Spring 2021

WHERE DO YOU LOOK? RELATING VISUAL ATTENTION TO LEARNING OUTCOMES AND URL PARSING

Niveta Ramkumar
University of New Hampshire, Durham

WHERE DO YOU LOOK? RELATING VISUAL ATTENTION TO LEARNING OUTCOMES AND URL PARSING

BY

NIVETA RAMKUMAR

Bachelor of Engineering in Robotics and Automation, PSG College of Technology, 2015

THESIS

Submitted to the University of New Hampshire
in Partial Fulfillment of
the Requirements for the Degree of

Master of Science

in

Electrical and Computer Engineering

May 2021

This thesis has been examined and approved in partial fulfillment of the requirements for the degree of Master of Science in Electrical and Computer Engineering by:

Thesis Director, Dr. Andrew L. Kun
Professor (Electrical and Computer Engineering)

Dr. MD Shaad Mahmud
Assistant Professor (Electrical and Computer Engineering)

Dr. Caitlin Mills
Assistant Professor (Psychology)

on 05/05/2021.

Original approval signatures are on file with the University of New Hampshire Graduate School.

DEDICATION

I dedicate this thesis to Mr. Aarthiraj Ramani, without whose guidance and encouragement my perusal of higher education would have been impossible.

ACKNOWLEDGMENTS

First and foremost, I have deep gratitude to my research advisor, Dr. Andrew L. Kun, for his generous support, motivation, and guidance during the course of this research. I am very fortunate to have him as my advisor, and I learned a great deal from him.

I want to thank my thesis committee members, Dr. MD Shaad Mahmud and Dr. Caitlin Mills, for their willingness to serve on the committee. Also, I am thankful to Dr. Orit Shaer of Wellesley College and Dr. Sean Smith of Dartmouth College for their collaboration, knowledge, and suggestions during the course of the research work.

I express my sincere gratitude to the Graduate school of the University of New Hampshire for supporting me with the Graduate Fellowship and Summer Teaching Assistant Fellowship (STAF'19). It helped me immensely in completing my research work.

My family is the most precious I have. I am tremendously grateful to my parents, Ramkumar and Kavitha, and to my sister Nikkethana for their love and support. You are my strength and motivation. Lastly, I am greatly indebted to my uncle Govindharajulu, and aunt Jayanthi. Thank you for everything that you have been doing for our family and me.

Special thanks to my close friends Janani and Gokul, who help me find the right perspective and walk through life and its ups and downs along my side.

TABLE OF CONTENTS

NOMENCLATURE

Abbreviations

3D	3 Dimension
ACM	Association for Computing Machinery
AOI	Area of Interest
APWG	Anti-Phishing Working Group
AR	Augmented Reality
GUI	Graphical User-Interface
HCI	Human-Computer Interaction
UNH	University of New Hampshire
URL	Uniform Resource Locators
VR	Virtual Reality

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

ABSTRACT

WHERE DO YOU LOOK? RELATING VISUAL ATTENTION TO LEARNING OUTCOMES AND URL PARSING

by

Niveta Ramkumar

University of New Hampshire, May 2021

Visual behavior provides a dynamic trail of where attention is directed. It is considered the behavioral interface between engagement and gaining information, and researchers have used it for several decades to study user's behavior. This thesis focuses on employing visual attention to understand user's behavior in two contexts: 3D learning and gauging URL safety. Such understanding is valuable for improving interactive tools and interface designs. In the first chapter, we present results from studying learners' visual behavior while engaging with tangible and virtual 3D representations of objects. This is a replication of a recent study, and we extended it using eye tracking. By analyzing the visual behavior, we confirmed the original study results and added more quantitative explanations for the corresponding learning outcomes. Among other things, our results indicated that the users allocate similar visual attention while analyzing virtual and tangible learning material. In the next chapter, we present a user study's outcomes wherein participants are

instructed to classify a set of URLs wearing an eye tracker. Much effort is spent on teaching users how to detect malicious URLs. There has been significantly less focus on understanding exactly how and why users routinely fail to vet URLs properly. This user study aims to fill the void by shedding light on the underlying processes that users employ to gauge the UR L's trustworthiness at the time of scanning. Our findings suggest that users have a cap on the amount of cognitive resources they are willing to expend on vetting a URL. Also, they tend to believe that the presence of "www" in the domain name indicates that the URL is safe.

CHAPTER 1

VISUAL BEHAVIOR WHILE LEARNING WITH VIRTUAL AND TANGIBLE 3D ARTIFACTS

This chapter is based on the publication [P1] presented in the proceedings of the 8th ACM International Symposium on Pervasive Displays (PerDis'19).

1.1 Introduction

Object-oriented learning focuses learner's interaction with physical artifacts as the center of the learning process [23]. It is predominantly used in the disciplines like anthropology, medicine, and archaeology [64]. Recent advancements in technologies like 3D printing, augmented and virtual reality has paved the way for a new realm in object-oriented learning. These technologies allow educators to replicate artifacts eliminating the need for the use of the original artifact. However, such technologies' availability is not enough for fostering object-based learning as we lack the understanding of how these new digital and tactile objects affect the learning process [24]. Pollalis et al. conducted a study recently that sheds light on the learning outcomes when interacting with archaeological artifacts created with three different technology; augmented reality, virtual reality, and 3D fabrication technology. They found that the medium of representation of the object affected the learning outcomes [53]. We replicated and extended their study by adding eye-trackers to provide a more quantitative explanation for these observed differences in the learning outcomes.

Visual attention is considered a window into one's mind [63, 34], and researchers have used eye-trackers for over a decade to get insights into the learning process [37]. By adding eye-trackers to Pollalis et al.'s study, we seek to comprehend how visual attention differs among the three conditions and how these different visual behavior translate to learning outcomes. These insights, in turn, will help us to design and better interactive learning tools.

Figure 1.1. Three different representation of the same scanned museum artifact. a) 3D printed replica, b) model displayed on a screen - sketchfab c) holographic replica - HoloLens

1.2 Overview of Related Research

Object-based learning facilitates interaction with an object to enhance the learner's observational, interpretation, and critical analyzing skills [32]. Also, it is found to be complementing and effectual when paired with regular lectures [23]. Augmented reality is an immersive technology that intends to blend computer-generated information into our physical world in real-time to assist us with additional information. Prior research indicates that AR has a positive effect on the learning process [5, 15]. Here in our study, we used HoloLens, an Optical See-Through (OST) head mount AR device designed by Microsoft. On the other side of the spectrum, we used 3D printed objects that provide intuitive interactions and tactile feedback. Also, it is proven to help learners to form concrete visualization [16].

Pollalis et al. conducted a between-subject experiment to understand the learning outcomes when users engage with three different representations of scanned museum artifacts - tangible 3D printed artifacts, 3D virtual digital models presented on a screen, and holographic artifacts presented on the HoloLens [54]. Figure 1.1 shows the same artifact represented in the above-mentioned forms. They evaluated the user's learning experience in terms of enjoyment, perceived task workload, spatial presence, and learning outcomes. Their results indicated that the 3D printed replicas lacked visual information and did not promote the user's conceptualization and critical

(a) (b)

Figure 1.2. a) Pupil Lab's eye tracker. b) HoloLens with eye-tracker mounted on it

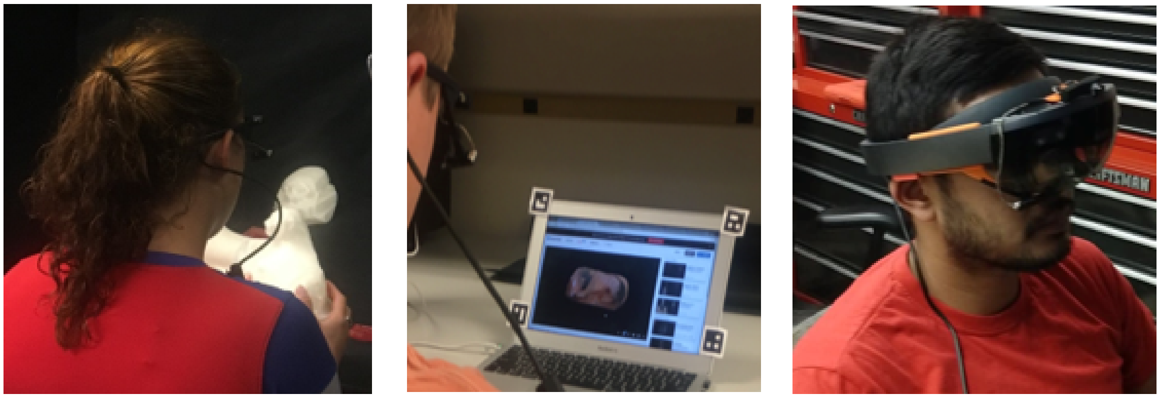
thinking. At the same time, the learning goals accomplished by the other two technologies were comparable. However, further studies are needed to understand how these technologies impact object-based learning processes. Moreover, the roles of the physicals and digital elements of the learning experience remain to be mapped out [4].

Extensive research has been conducted using eye-trackers to deconstruct complex tasks like reading and learning [36, 62, 72, 10, 61, 28, 29]. In this study, we replicated the original experiment by Pollalis et al. [54], while adding eye tracking to each condition. Our goal is to understand users' visual behavior while interact with these object and relate it to the learning outcomes reported in the original study. Most of prior research relating learning and visual behavior uses 2D study material [23, 39, 41, 51, 76]. This is one of the first attempts to use visual behavior in a 3D learning environment.

