** is: There is no statistically significant difference between the revisits per area of interest and an individual's susceptibility to the phishing attack.

**The alternative hypothesis (Ha) for this T-Test** is: There is a statistically significant difference between the revisits per area of interest and an individual's susceptibility to the phishing attack.

The results of the T-Test can be seen in figure 6.1 below.

Area of Interest	t-statistic	p-value	Significant
Pers	1.16	0.25	No
Iploc	0.70	0.48	No
Contentse	1.13	0.26	No
Adresse	<b>2.52</b>	<b>0.01</b>	<b>Yes</b>
Subjectse	-0.63	0.53	No

Figure 6.1 – T-test results

In a t-test, **the t-statistic measures the difference between the sample mean and the null hypothesis mean** in units of standard error. The higher the absolute value of the t-statistic, the more likely it is that the difference between the means is statistically significant. Generally any t-value with an absolute value of 2 or higher is acceptable. The higher the t-value, the more the confidence we have in the coefficient as a predictor.

On the other hand, **the p-value measures the probability of observing the observed or more extreme difference between the means, assuming the null hypothesis is true.** A low p-value (typically less than 0.05) indicates that the observed difference is statistically significant, meaning that the null hypothesis can be rejected in favor of the alternative hypothesis that there is a true difference between the means. A high p-value (typically greater than 0.05) indicates that the observed difference is not statistically significant, meaning that the null hypothesis cannot be rejected.

We can see that **there is a statistically significant difference in the revisit count between people that were tricked and people that recognized the attack.**

However, there is no statistically significant difference between field-dependent and independent people when it comes to the revisits on the address, or any other area of interest (Appendix 2 - Test 2).

**The same results and patterns can be found when performing the t-test on the fixation count and the time viewed (in seconds) for the users.**

**The t-test however, assumes that the data is normally distributed.** Using a python script (Appendix 1 – script 3) we tested our revisit data for normality and found that **only contentse was normal.** This is why we decided to use the non-parametric Mann-Whitney U test to confirm our results. The normality test results can be seen in the following histogram in Figure 6.2 and table in Figure 6.3.

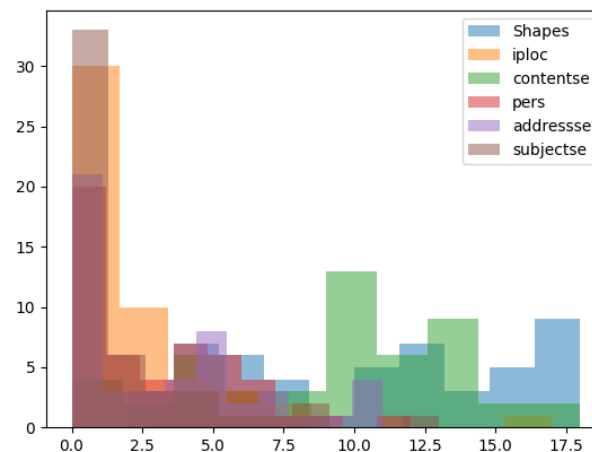


Figure 6.2 - Normality test results

Area of interest	W	p	Normally distributed
<b>iploc</b>	0.669	2.46e-09	no
<b>contentse</b>	0.962	0.109	yes
<b>pers</b>	0.888	0.0002	no
<b>adresse</b>	0.872	6.64e-05	no
<b>subjectse</b>	0.647	1.04e-09	no

Figure 6.3 - Normality test results

- **iploc**:  $W=0.669$ ,  $p<0.05$ : The p-value is less than 0.05, indicating that the null hypothesis of normality is rejected. The W value is also lower than 1, indicating that the data is not perfectly normally distributed.
- **contentse**:  $W=0.962$ ,  $p>0.05$ : The p-value is greater than 0.05, indicating that there is insufficient evidence to reject the null hypothesis of normality. The W value is also relatively close to 1, suggesting that the data is somewhat normally distributed.
- **pers**:  $W=0.888$ ,  $p<0.05$ : The p-value is less than 0.05, indicating that the null hypothesis of normality is rejected. The W value is also lower than 1, suggesting that the data is not perfectly normally distributed.
- **adresse**:  $W=0.872$ ,  $p<0.05$ : The p-value is less than 0.05, indicating that the null hypothesis of normality is rejected. The W value is also lower than 1, suggesting that the data is not perfectly normally distributed.
- **subjectse**:  $W=0.647$ ,  $p<0.05$ : The p-value is less than 0.05, indicating that the null hypothesis of normality is rejected. The W value is also lower than 1, suggesting that the data is not perfectly normally distributed.

Therefore, none of the areas of interest are normally distributed at a satisfactory level, which is why we move forward to confirm our finding using the Mann-Whitney U test.

## 6.2 The Mann-Whitney U test and area of interest revisits.

The Mann-Whitney U test is essentially a non-parametric equivalent of the two-sample t-test, and is used to compare the distribution of two independent samples. It was chosen

because it does not require an identical sample size for the 2 groups like the Wilcoxon Signed-Rank test or normalized data like the t-test as it is non-parametric.

To run the Mann-Whitney U test we used a python script (Appendix 2- script 4) where our 2 groups were those who were tricked by the email and clicked the link, and those who were not tricked.

**The Mann-Whitney U test has indeed confirmed our findings of the t-test as you can see below in Figure 6.4.**

Area of interest	U statistic	p-value	significant
Iploc	338.5	0.576	no
Contetse	358.5	0.353	no
Pers	381.5	0.164	no
Addressse	<b>439.5</b>	<b>0.011</b>	<b>yes</b>
subjectse	254.5	0.222	no

Figure 6.4 – results of the Mann-Whitney U test

The U statistic in the Mann-Whitney U test represents the rank-sum statistic. It is the smaller of the two rank-sums calculated between two samples. The rank-sum is the sum of the ranks of the observations in one sample. **In other words, the U statistic is the sum of the ranks of the observations in the smaller sample, adjusted by a correction factor to account for ties.**

In the Mann-Whitney U test, the sum of the ranks of the observations is a statistic that measures the degree of difference between two groups. The ranks are used to assign a unique order to the observations within each group. **The ranks represent the relative positions of the observations within the group, with the lowest observation assigned a rank of 1, the second lowest a rank of 2, and so on.** Ties in the data are assigned the average rank of their corresponding values.

In our study, **the observations are the fixation counts for each area of interest, grouped by whether an individual was tricked or not by the phishing email.** The ranks are assigned separately to the observations within each group, and the sum of the ranks is used to calculate the Mann-Whitney U statistic.

The U statistic is used to compare the ranks of two independent samples and is a non-parametric test for determining if there is a significant difference between the two groups. **A smaller U value indicates that one sample is generally higher ranked than the other,** while a larger U value indicates that the two samples are more similar in terms of their rankings.

In our case, the only statistically important differentiation between the rank-sums is for addresse due to the p-value that was explained in the previous section.

### 6.3 Histograms and explanations

**To understand this difference, let's look at the histogram** of the addresse are of interest in figure 6.5 below where 'yes' means the individual was tricked and clicked the link, and 'no' means that they were not.