1.3 Method

1.3.1 Tasks and Procedure

Pollalis et al. developed this experiment based on a task that an archaeology student might encounter in the class. The aim of this task is to practice the learner's critical analysis and observational skills. The task is to select two artifacts from an available inventory of six artifacts, explore them, and answer a corresponding artifact questionnaire. We conducted a between-subjects experiment. Three groups of participants completed the same learning task by interacting with artifacts



(a) (b) (c)

Figure 1.3. Participants performing the task wearing eye trackers a) 3D prints condition, b) sketchfab condition c) HoloLens condition

presented along the real-virtual continuum. Thus, one group used 3D printed replicas, another used 3D virtual replicas on the Sketchfab platform [52], and another used holographic objects through a Microsoft HoloLens application [53].

All three conditions used the same six archaeological artifacts replicated using the scans provided by the British Museum.. The inventory of artifacts, their order in the inventory, their descriptions, and the eye tracking method were consistent throughout all three conditions.

3D Artifacts: These are the same objects used in the original study. It is printed using the MakerBot Replicator 2X and Afinia H800, which are accessible in several educational spaces. Participants wore eye-tracker while interacting with the artifacts. Printed descriptions containing the description of the artifact were kept next to the artifact as shown in Figure 1.1.

Sketchfab: In the second condition, we used an online 3D modeling platform on the desktop. Users could choose any artifact from the inventory and manipulate them using a mouse. Also, they could read the description of the artifact by scrolling down the page.

HoloMuse: For the third condition, participants used HoloMuse, an AR application designed and used by Pollalis et al. The users had to use air gestures to pick and explore the artifact from the inventory. Unlike other conditions, the participant did not have to wear an eye-tracker separately.

Instead, it was already mounted on the HoloLens. Figure 1.3 - c shows a participant wearing the HoloLens and performing the study.

We collected data from 35 participants (10 female). All the participants were given a $10 gift card at the end of the experiment. We dismissed data from 5 participants due to a low eye tracker confidence (<70%). Thus, we report on data from 30 participants (10 participants per condition). All the participants signed the consent forms and completed a pre-task questionnaire regarding prior experience with visual analysis and AR or VR. They were given a brief training about their randomly assigned condition. Before beginning the task, we calibrated the eye-tracker using screen marker calibration [15]. After finish examining each artifact, participants were asked to fill out an object questionnaire. Other than these forms, they had to fill a post-task questionnaire concerning perceived task workload and spatial presence at the end of the study.

1.3.2 Eye tracking

In eye tracking, there are two different types of devices, head mount, and remote eye tracking device. Though remote eye tracking is less intrusive, it is not suitable for our application as the user movement is not restricted in the given space. So, we are using the head mount eye trackers developed by Pupil Labs. The eye tracker has three camera cameras; a world camera and two slide eye cameras. The eye camera records the pupil movement, and the world camera captures the information in front of the user. It uses the dark pupil detection algorithm to detect the eyes, and it calculates users' gaze based on their pupil movements. Then, it maps the calculated gaze onto the video from the world camera.

1.3.3 Data analysis

We used JMP Pro 14 for the statistical analysis. We checked the data's normality and used ANOVA (mean comparison), t-test and Tukey test (post hoc) for normally distributed data. Kruskal-Wallis test and the Wilcoxon test was used for non-normally distributed data. To better understand the distribution of users' visual attention, we defined three areas of interest (AOI); the artifact, the

Visual inspection	*Complex observation*	*Inferences*	*Interpretation*
Shape	Damage	Material	Aesthetic
Color	Detail	Size	Analysis
Texture	Facial Feature	Weight	Context

Table 1.1. Thematic codes used to analyze users' object questionnaires categorized into more general groups

description (where they could read more about the artifact), the manipulation (which they could use to manipulate the artifact). Any visual target other than these three categories were considered as "other surfaces". Due to the intuitive tactile nature of 3D objects, we did not have a separate manipulation AOI for 3D prints condition. We developed image processing algorithms using MATLAB to identify participants' visual targets over the AOIs. These algorithms and the statistical results is discussed in detail in the appendix [P1]).

1.4 Measures & Results

We used the following measures to understand the visual behavior and the learning outcomes in the study.

Time on task: We calculated the overall time spent and time spent on the three AOIs using the timestamps of each recorded data point. This helps in assessing engagement.

Result: The total time spend on completing the task was significantly different among the conditions [X^2(2)=18.7357, p<0.0001] and post-hoc results indicated that users spent more time in the HoloLens condition compared to the other two conditions. How ever, there was no significant difference between in the amount of time spent on the artifact AOI among the three conditions [X^2(2)=3.3626, p=0.1861]. As there was no manipulation AOI for 3D prints condition, we compared the time that participants spent looking at manipulation in the HoloLens and Sketchfab conditions. We found that the time spent on manipulation was significantly higher in HoloLens condition than Sketchfab condition [Z=14.2857,p<0.0002].

Fixations: Similar to the gaze areas of interest, the fixation targets fell into three AOIs; artifacts, descriptions and manipulation. Fixations is maintaining gaze at a particular target for a certain time. We extracted fixations with a minimum duration 100 ms [30, 38, 52] and 1°of dispersion angle [45]. We explored fixation in-terms of fixation rate (fixation count per minute) and overall fixation duration.

> ***Result:*** The overall fixation rate for the task was not significantly different among the conditions [X^2(2)=1.3871, p=0.4998].The fixation rate on the artifact was significantly different and post hoc analysis indicated that fixation rate for the HoloLens condition is lower than the other two conditions [X^2(2)=10.8891, p=0.0043]. In terms of fixation rate on the descriptions, 3D condition has higher fixation rate than the other two conditions [X^2(2)=19.4813, p<0.0001]. Since there was no manipulation AOI for the 3D prints condition, we compared the results between the other two condition. With respect to manipulation, HoloLens condition had a significantly higher fixation rate [Z =8.9195, p=0.0028] and fixation duration [Z=5.4896, p=<0.0006]. Also, we did not find any significant difference in the fixation duration on artifact [X^2(2)=2.2405, p=0.3262] or description AOI [X^2(2)=14.6945, p=0.0191].

Learning outcomes: Participants wrote down a down detailed explanations of the viewed artifacts in the object questionnaire. Content codes were used to demonstrate progress from observation to analysis, and to develop a preliminary review of learning outcomes.To analyze the open-ended questions, we used the same coding-scheme and process as the original study. Table- 1.1 shows the coding scheme used to analyse the open-ended questions. We acted as coders identifying the content codes in the questionnaire responses. Our inter-code reliability >95% and resolved disagreements by consensus.

> ***Result:*** We evaluated the learning outcomes in-terms of total number of content codes and the frequency of mentioning individual code. There was no significant difference in the total number of content codes mentioned in the responses between the conditions [F (2,27) = 1.552, p= 0.8570]. When analyzing the frequency of individual content code, we found

that facial feature and detail showed significant difference. Facial feature content code was used significantly less in the 3D condition than the other conditions. Also, Detail was also mentioned significantly less in the HoloLens than in the 3D prints and Sketchfab condition.

Perceived task workload and spatial presence: We used the NASA TLX questionnaire [26] to measure participants' perceived workload. To measure participants perceived spatial presence, we used a questionnaire based on MEC- SPQ standardized questionnaire [73]. Both these questionnaire is same as the ones used in the original study.

Result: Results suggested that the participants in the 3D prints condition experienced significantly higher effort [F (2, 27) = 5.7838, p= 0.0081]. Also, participants also claimed to think more intensely about the characteristics of the 3D printed artifacts [X^2(2)=6.7055, p=0.0350]. In the HoloLens condition, users felt as if the original artifact was physically present in their environment significantly more than users in the other two conditions [X^2(2)=6.3786, p=0.0412].

1.5 Discussions

In this section we combine the quantitative and qualitative results collected from the study and discuss our main findings:

Comparison with the Original Study: In the original study, they reported on the overall time spent on the task. By adding eye-trackers, we were able to understand how users distribute time between examining the object, manipulating it and reading about it. Our results confirm the results of the original paper in multiple aspects. Both studies found no significant difference in the complexity score of the open question responses about the artifact they viewed. Another common finding was that participants of both studies confirmed the immersive nature of AR.

3D printed artifacts were the only ones to include a sense of touch and to exist physically in the environment. Despite this participants ranked AR the highest when if they "felt as though the original ancient artifact was physically present in [their]environment". This result suggests a new line of investigation, reconsidering Montessori's finding emphasizing the role of physical material in learning [44].

All the visual information is collected during fixation [25, 66, 70]. The rate of fixations on the artifact was lowest for the HoloLens condition while the fixation duration was comparable among other conditions. Both of the studies had significantly fewer mentions of facial features in the open question responses for the HoloLens condition. It is likely that both the fewer mentions of facial features, and the lower fixation rate are related to two innate features of the holograms: they have lower resolution compared to the Sketchfab condition and are not tactile like the 3D prints condition.

Physical Artifact vs. Virtual Artifact: There is not much research done in comparing visual behavior while users engage in tasks involving similar physical and virtual objects [4, 40]. One of the main contributions of our study is making this comparison. The fixation duration and the total time spent on the objects is comparable among all the three conditions. Hence the visual behavior of the virtual and tangible learning material is similar. Also, participants in Hololens condition felt as though the original ancient artifact was physically present in environment. This shows that they appreciate the virtual objects similar to the tangible ones.

Interface and Design Implications: Participants spent significantly higher time looking at the manipulation AOI in the HoloLens condition. By visually inspecting the HoloLens videos, we observed that the higher gaze time and fixation duration for manipulation in the HoloLens condition was due to participant's difficulty performing the dragging gesture. So, a possible design improvement to the interface is to use click on a bar gesture instead of a dragging gestures to manipulate the object.

It is shown that interaction costs can lead to increased reflection on the material [25, 47, 67, 71] Moreover, Marshall claims that the easy manipulation of concrete objects can result in decreased reflection on the learning material. As discussed, the HoloLens condition introduced interaction constraint to the participants. However, the 3D printed condition users reported that they "thought intensely about the characteristics of the ancient artifact" considerably higher than the other conditions. This means that the manipulation effort the users made for the HoloLens condition might have been too much, considering that the manipulation gaze time and fixation duration was highest for this condition.

1.6 Conclusions and Future Work

By adding eye-tracking to the original study, we were able to replicate the original study results and provide a more in-depth understanding of the observed learning outcomes. Our results indicated the following, 1) Users' visual behavior towards virtual learning material is similar to tangible ones. 2) the total time spent on examining the artifact is independent of the medium of presentation and is more dependent on the manipulation options provided in that medium. Especially in the HoloLens condition, users spent significantly higher time on the manipulation, emphasizing the need to explore the role of physical manipulation in learning further. Another aspect that needs to be explored in the future is to analyze collaborative object-based learning and understand how these new technologies aid collaborative object-oriented learning.

CHAPTER 2

EYES ON URLS: RELATING VISUAL BEHAVIOR TO SAFETY DECISIONS

This chapter is based on the publication [P2], Eyes on URLs: relating visual behavior to safety decisions. It is published in the ACM Symposium on Eye Tracking Research and Applications (ETRA '20 Full Papers).

2.1 Introduction

A uniform resource locator (URL) is a string of characters that specifies how and where to retrieve a web resource. People encounter URLs everyday in various environments, such as emails, social media, web browsers, and other applications. On every encounter, they must determine whether these URLs are safe to click on or not. Incorrect evaluation of an URL can expose the individual or their organization to phishing attacks. Users assess the URL based on their knowledge of URL structure, the URL itself, their mood, the context under which the URL is presented, and other determinants. As many factors involve in this determination, nefarious actors exploit the user's inability to correctly interpret the information conveyed in URLs to carry phishing attacks [65].

According to the Anti-Phishing Working Group (APWG), there were 255,065 unique phishing attacks in 2016, and it is increasing day by day [1]. However, there has been significant effort spent on improving the user's ability to vet URLs, e.g., through user training [33, 57] or adoption of browser warnings [56, 42]. And there has also been vital research on understanding how and why users fall for phishing attacks more generally [17, 27].

Researchers have used eye tracking for many decades to obtain a comprehensive understanding of user behavior. To the best of our knowledge, this is one of the first attempts to understand visual attention while users assess URLs. This will help us create a descriptive model of the relationship between URL characteristics and visual behavior. This, in turn, will supplement our understanding of why certain succeed and how to develop better countermeasures and training techniques. We conducted a user study where users were asked to classify URLs as safe or unsafe while wearing an eye tracker. We evaluated the results in terms of total time spent, fixation count, fixation count on various areas of interest (AOIs), time spent on each AOI, pupil dilation, and backtrack fixation count. Our key finding is that users have a cap on the amount of time they spend on assessing a URL, and they depend more upon the `authority` component of URLs than any other component.

2.2 Overview of Related Research

Eye-tracking has been used for over a century in one form or the other to understand human behavior. Researches have used this technology to assess cognitive load [49, 50, 55, 77], reading strategies [10, 61, 28, 29], and design implications [22, 7]. When users make decisions about the safety of URLs, they do so by reading. A great deal of research has been conducted to establish the relationships between reading and eye movements [69, 11, 60, 77]. The most commonly used eye-tracking measures to describe the eye movements while reading are fixations, saccades, regressions, and backtracks [69, 11]. Also, Just and Carpenter proposed the "eye-mind" theory, according to which longer fixation indicates higher comprehensive load.

An interesting class of security challenges deals with ensuring that users properly interpret and operationalize computer security information. Also, a salient challenge within this class involves ensuring that users correctly vet URLs. Many phishing attacks rely on presenting malicious URLs that look safe to unsuspecting victims. Since the early 2000s, researchers have: studied who falls for phishing attacks and why they fall for them, e.g., [17, 68]; the effectiveness of defensive measures [75, 19]; categorized different types of phishing and URL obfuscation techniques, e.g., [18, 48]; examined specific flavors of attacks, e.g., [58]; developed explanatory models of how phishing attacks are conducted, e.g., [31]; proposed phishing detection systems, e.g., [20]; analyzed phishing detection systems, e.g., [2]; and analyzed anti-phishing training systems [35].

Eye-tracking has been used in usable security for training and understanding user behavior. Miyamoto et al. developed a system using an eye tracker to train users to look at the status bar on websites [43]. To understand how users gauge the website's legitimacy, Alsharnouby et al. conducted a user study that uses eye-trackers while participants classify websites as legitimate or illegitimate [3] . While similar in spirit to these earlier studies, our focus in this paper is solely on understanding how users process URLs. We are working at a different level of granularity; we are not concerned with how people visually process websites but rather how people visually process URLs. Working at this finer level of granularity enables us to dissect URLs into different parts and examine how people process each part. We seek to understand what parts of a URL people pay attention to, what parts they don't, when people give up, and how their eyes process different flavors of URLs, amongst other things.

2.2.1 URL Structure

Each URL in our corpus has the form: <scheme>://<authority><rest>. The below table - 2.1 indicates how a URL is disaggregated into these components. Here, we discuss the basics of the URL structure at an appropriate level of granularity to understand our work.

`scheme`	*delims.*	`authority`	`rest`
https	://	www.google.com	/maps/

Table 2.1. Disaggregation of a URL into its three components.

- **Scheme:** This corresponds to a scheme name, which indicates how to interpret the text following the colon [9, 8, 74].
- **Authority:** The `authority` component specifies a subset of the host, port, username, and password [8, 74]. For URLs in our corpus, the `authority` component has either the form `host` or `user@host` where `host` represents the host and `user` represents the username.

- **Rest:** It is a term we use to indicate any characters following the `authority` component. It could be empty or in certain cases it may comprise of the path or queries or fragments.

2.3 Method

2.3.1 URL Corpus and Classification

Our corpus contained a total of 64 URLs that fell into 8 categories. These eight URL categories are presented in Table 2.2 and each category has 8 URLs. This corpus is primarily designed to test the hypothesis discussed in the results section. To maintain uniformity among the categories, every URL has the same `scheme` component (*https*) and the top-level domain (*com*). The categories are defined by the following 4 features: (1) safety, (2) complexity, (3) a leading *www* in the `authority` component, and (4) the attack type for unsafe URLs.

- **Safety:** For safe URLs, we chose fully verified domain names priarily from the top 1,000US websites in the Quantcast Top One Million list. While unsafe URLs contained URLs that did not have a domain name server record, or were spoofed websites.

- **Complexity:** In complexity category, we grouped URLs into simple and complex. This classification is made based on the URL length and URL features. A URL is considered simple if it has an empty path (i.e no rest component) and complex if it has a non-empty rest component.

- **Presence of *www*:** URLs with the *www* attribute begin with *https://www*. URLs with the *non-www* attribute do not.

- **Attack Type:** We chose to explore four conditions for unsafe URLs: Positive, negative, substring and user@host. These attack types does not fully representative of real-world attacks. Our motive here was to explore a variety of conditions that may affect visual behaviors and/or classification. Positive attack types had positive, feel good phrases in the URL. On the other hand, negative attack type had negative or security related words like malware, antivirus. The substring attack uses safe domain name like *https://X.*

Category	Safety	Complexity	www	Attack Type	URL Length
C1	*safe*	*simple*	*www*	N/A	25.0 (4.8)
C2	*safe*	*simple*	*non-www*	N/A	19.8 (2.0)
C3	*safe*	*complex*	*www*	N/A	124.0 (13.2)
C4	*safe*	*complex*	*non-www*	N/A	105.3 (13.5)
C5	*unsafe*	*simple*	*www*	*positive*	28.5 (2.4)
C6	*unsafe*	*simple*	*www*	*negative*	29.3 (3.6)
C7	*unsafe*	*complex*	*non-www*	substring	96.0 (20.4)
C8	*unsafe*	*complex*	*www*	*user@host*	95.0 (17.4)

Table 2.2. Categories of URL corpus.

com and obfuscate it as *https://X.Y.com*. In user@host attack type, `authority` component has form *www.X.com@Y* where *https://www.X.com* is a legitimate URL.

2.3.2 Tasks and Procedure

We conducted a within-subject experiment with 20 participants (3 female). As a part of their coursework requirement, students participated in our study. We discarded data from 4 participants due to technical issues with the eye tracker. Hence, we report on the data from 16 participants (2 female).

The task is to classify the URLs shown on the screen as "Safe" or "Unsafe". Users were shown 64 URLs individually and asked the question, "Is the web address safe to visit?. It was accompanied by two response buttons that read "Safe" and "Unsafe". The aim is to analyze and understand how certain URL properties influence the decision-making and the visual behavior. Mood can affect a person's ability to comprehend text and their judgment [21, 12].To reduce the effect of mood in our study, which in turn improves study replicability, we used neutral mood induction by making the participants watch a neutral video.