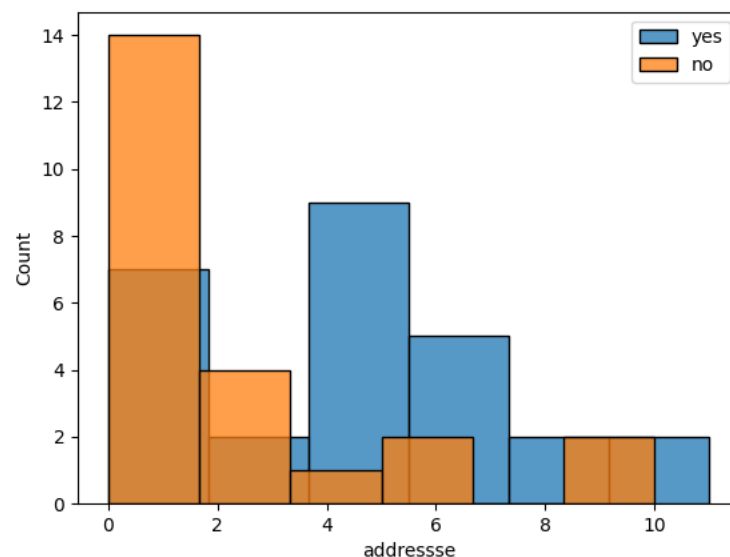


Figure 6.5 – Histogram of fixation count on addresses

To see the contrast with histograms that posed no significant statistical difference, see Figure 6.6 below with the other 4 histograms.

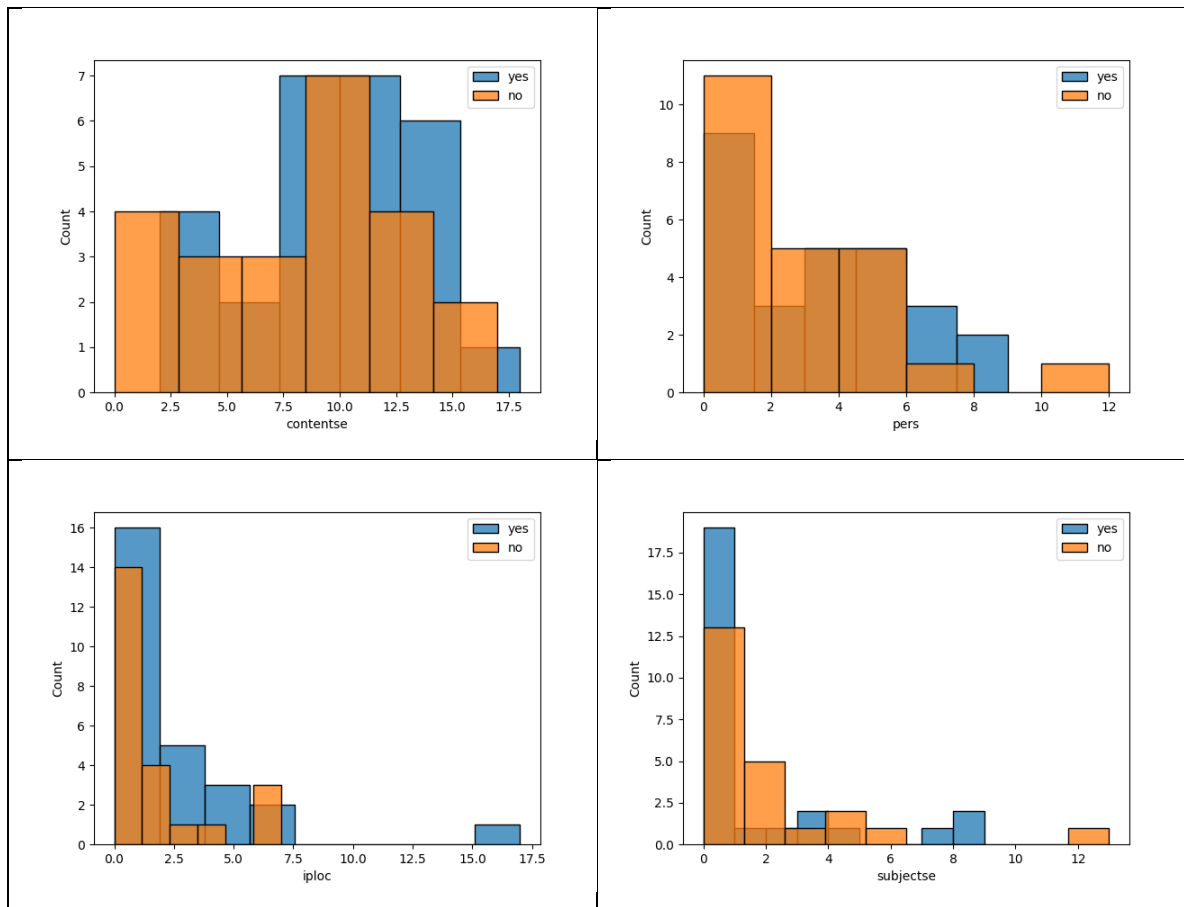


Figure 6.6 – The histograms for the rest of the AOIs.

This histogram tells us that **the more an individual revisits the address from which the email was sent, the more likely they are to trust the email and click the link.** In the context of our experiment, the address was security@facebook.com with the display name of Facebook Security. **This apparently gave confidence to users to trust the email.**



## 7. Think Aloud Protocol and Explanations

In order to explain the observations made so far, but also to examine if there is a dependency between the participants' thought processes and behavior, **we used a large language model (GPT 4) in order to conduct a sentiment analysis and categorize the participants comments in the following categories:**

- Sentiment.
- Executed a google search.
- Had similar past experiences.
- Relied on outside sources.
- Was willing to ask for help.
- Was suspicious.

A researcher cross-checked the categories and sentiment suggested by GPT-4. Some examples of sentiments can be seen in the following quotes:

*"It's a login from Paphos, I haven't been there, I won't click the link if I'm not sure the email is legit"* – User 11: categorized as cautious.

*"I clicked link only out of curiosity, nothing like this has happened before and I know it's a phishing attack"* User 8: categorized as curious.

On the Kaspersky email *"I clicked the link because I was scared"* – User 31: categorized as fearful.

We ran the **Chi-square test for independence** to examine the relationship between the aforementioned parameters extracted from the think aloud protocol and the following parameters:

- Spoofed Email actions.
- Kaspersky Email actions.
- Field Dependency.
- Sex.

The test used in the script (Appendix 1 – Script 7) is the Chi-square test for independence. **The Chi-square test is a statistical method used to determine if there is a significant**

**association between two categorical variables** in a sample. In other words, it checks whether the variables are related or independent.

The Chi-square value represents the test statistic, which is calculated based on the differences between the observed frequencies (counts) in the data and the expected frequencies (counts) that would occur if the variables were independent. **A higher Chi-square value suggests a greater difference between the observed and expected frequencies**, implying that the variables are more likely to be related.

The results for all the relationships can be seen in the table in Figure 7.1.