The URL corpus was split into two equal-sized sets presented over two sessions, such that four URLs from each category were represented in each set. For each session, the order in which URLs were presented was randomly determined but held fixed for all participants. However, the session

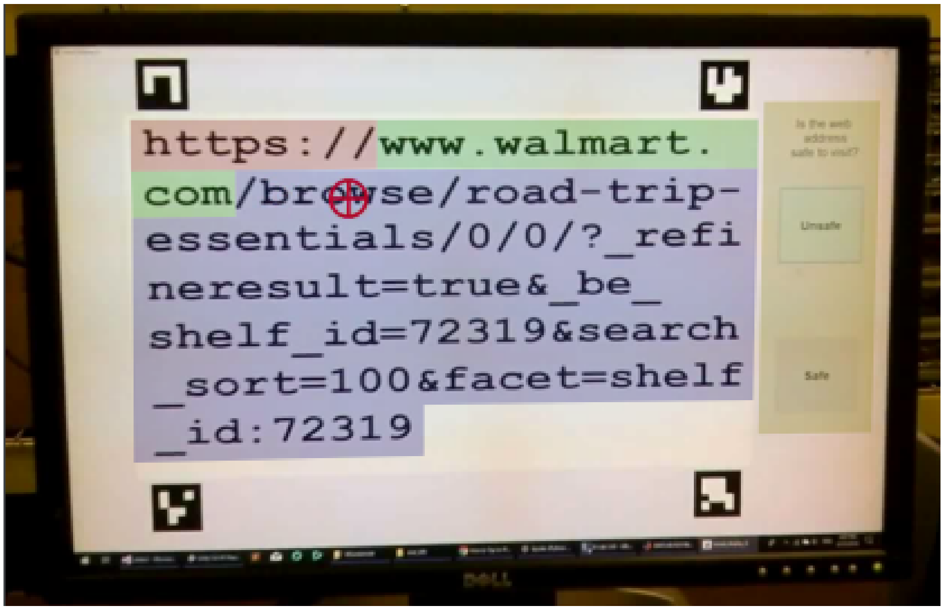


Figure 2.1. A processed frame from the eye tracker video. The red cursor indicates gaze position and the four colored boxes represent four AOIs: the scheme AOI(red), the authority AOI(green), the rest AOI(blue), and the response AOI(yellow).

order alternated between participants. At the end of the session, they were asked to fill a post-task questionnaire containing questions about demography and security knowledge.

2.3.3 Eye tracking & User Interface

To collect the gaze data, we used Dikablis eye trackers. It contains three cameras; two to record the pupil and one to record the scene in front of the user. Calibration is an essential aspect of eye-tracking as it establishes a mathematical relation between the eye and the target being focused. In our study, we used a four-point operator-controlled calibration method in our study [46].

The URLs were presented to the participants in a 24" monitor over an application created using MATLAB. Figure - 2.1 is a processed frame from the eye-tracker and it shows the GUI used in the study. We maintained the font and size of the text constant throughout the study. Four markers were embedded in the application to identify the surface plane to mark various AOIs during post-processing of the eye-tracking data. Along with the gaze data and post-task questionnaire, we also recorded response and the time of response.

scheme AOI	*authority AOI*	*rest AOI*
https://	www.google.com	/maps/

Table 2.3. Disaggregation of a URL in accordance with the first three AOIs. This differs from Table 2.1 in that the scheme AOI includes the "://" following the scheme.

2.3.4 Data analysis

MATLAB software was used for post-processing all our eye-tracking and post-task questionnaire data. For statistical analysis, we used JMP Pro 14 and R. Initially, the data were tested for normality using the Shapiro-Wilk test. Since all our data were non-normally distributed, we Kruskal-Wallis test and Wilcoxon test for analysis.

2.4 Measures

We used the following measures to understand the visual behavior and the decision outcomes in our study.).

Mood: Participant's mood was assessed along six emotional state : awake, pleasant, angry, fearful, happy, and sad. They ranked their current emotional state using a 10-point scale, where 10 indicated that they highly associated with their mood.

Score: Score is the total number of correctly classified URLs within a set. There was no penalty for incorrect classification.

Total Time Spent: The total time spent on a task is a proxy for cumulative effort and engagement. Here, the total time spent represents the time in seconds that user spend on each URL before clicking on a button to classify it.

We used the following measures to understand and compare the visual behavior while users complete the task.

Time Spent on Areas of Interest: We computed the percentage dwell time on five AOIs using the timestamps from the eye-tracker. This measure is used to assess distribution of users' visual attention and understand which URL components users use to gauge URL safety. These five AOIs are discussed in detail below and Table 2.3 gives a disaggregation of a URL in accordance with the AOIs that correspond to the URL.

- **Scheme AOI:** It consists of the `scheme` component and the delimiters following it. As all the URLs use "https" as the `scheme`, the ***scheme AOI*** in our study always corresponds to "https://" in the URL.

- **Scheme AOI:** It represents the `authority` component. For classes C1 through C7, the `authority` component is a fully qualified domain name, e.g., *`www.google.com`* is the `authority` component of *`https://www.google.com`*. For class C8, the `authority` component has form `user@host`, e.g., as in *`www.google.com@evil.com`*. To test ***P5***, the ***authority AOI*** was further split into two smaller AOIs, the ***user AOI*** and the ***host AOI*** corresponding to the `user` and `host` components. Indeed, the purpose of class 8 is precisely to explore how URLs of this form may break user expectations of what the domain name is.

- **Rest AOI:** It captures the `rest` component.

- **Response AOI:** The gaze targets on the response portion of the screen where the “Safe” and “Unsafe” buttons are located comes under this AOI.

- Anything other than the above mentioned four AOIs is captured in this last AOI and is not a part of the statistical analysis.

Fixations and Backtracking Fixations: Fixation is a cluster of gaze target indicating that the user is looking at a particular information for a certain duration of time. It represents the time where new information is gathered [59]. Whereas, backtracking is the process of revisiting information that was previously processed or skipped [13]. It usually occurs to re-establish previously processed information or it signifies a cognitive interest in an area with respect to the given task [14]. We measured the backtrack fixation count, i.e., the number of fixations involving backtracking.

As the number of characters in each URLs differ, we normalized these measures for comparison. For the overall comparison, we computed overall time spent per character (total time spent/total URL length), overall fixation count per character (total fixation count/total URL length), and backtrack fixation count per character(total backtrack fixations/total URL length). Similarly, the number of characters in the `scheme`, `authority`, and `rest` components of the URL may vary. Thus, for the corresponding AOIs, we calculated the time spent per character (total time spent on AOI divided by number of characters in AOI) and the fixation count per character (total number of fixations occurring on AOI divided by total number of characters in AOI).

2.5 Results & Discussion

The statistical results is discussed in detail in the appendix [P2]. We present here the important hypothesis and their results.

H1: Total time spent on processing a URL is longer for complex URLs than it is for simple URLs.

The overall time spent on classifying simple(and shorter) URLs (C1, C2, C5, C6) was less than the total time spent on classifying complex(and longer) URLs (C3, C4, C7, C8) [Z=3.4865, p=0.0005]. This weakly supports ***H1***, though follow-up work must be done to disentangle length from other complexity factors.

H2: Total time spent on processing a URL, normalized by the URL length, is shorter for complex URLs than it is for simple URLs.

For complex URLs, we found URL length negatively correlated with time spent per character and fixation count per character. This supports ***H2***. We did not observe a correlation between URL length and score.

H3: There exists a URL length threshold over which increasing URL length does not result in more time being spent on processing URLs.

Results indicate that at a threshold of approximately 100 characters, time spent stops increasing as we increase URL length. Similar trends were observed with fixation count per character and backtrack fixation count per character. We also observed no statistical difference between time spent on complex URLs under 100 characters and those above. One interpretation is captured by a notion similar to that of the compliance budget proposed by

Beautement et al. [6]: the user may only expend a finite budget of resources (here, time is a proxy for expended resources) to classify a URL, and, if the resources required to fully process a URL exceeds this budget, the user will not expend them.

H4: Total time spent on the `scheme` per character is less than that of the `authority` and `rest` components.

We observed a statistically significant difference in time spent per character between the ***scheme AOI*** and the ***authority AOI*** (with the latter being higher) [$X^2(2)$=30.4152, p<0.0001]; however, we did not observe such a difference for the ***scheme AOI*** and the ***rest AOI***. Therefore, we do not have evidence to support ***H4***.

The study suggests a sort of ceiling effect: as URL length increases, participants spend less time per character on vetting the URL until a cap on time spent is reached. It also provides visual evidence of user misperceptions regarding URL structure. These insights into how users process and perceive URLs suggest concrete steps and best practices for services to improve the perceived security - and, we argue, the *actual* security - associated with the URLs they serve. For example, from a purely technical standpoint, there is no intrinsic security benefit to serving a URL that is short, has a domain name that begins with *www*, and does not include unnecessary special characters. But if those URLs match users' safety expectations, users would be better at classifying both safe URLs served by the service and unsafe, obfuscated URLs served by the adversary.

Some *unsafe* URLs from our corpus were classified as safe because they exploited uncommon URL features that users rarely encounter in practice with legitimate services. Ironically, this makes such URLs easy to classify as risky by a computer. Surprisingly, we found that some - not all - web browsers offer no user protection against such URLs, even though simple-to-write parsers could easily detect them. This provides an opportunity to improve security at minimal cost.

Last, our findings can improve the quality of security awareness training programs. Our study identifies various misperceptions held by users. It also provides concrete evidence of where users look as they process URLs. This study's methods and data may help in assessing, comparing, and improving training modules that aim to help users correctly identify URLs.

2.6 Conclusions and Future Work

Eye tracking provides a window to examine security behavior. This paper is a first step toward developing a model that captures how users visually process, derive meaning from, and operationalize URL security information to gauge URL safety. We conducted a user study in which participants saw URLs and then classified them while wearing an eye tracker. The findings suggest that participants relied on poor security indicators such as presence of *www* to gauge URL legitimacy, that they spent more time and cognitive resources to vet longer URLs but only up to a point, and that, for the unsafe, *user@host* URLs, participants perceived the `user` component to be the `host` component. In future work, we plan to study other contextual factors such as mood, additional flavors of URL obfuscation, and the effectiveness of training the user.