Chi-Square Results	Spoofed Actions	Kaspersky Actions	Field Dependency	Sex
<b>Sentiment</b>				
Chi-square:	64.08	55.61	16.99	12.22
P-value:	0.00	0.00	0.07	0.27
Significant:	Yes	Yes	Maybe	No
<b>Google Search</b>				
Chi-square:	5.10	5.53	0.00	0.00
P-value:	0.08	0.06	1.00	1.00
Significant:	Maybe	Maybe	No	No
<b>Had Past Experiences</b>				
Chi-square:	2.96	6.93		0.01
P-value:	0.22	0.03	Unrelated	0.94
Significant:	No	Yes		No
<b>Outside Sources</b>				
Chi-square:	6.37	4.14	1.64	0.63
P-value:	0.04	0.12	0.19	0.42
Significant:	Yes	No	No	No
<b>Ask for help</b>				
Chi-square:	6.37	4.14	0.26	0.63
P-value:	0.04	0.13	0.61	0.42
Significant:	Yes	No	No	No
<b>Suspicious</b>				

Chi-square:	9.51	6.70	0.00	0.00
P-value:	0.01	0.03	1.00	1.00
Significant:	Yes	Yes	No	No

Figure 7.1 -Chi-Square analysis results

**The Chi-square value tells us how different the relationship we see in our data is from what we would expect if there were no relationship at all. A higher value means that it's more likely that the two things are related.**

### **7.1 Relationship between Field Dependency and Sentiment.**

Although the relationship between field dependency and Sentiment is not statistically significant at the conventional 0.05 significance level, the p-value is 0.0745, which is close enough compared to other parameter p-values and it is definitely worth mentioning. The p-value of 0.0745 indicates that there's a 7.45% chance of observing the data if the null hypothesis (i.e., no relationship between the two variables) were true. This p-value suggests that **there might be a weak association between an individual's field dependency and sentiment while interacting with the emails.**

Using the same script as before, we produced the stacked bar chart for field dependency and sentiment in the hopes of understanding this relationship better, the results can be seen in Figure 7.2.

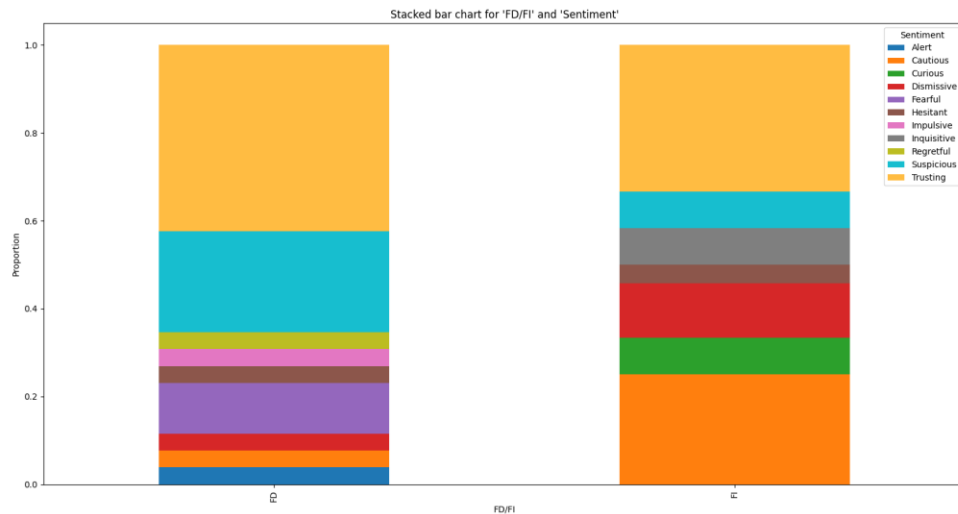


Figure 7.2 - Field dependency and sentiment stacked bar chart.

We can see that only some FI individuals exhibited curiosity, while they were significantly more cautious overall. Additionally, more FI participants tended to be dismissive of the emails. In contrast, only FD individuals displayed fearfulness. Although only some FD participants were impulsive (and no FI participants were), a greater number of them demonstrated suspicion when encountering the emails. These behavioral differences between FI and FD individuals provide valuable insights into their respective approaches to potentially phishing emails.

## 7.2 Relationship between sentiment and actions taken.

Now that we discussed the relationship between field-dependency and sentiment, it's time to look at why we care about this relationship. We can see from the table in Figure 7.1 that **there is a statistically significant dependence between sentiment and the actions taken on the emails**. It seems like it is the one most important factor that determines the way that people will behave and handle the emails.

### 7.2.1 Sentiment for Spoofed email

Let's take a closer look at the breakdown of sentiment per action taken for each email, first starting with the spoofed email in Figure 7.3.

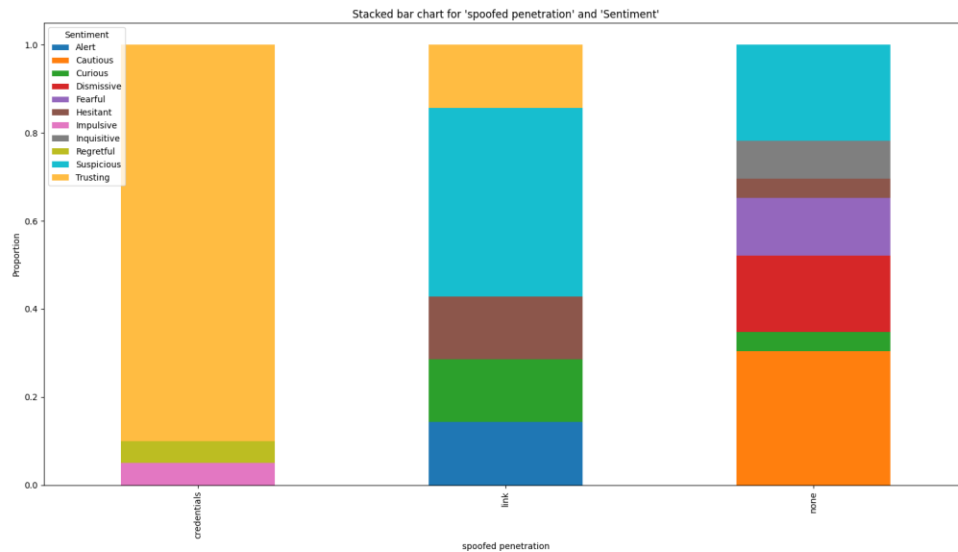


Figure 7.3 -Actions on spoofed and sentiment stacked bar chart.

**We can immediately see that participants who provided their credentials were more trusting, while people that made no actions were mainly cautious. We can also notice that some people that did no actions were fearful. Curiosity seems to drive people to click the link. Suspicion seems to stop some people from clicking the link, and some people from providing their credentials, but no suspicious person provided their credentials. Impulsiveness and trustfulness seem to be the sentiments that lead to most credentials provided. Alertness also stops people from providing their credentials after they've clicked the link.**

### 7.2.2 Sentiment for Kaspersky email

Now let's look at the sentiment breakdown per action taken for the Kaspersky email in Figure 7.4.

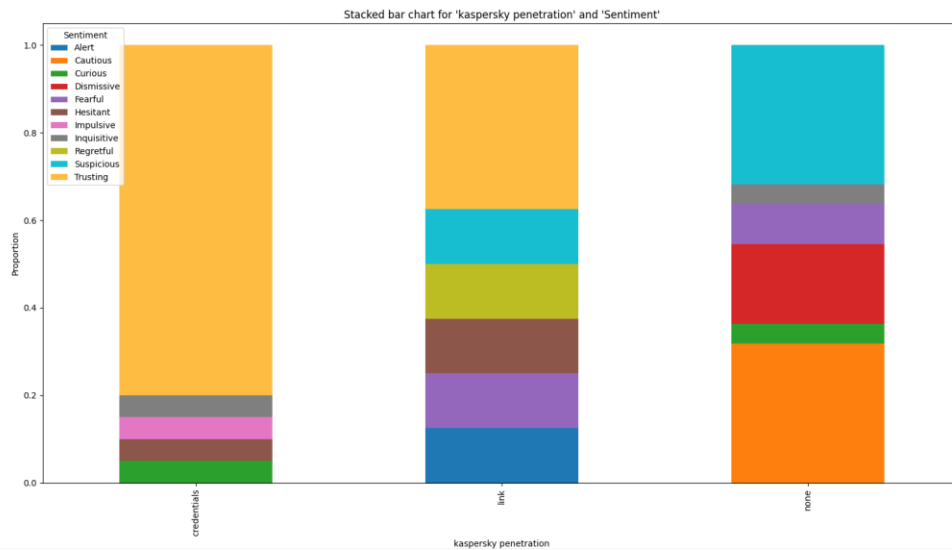


Figure 7.4 -Actions on Kaspersky and sentiment stacked bar chart.