LIST OF PUBLICATIONS AND OWN CONTRIBUTION

The two user studies presented in this thesis was conducted at the University of New Hampshire between 2017 and 2019. During this period, I collaborated with other researchers and students. The publication [P1] is an extend and replicate study of Pollalis et al. [54]. The second publication [P2] is a collaborative work with Dartmouth College. Both the user studies were conducted by me and my contribution to the research work is mentioned below each publication.

[***P1***]: Niveta Ramkumar, Nadia Fereydooni, Orit Shaer, and Andrew L. Kun. 2019.Visual Behavior During Engagement with Tangible and Virtual Representa-tions of Archaeological Artifacts. In *Proceedings of the 8th ACM International Symposium on Pervasive Displays* (PerDis '19), June 12–14, 2019, Palermo,Italy.ACM, New York, NY, USA, 7 pages. DOI: https://doi.org/10.1145/3321335.3324930

Own Contribution: My contribution was primarily to the study design, algorithm development, data analysis and publication writing.

[***P2***]: Niveta Ramkumar, Vijay Kothari, Caitlin Mills, Ross Koppel, Jim Blythe, Sean Smith, and Andrew L. Kun. 2020. Eyes on URLs: Relating Visual Behavior to Safety Decisions. In *Symposium on Eye Tracking Research and Applications (ETRA '20 Full Papers), June 2–5, 2020, Stuttgart, Germany*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3379155.3391328

Own Contribution: I largely contributed to the study design, implementation, data analysis and publication writing.

BIBLIOGRAPHY

[1] G. Aaron and R. Rasmussen. Global phishing survey: Trends and domain name use in 2016, 2017. https://apwg.org/globalphishingsurvey/.

[2] A. Almomani, B. Gupta, S. Atawneh, A. Meulenberg, and E. Almomani. A survey of phishing email filtering techniques. *IEEE communications surveys & tutorials*, 15(4):2070–2090, 2013.

[3] M. Alsharnouby, F. Alaca, and S. Chiasson. Why phishing still works: User strategies for combating phishing attacks. *International Journal of Human-Computer Studies*, 82:69–82, 2015.

[4] A. Alzayat, M. Hancock, and M. Nacenta. Quantitative measurement of virtual vs. physical object embodiment through kinesthetic figural after effects. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 2903–2912, New York, NY, USA, 2014. ACM.

[5] J. Bacca, S. Baldiris, R. Fabregat, S. Graf, and D. Kinshuk. Augmented reality trends in education: A systematic review of research and applications. *Educational Technology and Society*, 17:133–149, 10 2014.

[6] A. Beautement, M. A. Sasse, and M. Wonham. The compliance budget: managing security behaviour in organisations. In *Proceedings of the 2008 New Security Paradigms Workshop*, pages 47–58. ACM, 2009.

[7] J. R. Bergstrom and A. Schall. *Eye tracking in user experience design*. Elsevier, 2014.

[8] T. Berners-Lee, R. Fielding, L. Masinter, et al. Uniform resource identifiers (uri): Generic syntax, 1998. https://tools.ietf.org/html/rfc3986.

[9] T. Berners-Lee, L. Masinter, M. McCahill, et al. Uniform resource locators (url). 1994. https://tools.ietf.org/html/rfc1738.

[10] D. Beymer, D. Russell, and P. Orton. An eye tracking study of how font size and type influence online reading. In *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction-Volume 2*, pages 15–18. BCS Learning & Development Ltd., 2008.

[11] D. Beymer and D. M. Russell. Webgazeanalyzer: a system for capturing and analyzing web reading behavior using eye gaze. In *CHI'05 extended abstracts on Human factors in computing systems*, pages 1913–1916. ACM, 2005.

[12] C. M. Bohn-Gettler and D. N. Rapp. Depending on my mood: Mood-driven influences on text comprehension. *Journal of Educational Psychology*, 103(3):562, 2011.

[13] D. Bruneau, M. A. Sasse, and J. McCarthy. The eyes never lie: the use of eyetracking data in hci research. ACM, 2002.

[14] C. Burton and M. Daneman. Compensating for a limited working memory capacity during reading: Evidence from eye movements. *Reading Psychology*, 28(2):163–186, 2007.

[15] K.-E. Chang, C.-T. Chang, H.-T. Hou, Y.-T. Sung, H.-L. Chao, and C.-M. Lee. Development and behavioral pattern analysis of a mobile guide system with augmented reality for painting appreciation instruction in an art museum. *Comput. Educ.*, 71:185–197, Feb. 2014.

[16] H. J. Chatterjee, L. Hannan, and L. Thomson. An introduction to object-based learning and multisensory engagement. *Engaging the senses: Object-based learning in higher education*, pages 1–18, 2015.

[17] R. Dhamija, J. D. Tygar, and M. Hearst. Why phishing works. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 581–590. ACM, 2006.

[18] C. E. Drake, J. J. Oliver, and E. J. Koontz. Anatomy of a phishing email. In *CEAS*. Citeseer, 2004.

[19] S. Egelman, L. F. Cranor, and J. Hong. You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1065–1074. ACM, 2008.

[20] I. Fette, N. Sadeh, and A. Tomasic. Learning to detect phishing emails. In *Proceedings of the 16th international conference on World Wide Web*, pages 649–656. ACM, 2007.

[21] J. P. Forgas. Mood effects on decision making strategies. *Australian Journal of Psychology*, 41(2):197–214, 1989.

[22] J. H. Goldberg, M. J. Stimson, M. Lewenstein, N. Scott, and A. M. Wichansky. Eye tracking in web search tasks: Design implications. In *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*, ETRA '02, pages 51–58, New York, NY, USA, 2002. ACM.

[23] R. Gu, J. Zhang, J. Zhou, and M. Tong. The baroque music's influence on learning efficiency based on the research of eye movement. In *2014 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)*, pages 511–515, Dec 2014.

[24] M. Hachet, B. Bossavit, A. Cohé, and J.-B. de la Rivière. Toucheo: Multitouch and stereo combined in a seamless workspace. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, UIST '11, pages 587–592, New York, NY, USA, 2011. ACM.

[25] S. S. Hacisalihzade, L. W. Stark, and J. S. Allen. Visual perception and sequences of eye movement fixations: a stochastic modeling approach. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(3):474–481, May 1992.

[26] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati, editors, *Human Mental Workload*, volume 52 of *Advances in Psychology*, pages 139 – 183. North-Holland, 1988.

[27] K. W. Hong, C. M. Kelley, R. Tembe, E. Murphy-Hill, and C. B. Mayhorn. Keeping up with the joneses: Assessing phishing susceptibility in an email task. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 57, pages 1012–1016. SAGE Publications Sage CA: Los Angeles, CA, 2013.

[28] J. Hyönä, R. F. Lorch Jr, and J. K. Kaakinen. Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. *Journal of Educational Psychology*, 94(1):44, 2002.

[29] J. Hyönä, R. F. Lorch Jr, and M. Rinck. Eye movement measures to study global text processing. In *The Mind's Eye*, pages 313–334. Elsevier, 2003.

[30] D. E. Irwin and G. J. Zelinsky. Eye movements and scene perception: Memory for things observed. *Perception & Psychophysics*, 64(6):882–895, Aug 2002.

[31] M. Jakobsson. Modeling and preventing phishing attacks. In *Financial Cryptography*, volume 5, 2005.

[32] T. Kador, H. Chatterjee, and L. Hannan. *The materials of life: Making meaning through object-based learning in twenty-first century higher education*, pages 60–74. UCL Press, 2017.

[33] KnowBe4. Kevin Mitnick Security Awareness Training. `https://www.knowbe4.com/products/kevin-mitnick-security-awareness-training/`, 2019.

[34] P. König, N. Wilming, T. C. Kietzmann, J. P. Ossandón, S. Onat, B. V. Ehinger, R. R. Gameiro, and K. Kaspar. Eye movements as a window to cognitive processes. *J. Eye Mov. Res*, 9(5):1–16, 2016.

[35] P. Kumaraguru, J. Cranshaw, A. Acquisti, L. Cranor, J. Hong, M. A. Blair, and T. Pham. School of phish: a real-world evaluation of anti-phishing training. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, page 3. ACM, 2009.

[36] A. L. Kun. Human-machine interaction for vehicles: Review and outlook. *Foundations and Trends® in Human-Computer Interaction*, 11:201–293, 01 2018.

[37] M.-L. Lai, M.-J. Tsai, F.-Y. Yang, C.-Y. Hsu, T.-C. Liu, S. W.-Y. Lee, M.-H. Lee, G.-L. Chiou, J.-C. Liang, and C.-C. Tsai. A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational research review*, 10:90–115, 2013.

[38] C. Lander, M. Speicher, F. Kerber, and A. Krüger. Towards fixation extraction in corneal imaging based eye tracking data. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, pages LBW558:1–LBW558:6, New York, NY, USA, 2018. ACM.

[39] D. Lasky. Learning from objects: A future for 21st century urban arts education. *Penn GSE Perspectives on Urban Education*, 6(2):72–76, 2009.

[40] P. Marshall. Do tangible interfaces enhance learning? In *Proceedings of the 1st international conference on Tangible and embedded interaction*, pages 163–170. ACM, 2007.

[41] S. Mathe and C. Sminchisescu. Actions in the eye: Dynamic gaze datasets and learnt saliency models for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 37(7):1408–1424, 2015.

[42] Microsoft. Office 365 ATP Safe Links. `https://docs.microsoft.com/en-us/office365/securitycompliance/atp-safe-links`, March 2019.