**The most noteworthy observation is that fearful people tended to sacrum to the emotional response that the Kaspersky email caused, making them click the link, but not provide their credentials.**

Other than that, we can see roughly the same patterns as the spoofed email.

### 7.3 Other factors

It seems like relying on outside factors, using google searches and being suspicious of the emails have an effect on whether someone interacts with the email or not, however, this is to be expected, but there seems to be **no relationship between these actions and an individual's field dependency or sex.**

All the observations that we can make are:

1. Using google seems to have a weak relationship with the actions taken in both emails.
2. Having experience seems to have a statistically significant impact on the actions taken in the Kaspersky email, but not the more convincing spoofed email. (Figure 7.5)

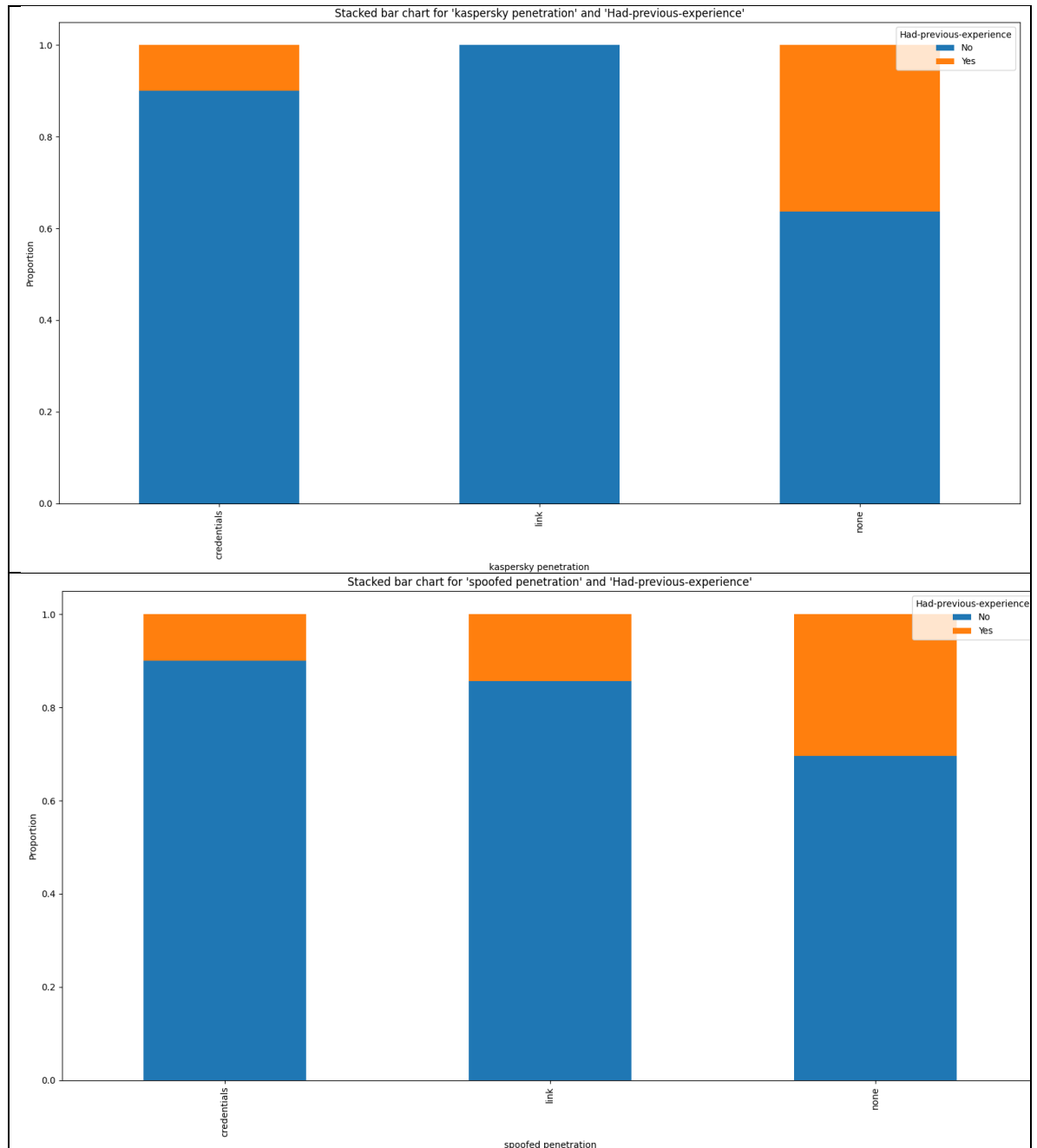


Figure 7.5 -Past experience and actions taken.

3. Relying on outside sources and willingness to ask for help seems to have a statistically significant effect on the actions taken regarding the spoofed email but not the Kaspersky email.
4. The suspicious nature of participants seems to have a statistically significant effect on how participants react to the emails.

It is worth mentioning that although **not statistically significant**, more field-independent people relied on outside sources, and that females were more willing to ask for help as seen in Figure 7.6 and Figure 7.7.

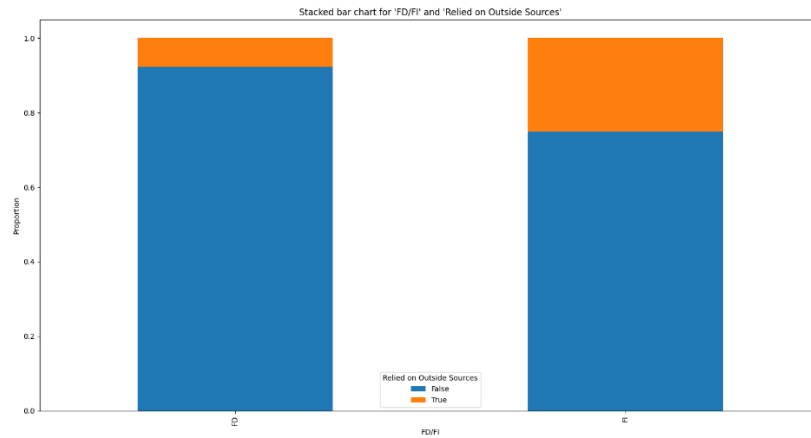


Figure 7.6 – FD/FI and outside sources.