[43] D. Miyamoto, T. Iimura, G. Blanc, H. Tazaki, and Y. Kadobayashi. Eyebit: Eye-tracking approach for enforcing phishing prevention habits. In *2014 Third International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS)*, pages 56–65. IEEE, 2014.

[44] M. Montessori and H. W. Holmes. *The Montessori Method: Scientific Pedagogy as Applied to Child Education in" The Children's Houses".* Frederick A. Stokes Company, 1912.

[45] S. M. Munn, L. Stefano, and J. B. Pelz. Fixation-identification in dynamic scenes: Comparing an automated algorithm to manual coding. In *Proceedings of the 5th symposium on Applied perception in graphics and visualization*, pages 33–42. ACM, 2008.

[46] M. Nyström, R. Andersson, K. Holmqvist, and J. van de Weijer. The influence of calibration method and eye physiology on eyetracking data quality. *Behavior Research Methods*, 45(1):272–288, Mar 2013.

[47] K. P. O'Hara and S. J. Payne. The effects of operator implementation cost on planfulness of problem solving and learning. *Cognitive psychology*, 35(1):34–70, 1998.

[48] G. Ollmann. The phishing guide–understanding & preventing phishing attacks. *NGS Software Insight Security Research*, 2004.

[49] O. Palinko, A. L. Kun, A. Shyrokov, and P. Heeman. Estimating cognitive load using remote eye tracking in a driving simulator. In *Proceedings of the 2010 symposium on eye-tracking research & applications*, pages 141–144. ACM, 2010.

[50] D. Pappusetty, V. V. R. Chinta, and H. Kalva. Using pupillary response to assess video quality. In *2017 IEEE International Conference on Consumer Electronics (ICCE)*, pages 64–65, Jan 2017.

[51] S. Paracha, T. Takahara, and A. Inoue. Interactive screening of learning issues via eye-tracking technology. In *2016 13th Learning and Technology Conference (L&T)*, pages 1–6. IEEE, 2016.

[52] C. Pollalis. HoloMuse Study, 2017.

[53] C. Pollalis, W. Fahnbulleh, J. Tynes, and O. Shaer. Holomuse: Enhancing engagement with archaeological artifacts through gesture-based interaction with holograms. In *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction*, pages 565–570. ACM, 2017.

[54] C. Pollalis, E. J. Minor, L. Westendorf, W. Fahnbulleh, I. Virgilio, A. L. Kun, and O. Shaer. Evaluating learning with tangible and virtual representations of archaeological artifacts. In *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction*, pages 626–637. ACM, 2018.

[55] M. Pomplun and S. Sunkara. Pupil dilation as an indicator of cognitive workload in human-computer interaction. In *Proceedings of the International Conference on HCI*, volume 2003, 2003.

[56] Proofpoint. Proofpoint Essentials URL Defense: Advanced Protection with Proofpoint's Targeted Attack Protection. `https://www.proofpoint.com/us/resources/data-sheets/essentials-url-defense`, 2019.

[57] Proofpoint. Proofpoint Security Awareness Training. `https://www.proofpoint.com/us/product-family/security-awareness-training`, 2019.

[58] F. Quinkert, T. Lauinger, W. Robertson, E. Kirda, and T. Holz. It's not what it looks like: Measuring attacks and defensive registrations of homograph domains.

[59] N. Ramkumar, N. Fereydooni, O. Shaer, and A. L. Kun. Visual behavior during engagement with tangible and virtual representations of archaeological artifacts. In *Proceedings of the 8th ACM International Symposium on Pervasive Displays*, PerDis '19, pages 21:1–21:7, New York, NY, USA, 2019. ACM.

[60] K. Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124 3:372–422, 1998.

[61] K. Rayner, K. H. Chace, T. J. Slattery, and J. Ashby. Eye movements as reflections of comprehension processes in reading. *Scientific studies of reading*, 10(3):241–255, 2006.

[62] M. A. Recarte and L. M. Nunes. Effects of verbal and spatial-imagery tasks on eye fixations while driving. *Journal of experimental psychology: Applied*, 6(1):31, 2000.

[63] E. D. Reichle, A. Pollatsek, and K. Rayner. Using ez reader to simulate eye movements in nonreading tasks: A unified framework for understanding the eye–mind link. *Psychological review*, 119(1):155, 2012.

[64] P. Reuter, G. Riviere, N. Couture, N. Sorraing, L. Espinasse, and R. Vergnieux. Archeotui - a tangible user interface for the virtual reassembly of fractured archeological objects. In *Proceedings of the 8th International Conference on Virtual Reality, Archaeology and Intelligent Cultural Heritage*, VAST'07, pages 15–22, Aire-la-Ville, Switzerland, Switzerland, 2007. Eurographics Association.

[65] J. Reynolds, D. Kumar, Z. Ma, R. Subramanian, M. Wu, M. Shelton, J. Mason, E. Stark, and M. Bailey. Measuring identity confusion with uniform resource locators. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–12, New York, NY, USA, 2020. Association for Computing Machinery.

[66] D. D. Salvucci and J. H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pages 71–78. ACM, 2000.

[67] K. Sedig, M. Klawe, and M. Westrom. Role of interface manipulation style and scaffolding on cognition and concept learning in learnware. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 8(1):34–59, 2001.

[68] S. Sheng, M. Holbrook, P. Kumaraguru, L. F. Cranor, and J. Downs. Who falls for phish?: a demographic analysis of phishing susceptibility and effectiveness of interventions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 373–382. ACM, 2010.

[69] J. L. Sibert, M. Gokturk, and R. A. Lavine. The reading assistant: eye gaze triggered auditory prompting for reading remediation. In *Proceedings of the 13th annual ACM symposium on User interface software and technology*, pages 101–107. ACM, 2000.

[70] R. Singh, S. Chandra, K. Dhusia, and G. Sharma. Capacitating surveillance and situational awareness with measure of visual engagement using eyetracker. In *2016 International Conference on Computing, Communication and Automation (ICCCA)*, pages 1161–1165. IEEE, 2016.

[71] G. B. Svendsen. The influence of interface style on problem solving. *International Journal of Man-Machine Studies*, 35(3):379–397, 1991.

[72] H. van der Meulen, P. Varsanyi, L. Westendorf, A. L. Kun, and O. Shaer. Towards understanding collaboration around interactive surfaces: Exploring joint visual attention. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 219–220. ACM, 2016.

[73] P. Vorderer, W. Wirth, F. R. Gouveia, F. Biocca, T. Saari, F. Jäncke, S. Böcking, H. Schramm, A. Gysbers, T. Hartmann, et al. Mec spatial presence questionnaire (mec-spq): Short documentation and instructions for application. *Report to the European community, project presence: MEC (IST-2001-37661)*, 3, 2004.

[74] W. H. A. T. W. G. (WHATWG). Url living standard, August 2019. `https://url.spec.whatwg.org`.

[75] M. Wu, R. C. Miller, and S. L. Garfinkel. Do security toolbars actually prevent phishing attacks? In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 601–610. ACM, 2006.

[76] E. Yecan and K. Cagiltay. Cognitive styles and students' interaction with an instructional web-site: Tracing users through eye-gaze. In *Sixth IEEE International Conference on Advanced Learning Technologies (ICALT'06)*, pages 340–342. IEEE, 2006.

[77] J. Zagermann, U. Pfeil, and H. Reiterer. Measuring cognitive load using eye tracking technology in visual computing. In *Proceedings of the sixth workshop on beyond time and errors on novel evaluation methods for visualization*, pages 78–85. ACM, 2016.

APPENDIX A

PUBLICATION [P1]

This research paper has been presented and published in the proceedings of the Association for Computing Machinery (ACM) Symposium on Pervasive Displays (PerDis ’19).

Visual Behavior During Engagement with Tangible and Virtual Representations of Archaeological Artifacts

Niveta Ramkumar
University of New Hampshire
Durham, New Hampshire
nr1063@wildcats.unh.edu

Nadia Fereydooni
University of New Hampshire
Durham, New Hampshire
nf1021@wildcats.unh.edu

Orit Shaer
Wellesley College
Wellesley, Massachusetts
oshaer@wellesley.edu

Andrew L. Kun
University of New Hampshire
Durham, New Hampshire
andrew.kun@unh.edu

ABSTRACT

In this paper, we present results from a study of users' visual behavior while engaging with tangible and virtual representations of archaeological artifacts. We replicated and extended a recent study that introduced an augmented reality system implemented using HoloLens, for engaging with the artifacts. Our study goes beyond the original study to estimate the distribution of users' visual attention for both tangible and virtual representations of the artifacts. Our study confirmed the results of the original study in various aspects. Specifically, participants in both studies confirmed the immersive nature of the HoloLens condition and showed similar learning outcomes in terms of post-task open questions. Additionally, our findings indicate that users allocate their visual attention in similar ways when interacting with virtual and tangible learning material, in terms of total gaze duration, gaze on object duration, and object fixation duration.

CCS CONCEPTS

• **Human-centered computing** → **Mixed / augmented reality**; *Gestural input*; • **Computing methodologies** → Object recognition.