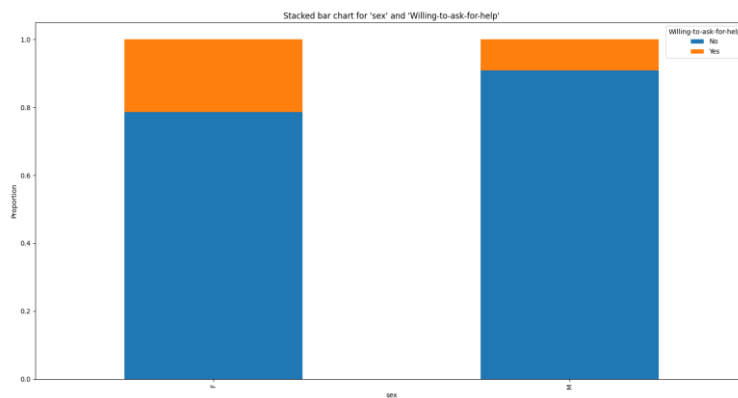


Figure 7.7 – Sex and willingness to ask for help.



## 8. Conclusions

### 8.1 Summary of Findings

Through this experiment, we were able to confirm some of our theories, reject others, and find new knowledge in the gathered data. The main findings of this study are 5 solid points.

**(i) Field-independent people indeed seem better at recognizing phishing attacks**, with an 18% improvement rate compared to field-dependent people: This could be because FI individuals tend to be more focused on individual elements rather than the overall context, which allows them to identify inconsistencies or suspicious elements in phishing emails.

**(ii) People tend to examine the email address** from which an email was received **more often than other areas**; this is an important finding when you consider that we also found that **(iii) users who examine the email address more often were also more likely to be tricked by the phishing attack**, keep in mind, in our case the email address was the same one as Facebook's so it may have provided a false sense of security: This emphasizes the importance of incorporating other indicators of phishing, such as inconsistencies in the email content or design, in addition to the sender's email address.

In general, there is no statistically significant difference between which areas FD and FI people examine: This highlights that cognitive style influences how individuals process and interpret the information they examine, rather than the specific areas they focus on.

However, we did find that **(iv) males tend to make more fixations in total than females**: The finding that males made more fixations than females could be attributed to various factors, such as differences in visual processing or decision-making strategies between the genders. However, this study does not provide enough evidence to determine the underlying cause of this difference.

Finally (v), by studying people's comments from the think aloud protocol, we were able to find the following patterns:

**1. More field-independent people were cautious, and cautious people in our experiment never interacted with the emails:** Being more detail-oriented, field-independent individuals might notice small discrepancies in the emails, such as unusual wording, inconsistencies in design, or slight deviations from the expected format. This heightened attention to detail could cause them to be more cautious when interacting with emails, particularly those with potentially suspicious content.

**2. More field-dependent people were suspicious, and some suspicious people clicked the link, but they never provide credentials:** Field-dependent individuals might be more prone to being suspicious, as they might sense that something is off in the email but struggle to pinpoint the exact inconsistency due to their focus on the broader context.

Suspicious individuals, while having a general sense of doubt about the emails, might still click on the links out of curiosity or to gather more information to confirm their suspicions. Their suspicion might not be strong enough to prevent them from clicking the links, but it is still sufficient to make them wary of providing their personal credentials.

**3. Only FD people were fearful and fearful people only clicked the link if the content of the email was threatening. If not, they didn't interact with the email:**

The reason why only FD people were identified as fearful in this study could be because their cognitive style makes them more vulnerable to the emotional triggers used in phishing emails. This susceptibility might lead them to experience a heightened sense of fear when confronted with a potentially threatening situation.

Fearful FD individuals might click on the link only if the email content was threatening because the fear-inducing tactics used in the email could create a sense of urgency or concern. In such cases, the individuals might feel compelled to take immediate action, such as clicking the link, in order to alleviate their fear or to address the perceived threat. The emotional response elicited by the threatening email content may override their caution or suspicion, leading them to engage with the phishing attack to a certain extent. However, it is important to note that the fearful FD

individuals did not go as far as providing their credentials, which might indicate a residual level of wariness or awareness of potential risks.

4. **Only FI people were inquisitive and inquisitive people never interacted with the email:** The reason why only FI people were identified as inquisitive in this study could be because their cognitive style enables them to more effectively scrutinize and evaluate the information presented in the phishing emails. As a result, they might be more prone to questioning the authenticity of the email and examining its elements critically.

**Other statistically significant findings are:**

1. **Having experience seems to have a statistically significant impact on the actions taken in the Kaspersky email, but not the more convincing spoofed email:** This highlights the need for continuous education and training in recognizing phishing attacks. (Figure 7.5)
2. **Relying on outside sources and willingness to ask for help seems to have a statistically significant effect on the actions taken regarding the spoofed email but not the Kaspersky email:** The Kaspersky email was less convincing, which might have made it easier for participants to recognize it as a phishing attempt without needing external input. In this case, their prior knowledge or experience could be sufficient to identify the phishing email, reducing the need to rely on outside sources or ask for help.

## **8.2 Implications**

This study has several important implications for enhancing online security and minimizing the risk of phishing attacks. By identifying the factors that contribute to users' susceptibility, our research can inform the development of targeted strategies for individuals with varying cognitive styles and visual behaviors.

Firstly, the findings suggest that field-independent individuals are better equipped to recognize phishing attacks. As such, it **may be beneficial to further investigate the factors that make field-independent people more likely to identify an attack** in order to develop better targeted training programs and tools to help people be safer online.

Secondly, **our research highlights the need for improved email security features**. As users tend to focus on the sender's email address, in our case the email address was the same as the legitimate one, therefore misguiding the users; incorporating additional visual cues or warnings related to potentially suspicious addresses could reduce the likelihood of successful phishing attacks.

Additionally, our study reveals that the strategies and concerns of field-dependent and field-independent individuals differ significantly. **This insight can be utilized to design more effective interventions that cater to these unique cognitive styles**, taking into account their specific patterns of thought and behavior.

Finally, our data imply that **gender may influence visual activity during email assessment**. This data might be utilized to further adapt awareness campaigns and training activities, ensuring that they resonate with various groups and meet their individual needs.

By taking these implications into account, we may help to support ongoing efforts to safeguard users from phishing attacks and promote a safer online environment for everybody.

### **8.3 Limitations**

While our research gives important insights into the factors that determine sensitivity to phishing emails, it is critical to acknowledge the limitations of our findings. Recognizing these limits not only contextualizes our findings, but also aids in identifying future study and development opportunities.

1. **Sample size and demographics:** Our study involved a relatively small sample size of 50 participants, which may limit the generalizability of our findings. Additionally, the demographics of our participants might not fully represent the broader population, potentially affecting the external validity of our results. Our demographics lack in the variety of age as we have a convenient sample of people aged 18-25, furthermore, only 2 neurodivergent individuals took part in the study.
2. **Controlled environment:** The phishing attack experiment was conducted in a controlled environment, which may not accurately replicate real-world conditions. Participants might behave differently when faced with actual phishing attacks in their day-to-day online activities.
3. **Limited scope of phishing attacks:** In our study, we used two specific phishing attack scenarios based on a fake Facebook email. Future research could benefit from exploring a wider range of phishing attack scenarios, incorporating different platforms, email formats, and strategies to better understand users' susceptibility to various types of phishing attempts.
4. **Think-aloud protocol bias:** The think-aloud protocol may have introduced bias into the study, as participants might have been more cautious or attentive while articulating their thoughts. This could potentially affect their behavior during the experiment.