KEYWORDS

Human-Centered Computing, Mixed/Augmented reality, Gesture input, Object recognition, Eye tracking, Object-based learning

ACM Reference Format:
Niveta Ramkumar, Nadia Fereydooni, Orit Shaer, and Andrew L. Kun. 2019. Visual Behavior During Engagement with Tangible and Virtual Representations of Archaeological Artifacts. In *Proceedings of the 8th ACM International Symposium on Pervasive Displays (PerDis '19), June 12–14, 2019, Palermo, Italy.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3321335.3324930

PerDis '19, June 12–14, 2019, Palermo, Italy

ACM ISBN 978-1-4503-6751-6/19/06…$15.00
https://doi.org/10.1145/3321335.3324930

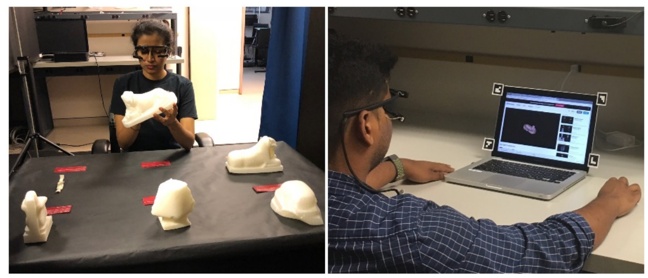


Figure 1: Inventory of 3D artifacts and the participant exploring their chosen artifact (left). A participant wearing eye tracker working with the Sketchfab platform (right)

1 INTRODUCTION

Object-based learning emphasizes the student's interaction with physical artifacts in the learning process. This pedagogical approach has been found to be more effective than relying exclusively on lectures[7]. This approach is well established in various fields including archaeology, art history, and anthropology[33]. The advent of technologies such as virtual reality (VR), augmented reality (AR), and 3D fabrication, has created opportunities for implementing object-based learning without the need to access the original physical artifacts[8]. These technologies allow educators to create tactile and virtual models of the artifacts so that students can learn by exploring and analyzing these models [30]. However, relatively little is known about how tactile and virtual models can be used in object-based learning. Recently, Pollalis et al.[30] conducted an experiment to evaluate learning with three different representations of ancient artifacts. Users interacted with artifacts represented as 3D models on a computer screen, as 3D virtual models in augmented reality, and as 3D fabricated tangible objects. Pollalis et al. found that there were differences in learning outcomes for the three types of presentations. The study we present here replicates core aspects of this study but extend it by asking the following two questions. First, how does visual attention vary among the three conditions? Second, how are differences in visual attention related to the differences in learning outcomes?

Replicating studies in HCI is important because practitioners and researchers could better trust and build upon results from the studies of novel technologies that can be, and have been, replicated.

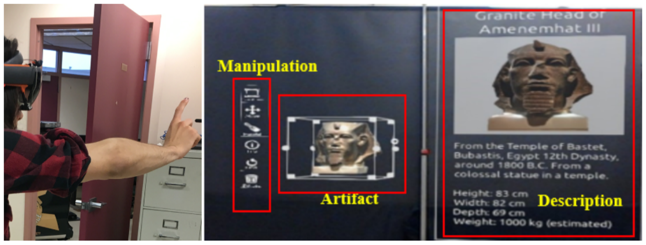


Figure 2: Participant wearing HoloLens with the add-on eye tracker(right). Screenshot of the hologram indicating the three areas of interest (AOI) in the Hololens condition(left).

"Replicate and extend" studies in particular, test the limits and pertinence of previous results[6, 11]. In this study, we replicate the tasks and experimental designs utilized by Pollalis et al. while extending it by collecting eye tracking data to study the visual behavior of participants. By observing visual behaviors, we can provide a quantitative measure of users' interaction with the artifacts, which can in turn help us understand the reasons for the observed learning outcomes. This understanding can then assist us in improving existing and future interactive learning tools. Here we report the findings form a study comparing how users interact with three representations of objects: 3D printed replicas of museum artifacts, holographic replicas, and 3D digital models displayed on a screen. We used an eye tracker to follow users' gaze movements to study their visual behavior. Our contributions beyond replicating the results of the original study include: 1) a new study using eye tracking to analyze users' visual behavior while learning about 3D objects; 2) computational methods for analyzing visual behavior around 3D objects and digital 3D digital models; and 3) understanding of users' visual behavior and how it is related to learning outcomes in an object-based learning activity.

2 BACKGROUND AND RELATED WORK

2.1 Object-Based Learning

Object based learning pedagogy views the learner's interaction with objects as critical for the learning process. Direct interaction with objects allows learners to take charge of their learning process and construct meanings to enhance their critical thinking skills[15]. Much research indicates the benefits of AR for learning and problem solving [3, 18, 27]. AR has been shown to be useful in motivating students in the learning process. Taking advantage of the features of this technology allows educators to improve students' educational experience, their engagement, and their academic achievement [2, 3]. On the other end, evidence shows that concrete visual models, such as 3D printed replicas, could capture students' attention and provide physical context in which to think about concepts. This makes students feel more comfortable visualizing and describing the material [4, 13].

2.1.1 The original study. Pollalis et al. conducted experiments to understand the learning outcomes when users are engaged with three kinds of replicas of museum artifacts: tangible 3D printed artifacts, 3D virtual models presented on a screen, and holographic artifacts [31]. These specific types of objects on the tangible-virtual spectrum were chosen due to their increasing availability in higher education[14]. The authors assessed users' enjoyment, perceived task workload, spatial presence, and learning outcomes. They observed that object-based learning goals were accomplished comparably with holographic artifacts and with the digital 3D models, while 3D replicas lacked visual information impeding learners' contextualization and critical thinking.

However, further studies are needed to understand how these technologies impact object-based learning processes. Moreover, the roles the physicals and digital elements of the learning experience remain to be mapped out [1]. In this study, we replicated the original experiment by Pollalis et al. [31], while adding eye tracking to each condition. Our goal is to build an understanding of users' visual behavior and how it is related to the learning outcomes reported in the original study.

2.2 Learning and Visual Behavior

Several studies use eye movements to describe users' visual attention, as they are considered the behavioral interface between attention and gaining information from the surrounding environment [17, 32, 40] When considering the learning process and its outcomes, many studies use eye tracking to track how learners interact with the learning material [28], predict their level of comprehension [20], and their learning efficiency [7]. Daraghmi et al. developed an on-screen learning system using eye tracking to give learners feedback about their learning [5]. However, the majority of literature that addresses users' visual behavior in learning focuses on learning material that is presented on a 2D surface [7, 21, 23, 28, 40, 42].

Van der Meulen et al. developed a method to combine eye tracking data with head-tracking data provided by HoloLens in order to improve our ability to assess the gaze location of HoloLens users. We are using this method in our study. However, the AR targets in their study were 2D[39]. In this study, we evaluate visual behavior of learners while learning about artifacts replicated in three different and increasingly available methods of creation: 3D printed physical objects, on-screen digital 3D models, and AR visualization of 3D holograms. To our knowledge, this is the first study using eye tracking to analyze users' visual behavior while learning about 3D objects with these modalities.

3 STUDY

3.1 Experimental Design and Tasks

We conducted a between-subjects experiment in which three groups of participants completed the same learning task. The learning task was developed by Pollalis et al. [31]. It is a task that students might encounter in an archaeology class, and its aim is to enhance students' observational skills and their critical analysis skills. The task consists of selecting two artifacts from an available inventory of six artifacts, exploring them, and answering the corresponding artifact questionnaire. For each object, participants were asked to indicate the first detail they noticed, all the details they observed, and what characteristic about the object made it unique or similar to the other artifacts in the set. We did not impose a time limit on the task. Participants completed the learning task either using tangible 3D replicas, virtual 3D replicas, or holographic objects [29].

Across the 3 conditions we used different replicas of the same 6 archaeological artifacts.

3.1.1 Tangible 3D replicas (3D prints). In the tangible 3D printed condition, subjects could choose an artifact from a 3D printed gallery of six objects. These objects are the same 3D printed artifacts used in the original study [31]. The Printed descriptions next to the artifacts are identical to the ones in the original study (Figure 1). Participants were free to choose and explore the replicas while wearing an eye tracker.

3.1.2 Sketchfab Condition. Participants in the second condition used Sketchfab, an online 3D modeling platform on a desktop[29]. They could choose any of the 3D models in an inventory of six objects, manipulate the chosen artifact using a mouse, and read more about it in a section below it. The platform was configured in a similar fashion to Pollalis et al. [31]. In the original study, the description section was right next to the artifact section, so the participant did not have to scroll down to view the descriptions. However, in the new version of the Sketchfab which we used the user has to scroll down the page to read the description.

3.1.3 Holographic objects (HoloMuse). The third condition consisted of participants using HoloMuse [30], an AR application on Microsoft HoloLens, which was developed and used by Pollalis et. al[31]. It introduces subjects to an inventory of six holographic objects. They could use air gestures to pick and handle the artifacts by moving, scaling and rotating them (see Figure 2). Users were also able to remove the artifact's material to view its surface and reveal supplementary information about the artifact.

3.1.4 Eye Tracking. We used Pupil Labs head-mounted eye trackers [16] to track users' gaze during the experiment. This eye tracker has a world camera capturing the users' environment, and two slide cameras for users' pupils. It then calculates users' gaze based on their pupil movements and maps it onto the video from the world camera to display the target the user is looking at. Participants who interacted with 3D artifact and Sketchfab wore the eye tracker. In the HoloLens condition we used an eye tracker add-on[39]. The inventory of artifacts, their order in the inventory, their descriptions, and the eye tracking method were consistent throughout all three conditions.

3.2 Participants

We collected data from 35 participants (10 female, average age = 23.5, SD=3.2); 12 participants in 3D prints condition (3 female), 10 participants for Sketchfab condition (3 female), and 13 participants for HoloLens condition (4 female). All the participants were given a $10 gift card at the end of the experiment. We dismissed data from 2 participants in the 3D prints condition and 3 participants in the HoloLens condition due to a low eye tracker confidence (<70%). The low confidence resulted from a suboptimal angle of the eye tracker with respect to the subject's pupils. Thus, we report on data from 30 participants (10 female).