By addressing these limitations in future research, we can continue to refine our understanding of the factors that contribute to users' vulnerability to phishing attacks and develop more effective strategies for combating this prevalent online security threat.

## **8.4 Future Research Directions:**

Building on our study's findings and limitations, several avenues for future research can be explored to further our understanding of phishing attack susceptibility and to develop more effective preventative measures. Some potential directions include:

1. **Expanding the sample size and demographics:** Conducting research with larger and more diverse participant samples, especially age wise, can help to increase the generalizability and external validity of the findings, providing a more comprehensive understanding of phishing attack susceptibility across different populations.
2. **Investigating various phishing attack scenarios:** Exploring a broader range of phishing attack scenarios, including different platforms, email formats, and strategies, can provide insights into how users respond to various types of phishing attempts and which factors contribute to their susceptibility.
3. **Assessing the impact of interventions:** Evaluating the effectiveness of targeted interventions, such as awareness campaigns or training programs tailored to specific cognitive styles or demographics, can help to determine the best strategies for reducing phishing attack susceptibility.
4. **Examining real-world phishing attack situations:** Conducting research in more realistic, uncontrolled settings can help to better understand users' behavior and susceptibility to phishing attacks in their daily online activities. Perhaps on the same sample that has already been categorized based on its field-dependency.

## 8.5 Final Thoughts

Our research has shed light on the elements that make users vulnerable to phishing attempts, highlighting the relevance of cognitive style variations, visual behavior, and gender. Future research can increase our understanding of phishing attack vulnerability and inform the development of successful measures for protecting users and fostering a safer online environment by expanding on these findings and resolving the noted limitations.

Finally, it is critical that we continue to engage in research and awareness projects that enable people to make informed decisions and confidently traverse the digital environment. Knowledge acquired from research like ours can help to build a more secure online environment by reducing the risks connected with phishing attempts and other cyber dangers.

## Bibliography

- [1] Baudrillard, J. (2003). Passwords. Verso.
- [2] Stajano, F. (2011, March). Pico: No more passwords!. In International Workshop on Security Protocols (pp. 49-81). Springer, Berlin, Heidelberg.
- [3] Dedenok, R., Buxton, D., Kaminsky, S., Starikova, A., Team, K., & Grustniy, L. (n.d.). Fake copyright violation notice aimed at stealing Facebook accounts. Daily English Global blogkasperskycom. Retrieved January 3, 2023, from <https://www.kaspersky.com/blog/facebook-account-hijack-through-notes/38571/>
- [4] Gupta, B. B., Tewari, A., Jain, A. K., & Agrawal, D. P. (2017). Fighting against phishing attacks: state of the art and future challenges. *Neural Computing and Applications*, 28(12), 3629-3654.
- [5] Ramzan, Z. (2010). Phishing attacks and countermeasures. *Handbook of information and communication security*, 433-448.
- [6] Hong, J. (2012). The state of phishing attacks. *Communications of the ACM*, 55(1), 74-81.
- [7] Kepner, M. D., & Neimark, E. D. (1984). Test-retest reliability and differential patterns of score change on the Group Embedded Figures Test. *Journal of personality and social psychology*, 46(6), 1405.
- [8] Carter, H., & Loo, R. (1980). Group Embedded-figures Test: psychometric data. *Perceptual and Motor Skills*, 50(1), 32-34.
- [9] Khatib, M., & Hosseinpour, R. M. (2011). On the Validity of the Group Embedded Figure Test (GEFT). *Journal of Language Teaching & Research*, 2(3).
- [10] Jääskeläinen, R. (2010). Think-aloud protocol. *Handbook of translation studies*, 1, 371-374.
- [11] Fonteyn, M. E., Kuipers, B., & Grobe, S. J. (1993). A description of think aloud method and protocol analysis. *Qualitative health research*, 3(4), 430-441.
- [12] Séguinot, C. (1996). Some thoughts about think-aloud protocols. *Target. International Journal of Translation Studies*, 8(1), 75-95.

- [13] Dedenok, R., Buxton, D., Kaminsky, S., Starikova, A., Team, K., & Grustniy, L. (n.d.). Fake copyright violation notice aimed at stealing Facebook accounts. Daily English Global blogkasperskycom. Retrieved January 3, 2023, from <https://www.kaspersky.com/blog/facebook-account-hijack-through-notes/38571/>
- [14] Van Der Heijden, A., & Allodi, L. (2019). Cognitive triaging of phishing attacks. In 28th USENIX Security Symposium (USENIX Security 19) (pp. 1309-1326).
- [15] Musuva, P. M., Getao, K. W., & Chepken, C. K. (2019). A new approach to modelling the effects of cognitive processing and threat detection on phishing susceptibility. *Computers in Human Behavior*, 94, 154-175.
- [16] Wang, J., Herath, T., Chen, R., Vishwanath, A., & 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.
- [17] Hakim, Z. M., Ebner, N. C., Oliveira, D. S., Getz, S. J., Levin, B. E., Lin, T., ... & Wilson, R. C. (2021). The Phishing Email Suspicion Test (PEST) a lab-based task for evaluating the cognitive mechanisms of phishing detection. *Behavior research methods*, 53(3), 1342-1352.
- [18] Vishwanath, A., Herath, T., Chen, R., Wang, J., & 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.
- [19] Goodenough, D. R. (2013). History of the field dependence construct. In *Field dependence in psychological theory, research, and application* (pp. 5-13). Routledge.
- [20] Bertini, M. (2013). Some implications of field dependence for education. *Field dependence in psychological theory, research, and application*, 93-106.
- [21] Jantan, D. H. (2014). Relationship between students' cognitive style (field-dependent and field-independent cognitive styles) with their mathematic achievement in primary school. *International Journal of Humanities Social Sciences and Education (IJHSSE)*, 1(10), 88-93.



- [22] Kozhevnikov, M. (2007). Cognitive styles in the context of modern psychology: toward an integrated framework of cognitive style. *Psychological bulletin*, 133(3), 464.
- [23] Oh, E., & Lim, D. (2005). Cross relationships between cognitive styles and learner variables in online learning environment. *Journal of Interactive Online Learning*, 4(1), 53-66.
- [24] Jakobsson, M., & Myers, S. (Eds.). (2006). *Phishing and countermeasures: understanding the increasing problem of electronic identity theft*. John Wiley & Sons.
- [25] Wright, R., Chakraborty, S., Basoglu, A., & Marett, K. (2010). Where did they go right? Understanding the deception in phishing communications. *Group Decision and Negotiation*, 19, 391-416.
- [26] Stojnic, T., Vatsalan, D., & Arachchilage, N. A. (2021). Phishing email strategies: understanding cybercriminals' strategies of crafting phishing emails. *Security and Privacy*, 4(5), e165.
- [27] Darwish, A., El Zarka, A., & Aloul, F. (2012, December). Towards understanding phishing victims' profile. In *2012 International Conference on Computer Systems and Industrial Informatics* (pp. 1-5). IEEE.
- [28] Mannaru, P., Balasingam, B., Pattipati, K., Sibley, C., & Coyne, J. T. (2017). Performance evaluation of the gazeport GP3 eye tracking device based on pupil dilation. In *Augmented Cognition. Neurocognition and Machine Learning: 11th International Conference, AC 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9-14, 2017, Proceedings, Part I 11* (pp. 166-175). Springer International Publishing.
- [29] Cuve, H. C., Stojanov, J., Roberts-Gaal, X., Catmur, C., & Bird, G. (2022). Validation of Gazeport low-cost eye-tracking and psychophysiology bundle. *Behavior research methods*, 54(2), 1027-1049.
- [30] Brand, J., Diamond, S. G., Thomas, N., & Gilbert-Diamond, D. (2021). Evaluating the data quality of the Gazeport GP3 low-cost eye tracker when used independently by study participants. *Behavior Research Methods*, 53, 1502-1514.
- [31] Lawson, B., & Sharp, R. (2011). *Introducing html5*. New Riders.

- [32] Meyer, E. A. (2006). *CSS: The Definitive Guide: The Definitive Guide*. "O'Reilly Media, Inc."
- [33] Python, W. (2021). Python. *Python Releases Wind*, 24.
- [34] Orquin, J. L., Ashby, N. J., & Clarke, A. D. (2016). Areas of interest as a signal detection problem in behavioral eye-tracking research. *Journal of Behavioral Decision Making*, 29(2-3), 103-115.
- [36] Etikan, I., & Bala, K. (2017). Sampling and sampling methods. *Biometrics & Biostatistics International Journal*, 5(6), 00149.
- [37] Sharma, G. (2017). Pros and cons of different sampling techniques. *International journal of applied research*, 3(7), 749-752.
- [38] Brown, G. H. (1947). A comparison of sampling methods. *Journal of marketing*, 11(4), 331-337.
- [39] Mathew, L., & Bindu, V. R. (2020, March). A review of natural language processing techniques for sentiment analysis using pre-trained models. In *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 340-345). IEEE.
- [40] Belk, M., Fidas, C., Germanakos, P., & Samaras, G. (2015). Do human cognitive differences in information processing affect preference and performance of CAPTCHA?. *International Journal of Human-Computer Studies*, 84, 1-18.
- [41] McKnight, P. E., & Najab, J. (2010). Mann-Whitney U Test. *The Corsini encyclopedia of psychology*, 1-1.

## Appendix 1 – Python scripts for statistical analysis

### Script 1: Formatting Data

This script was used to format the data in my excel sheet.

```
import pandas as pd

# Read the Excel file
df = pd.read_excel('file.xlsx')

# Pivot the data so that the rows are the user names and the columns are the AOI names
pivoted = df.pivot(index='User Name', columns='AOI Name', values='Fixations (#)')

# Sort the pivot table by user name
sorted = pivoted.sort_index()

# Export the pivot table to a new Excel file
sorted.to_excel('sorted.xlsx')
```

### Script 2: T-Test and histogram per area of interest

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import scipy.stats as stats

# Load the data from the xlsx file
data = pd.read_excel('data-revisits.xlsx')

# Plot the distribution of revisits for each area of interest for FD and FI users
for col in ['iploc', 'contentse', 'pers', 'adresse', 'subjectse']:
    sns.histplot(data[data['Tricked'] == 'yes'][col], label='yes')
    sns.histplot(data[data['Tricked'] == 'no'][col], label='no')
    plt.legend()
    plt.show()
```

```

    # Perform a t-test to see if the revisits for each area of interest for FD and FI
    users is significantly different

    t_test_result = stats.ttest_ind(data[data['Tricked'] == 'yes'][col],
                                    data[data['Tricked'] == 'no'][col])

    print("T-test results for {}: t-statistic = {}, p-value = {}".format(col,
t_test_result.statistic, t_test_result.pvalue))

    if t_test_result.pvalue < 0.05:
        print("For", col, "The results are significant with a p-value of",
t_test_result.pvalue)
    else:
        print("For", col, "The results are not significant with a p-value of",
t_test_result.pvalue)

```

### Script 3: Normality test and histogram script

```

import pandas as pd
import matplotlib.pyplot as plt
import scipy.stats as stats

# Load the data from the excel file into a pandas dataframe
df = pd.read_excel("data-revisits.xlsx")

# Plot histograms for all columns except for User Name and Tricked (which are non-
numeric columns)
for column in df.columns[3:]:
    plt.hist(df[column], bins=10, alpha=0.5, label=column)

# Add a legend and display the plot
plt.legend()
plt.show()

# Perform the Shapiro-Wilk test for normality on each column and print the results
for column in df.columns[3:]:
    shapiro_result = stats.shapiro(df[column])
    print(f"Shapiro result for {column}: W = {shapiro_result[0]}, p =
{shapiro_result[1]}")
    if shapiro_result[1] > 0.05:

```

```
        print(f"{column} is normally distributed.\n")
    else:
        print(f"{column} is not normally distributed.\n")
```

#### Script 4: Mann-Whitney U test script for group (tricked/sex/etc.) and area of interest.

```
import pandas as pd
from scipy import stats

# Load data from excel file
df = pd.read_excel("data-revisits.xlsx")

# Select columns of interest
df = df[["Tricked", "iploc", "contentse", "pers", "addressse", "subjectse"]]

# Divide data into two groups: Tricked and Not Tricked
tricked = df[df["Tricked"] == "yes"]
not_tricked = df[df["Tricked"] == "no"]

# Calculate U statistics and p-values for each area of interest
for aoi in ["iploc", "contentse", "pers", "addressse", "subjectse"]:
    stat, p = stats.mannwhitneyu(tricked[aoi], not_tricked[aoi])
    print("Mann-Whitney U test for", aoi)
    print("U statistic:", stat)
    print("p-value:", p)
    if p < 0.05:
        print("There is a significant difference in", aoi, "between the Tricked and Not Tricked groups.")
    else:
        print("There is no significant difference in", aoi, "between the Tricked and Not Tricked groups.")
    print()
```

#### Script 5: Heatmap plotting of fixation count.

```
import pandas as pd
```

```

import seaborn as sns
import matplotlib.pyplot as plt

# Load the Excel sheet
data = pd.read_excel('data-fixations - FI.xlsx', header=[0, 1], index_col=0)

# Get the fixation counts for each area of interest
fixation_counts = data.xs('fixation count', axis=1, level=0)
# Create a heatmap of fixation counts
sns.heatmap(fixation_counts, cmap="YlGnBu")
plt.title("Fixation Counts by Area of Interest")
plt.xlabel("Area of Interest")
plt.ylabel("User")
plt.show()

```

**Script 6: Mann-Whitney U test to determine whether total average difference is statistically significant.**

```

import numpy as np
from scipy.stats import mannwhitneyu

# Enter the data
male_fixation_counts = np.array([14.86] * 22) # 22 males
female_fixation_counts = np.array([11.64] * 28) # 28 females

# Perform Mann-Whitney U test
u_statistic, p_value = mannwhitneyu(male_fixation_counts, female_fixation_counts)

# Print the results
print("Mann-Whitney U statistic:", u_statistic)
print("p-value:", p_value)
if p_value < 0.05:
    print("The difference in fixation counts between males and females is statistically significant.")
else:

```

```
print("The difference in fixation counts between males and females is not statistically significant.")
```

## Script 7: Chi-square test for independence

```
import pandas as pd
from scipy.stats import chi2_contingency
import matplotlib.pyplot as plt

# Read the data from the Excel file
file_path = "think.xlsx"
df = pd.read_excel(file_path)

# Columns to analyze relationships
cols1 = [
    "First email received",
    "spoofed penetration",
    "kaspersky penetration",
    "FD/FI",
    "sex",
]
cols2 = [
    "Sentiment",
    "Mentioned Google",
    "Relied on Past Experiences",
    "Relied on Outside Sources",
    "Was Curious",
    "Was Suspicious",
]

def analyze_relationship(col1, col2):
    # Create a cross-tabulation (contingency table)
    ct = pd.crosstab(df[col1], df[col2])

    # Perform the chi-square test
    chi2, p, _, _ = chi2_contingency(ct)
```

```

# Print the results
print(f"Relationship between '{col1}' and '{col2}':")
print(f"Chi-square value: {chi2:.2f}")
print(f"P-value: {p:.4f}")

# Interpret the p-value
significance_level = 0.05
if p < significance_level or col1=='FD/FI':
    print("The relationship is statistically significant.")
    print_contingency_table(col1, col2)
    calculate_conditional_probabilities(col1, col2)
    plot_stacked_bar_chart(col1, col2)

else:
    print("The relationship is not statistically significant.")
print("\n")

def print_contingency_table(col1, col2):
    ct = pd.crosstab(df[col1], df[col2])
    print(f"Contingency table for '{col1}' and '{col2}':")
    print(ct)
    print("\n")

# Add this function call within the analyze_relationship function for significant
relationships.

def calculate_conditional_probabilities(col1, col2):
    ct = pd.crosstab(df[col1], df[col2], normalize='index')
    print(f"Conditional probabilities for '{col1}' given '{col2}':")
    print(ct)
    print("\n")

# Add this function call within the analyze_relationship function for significant
relationships.

def plot_stacked_bar_chart(col1, col2):

```



```

ct = pd.crosstab(df[col1], df[col2], normalize='index')

# Specify custom colors
colors = [
    '#1f77b4', # Blue
    '#ff7f0e', # Orange
    '#2ca02c', # Green
    '#d62728', # Red
    '#9467bd', # Purple
    '#8c564b', # Brown
    '#e377c2', # Pink
    '#7f7f7f', # Gray
    '#bcbd22', # Olive
    '#17becf', # Teal
    '#ffbc42', # Gold
]

ct.plot.bar(stacked=True, color=colors)
plt.title(f"Stacked bar chart for '{col1}' and '{col2}'")
plt.xlabel(col1)
plt.ylabel('Proportion')
plt.show()

# Add this function call within the analyze_relationship function for significant
relationships.

# Analyze the relationships
for col1 in cols1:
    for col2 in cols2:
        analyze_relationship(col1, col2)

```

## Appendix 2: Additional tests and graphs

### Test 1: Relationship between shapes and spoofed email recognition.

To determine whether there is a relationship between how many shapes a user identifies in the GEFT test and whether they are tricked by the spoofed email, the following Python script was used:

```
import seaborn as sns
import matplotlib.pyplot as plt
import scipy.stats as stats

# Load the data from the xlsx file
data = pd.read_excel('data.xlsx')

# Plot the distribution of shapes identified by tricked and not tricked users
sns.histplot(data[data['Tricked'] == 'yes']['Shapes'], label='yes')
sns.histplot(data[data['Tricked'] == 'no']['Shapes'], label='no')
plt.legend()
plt.show()

# Perform a t-test to see if the number of shapes identified by tricked and not tricked
users is significantly different
t_test_result = stats.ttest_ind(data[data['Tricked'] == 'yes']['Shapes'],
                                data[data['Tricked'] == 'no']['Shapes'])
print("T-test results: t-statistic = {}, p-value = {}".format(t_test_result.statistic,
t_test_result.pvalue))

if t_test_result.pvalue < 0.05:
    print("The results are significant with a p-value of", t_test_result.pvalue)
else:
    print("The results are not significant with a p-value of", t_test_result.pvalue)

# visualization
sns.swarmplot(x='Tricked', y='Shapes', data=data, color='black')
plt.legend()
plt.show()
```

Which produced the following output:

T-test results: t-statistic = -0.3848830376256119, p-value = 0.7020242346977787

The results are not significant with a p-value of 0.7020242346977787

The boxplot, swarm plot, and histogram can be seen below in Figure A2.1 where yes means the individual was tricked and clicked the link of the spoofed email, and no means that they were not.

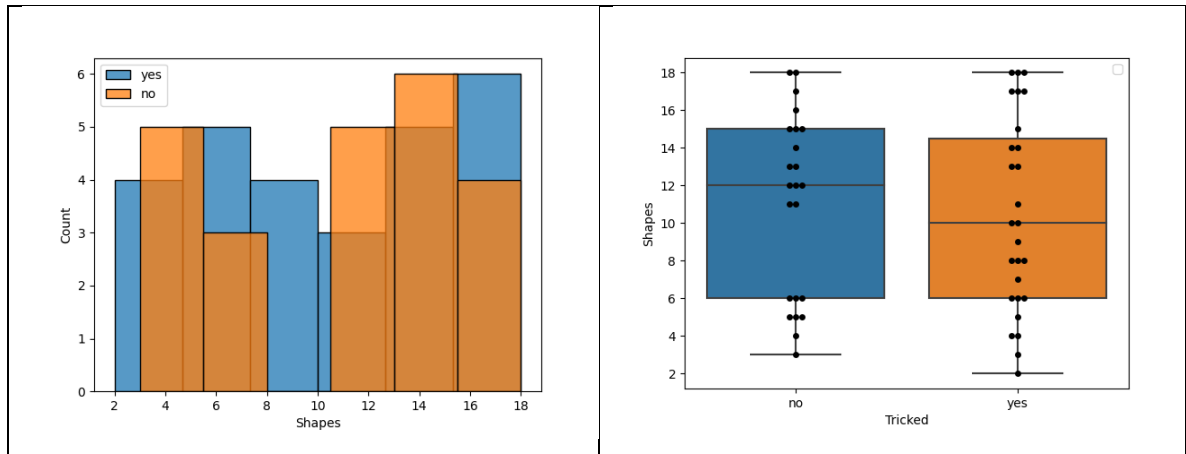


Figure A2.1 – Histogram, Boxplot, Swarmplot of shapes-tricked

## Appendix 3: Consent Form

### What this study is about

The purpose of this study is to understand how people interact with interactive systems. Your participation in this study will help us improve research in usability and security of interactive systems.

### Your participation in this study is voluntary

You can take a break at any time. Just tell the researcher if you need a break. You can leave at any time without giving a reason.

### Information we want to collect

We will watch how you complete an email reading task, track your interaction data, track your eye gaze behavior, we will ask you to respond to some questions and a paper and pencil test, and at the end we will ask you some questions. We will take notes to record your comments and actions.

### How we ensure your privacy

We may publish research papers and reports that may include your comments and actions, but your data will be anonymous. This means your name and identity will not be linked in our research reports to anything you say or do.

### Your consent

Please sign this form showing that you consent to us collecting these data.

I give my consent (please tick all that apply):

- ☒ For people to observe me during the research.
- ☒ For my interaction data to be tracked.
- ☒ For my responses to a questionnaire and a paper and pencil test to be recorded.

If you want to withdraw your consent in the future, contact the persons named below who will destroy any personal data we hold about you. Otherwise, we will delete your personal data after two years.

Taoufik Sousak – [tsousa01@cs.ucy.ac.cy](mailto:tsousa01@cs.ucy.ac.cy)

Marios Belk – [belk@cs.ucy.ac.cy](mailto:belk@cs.ucy.ac.cy)

Participant Name

Signature

Date