3.3 Procedure

After signing the consent forms, the participants filled out a pre-task questionnaire stating prior practice with visual analysis (e.g. art history class), and specifying former experience with 3D modeling software, AR, or VR. Depending on the condition they were randomly assigned to, participants were asked to wear the eye tracker or the HoloLens, were shown an inventory of six artifacts, and were given a brief training on how to choose an object and handle it. Before starting the task, the worn eye tracker was calibrated using screen marker calibration [34]. The HoloLens condition included an additional step in which we connected the HoloLens and its eye tracker to a server computer to synchronize time on both devices, so that we get real time data about the user: their position, head rotation, name of the hologram they are viewing, and their gaze information. In order to minimize the movement of the headset, we ensured that the HoloLens's headband was secured on the user's head. HoloLens condition participants were trained on how to use the device as well. Following the initial stage of the study, participants were given the task of choosing and studying two artifacts. They were asked to fill out an object questionnaire for each chosen object using a laptop we provided. After finishing the task, they were given a post-task questionnaire to fill. This form consisted of 15 questions, each being a 5-point Likert-type ratings ranging from "Strongly Disagree" to "Strongly Agree". A NASA TLX questionnaire [10] and four open-ended questions were also part of the post-task form. HoloLens users were also asked if they experienced any discomfort while performing the task. Collected data includes: questionnaire responses and eye tracking; for the HoloLens condition we logged data from the server and recorded videos from the HoloLens camera using its online portal.

4 DATA ANALYSIS

We used JMP Pro 14 for the statistical analysis of the results. The collected data was initially tested for normality using Shapiro-Wilk test. For the normally distributed data we used ANOVA for mean comparison, t-test and Tukey test for post hoc analysis. For the non-normally distributed data we used non-parametric Kruskal-Wallis test and the Wilcoxon test for post-hoc analysis. To analyze the open-ended questions, we used the same coding-scheme and process as the original study. Our participants were free to interact with the artifacts and manipulate them with no time limit or any restrictions on how to study the artifact. To better understand the distribution of users' visual attention, we defined three areas of interest(AOI): the artifact, the description (where they could read more about the artifact), the manipulation (which they could use to manipulate the artifact). Any visual target other than these three were categorized as "other surfaces" and will focus on the main three AOIs for statistical analysis. We developed algorithms using MATLAB (discussed below) to identify participants' visual targets over the time.

4.1 3D Prints Condition

Initially the image from the world camera (see Figure 3) is converted into grayscale image. The resulting image then goes through binary conversion using thresholding, and noise removal. Every frame goes through this two-stage processing to detect objects and descriptions separately. Subsequently, we developed an algorithm to mark a perimeter around each item using its centroid, as shown by blue stars in Figure 3. In this figure, the green star represents

Figure 3: A frame from the eye tracker's world camera with the artifact (white) and the description (red)(right). (b) The same image after being processed to identify the artifact and description AOI(left)

the gaze location of the participant at the artifact. If the gaze by a participant was inside the box around an item, we concluded that they were looking at that item. Otherwise, we marked that gaze as "other surfaces" as it was not aimed at a place of interest (artifact or description). Unlike the other two conditions, there is no manipulation AOI for the 3D prints; participants use their hands to directly manipulate the tactile objects. We validated the above algorithm by visually inspecting 1000 randomly selected frames, we found that the accuracy of the detection algorithm was greater than 95%.

4.2 Virtual 3D replicas (Sketchfab)

Similar to the 3D condition, the video of the world camera was analyzed frame by frame to identify the three AOI. Markers were used to mark the boundaries of the laptop screen (see figure 4). Our algorithm first filtered out the regions on the screen that were not of interest (using the markers) and separated the scroll bar which is the manipulation AOI. Then the remaining section of the screen undergoes binary conversion and noise removal to identify the artifact/description AOI based on their color; artifacts' background is black, and the description background is white (see figure 4).

4.3 HoloLens Condition

The eye tracker add-on camera could not see the holograms displayed by the HoloLens. Similar to Van der Meulen et al. work, we developed an algorithm to map users' gaze position into the holographic environment [39]. We collected the users' gaze information from the eye tracker, and holographic environment details from HoloLens. We logged information from a server that synchronizes the eye tracker and the HoloLens. Then we combined this information to find the gaze target of the participant in the augmented space; the holographic artifact (with its name), description, manipulations (move, rotate, etc.), or other surfaces.

4.4 Measures and Indicators

4.4.1 Time on Task. Using the timestamp of each data point recorded by the eye tracker, we calculated the time spent exploring each artifact and other AOI for all participants. We used this measure to assess meaningful engagement and determine how it is affected by the interaction styles.

4.4.2 Fixations. We used fixations to evaluate the visual behavior of the participants as they attended to different AOI (e.g. artifacts and descriptions). Fixations are the state of maintaining the gaze at a target for a specific amount of time. We extracted fixations with a minimum duration 100 ms[12, 19, 29] and 1°of dispersion angle [25]. We modified the above-mentioned gaze algorithms to find the fixation target. Similar to the gaze AOI, the fixation targets fell into the categories of artifact, manipulation, description, and other surfaces. We explored fixation in terms of fixation rate (fixation count per minute) and duration on the above-mentioned AOI and its overall values.

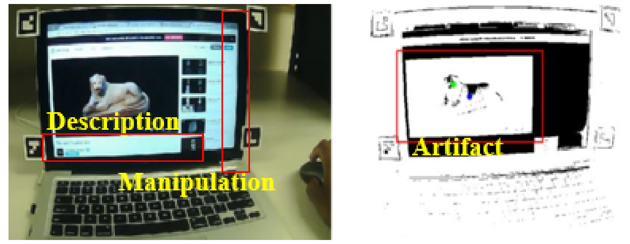


Figure 4: A frame from the eye tracker's world camera (right) and the same image after being processed to identify the artifact and descriptions (left) The red box identifies the artifact region and the green dot shows the user's gaze

4.4.3 Learning Outcomes. Study participants were expected to write down detailed explanations of the viewed artifacts. Content codes were used to demonstrate progress from observation to analysis, and to develop a preliminary review of learning outcomes. We used the codes developed by Pollalis et al.[31] since the questionnaire used in our study was the same as the one they composed. The content code are : texture, color, detail, facial feature, damage, material, weight, size, analysis and context. The first two authors acted as coders identifying the content codes in the questionnaire responses. Their inter-code reliability >95%. Disagreements were resolved by consensus.

4.4.4 Perceived Task Workload and Spatial Presence. We used the NASA TLX questionnaire [10][10] used in the original study to measure participants' perceived workload. Another series of questions used in the original study was utilized to measure participants perceived spatial presence. These questions were roughly based on the MEC- SPQ standardized questionnaire [41].

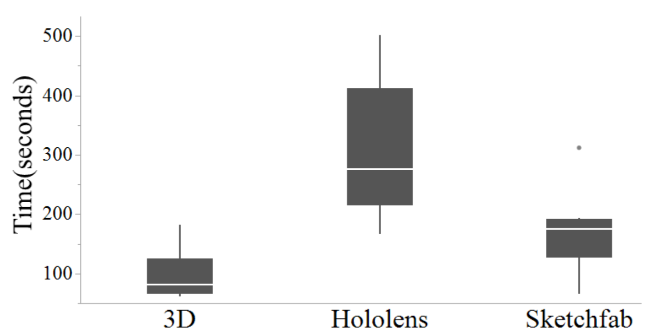


Figure 5: Time participants spent on the task for each condition. The white bar indicates the average time.

5 RESULTS

5.1 Time on Task

The total time participants spent to complete the task can be found in Figure 5. Kruskal-Wallis test indicated that there was a significant effect of condition on total time spent on task [X^2(2)=18.7357, p<0.0001]. The Post hoc comparison showed that the time spent to complete the HoloLens condition task was significantly higher than the other two conditions. This matches the results from the original study for the time on task. Percentage of time distribution among the AOI can be found in Figure 6. There was no significant difference between the time spent on the artifact itself among the three conditions [X^2(2)=3.3626, p=0.1861]. However, the time spent on the description of the artifact differ significantly based on the condition [X^2(2)=13.0477,p=0.0015]. Post hoc testing identified that the time spent on the description was significantly higher in the Sketchfab condition than in the other two conditions. There was a significant difference in the time spent on other surfaces [X^2(2)=19.3652,p<0.0001] and the post hoc analysis indicated significantly higher time spent in the HoloLens condition. Given that there was no manipulation AOI for 3D prints condition, we used the Wilcoxon Test to compare the time that participants spent looking at manipulation in the HoloLens and Sketchfab conditions. We found that the time spent on manipulation was significantly higher in HoloLens condition than Sketchfab condition [Z=14.2857,p<0.0002].

5.2 Fixations

We analyzed fixations from two perspectives; fixation duration, and the fixation rate.

5.2.1 Fixation Rate. The fixation rate (number of fixations per minute) for each condition can be found in Figure 7. The overall fixation rate for the task was not significantly different among the conditions [X^2(2)=1.3871, p=0.4998]. For distribution of fixations among various AOI we report on 28 participants due to momentary server failure for 2 participants in the HoloLens condition (time mismatch). The fixation rate on the artifact was significantly different [X^2(2)=10.8891, p=0.0043]. Post hoc analysis indicated that fixation rate for the HoloLens condition is lower than the other two conditions. Also, the fixation rate on the description was significantly different overall [X^2(2)=19.4813, p<0.0001]. The post hoc analysis indicated that the 3D condition has higher fixation rate on the description than the other two conditions. Participants also had significantly higher fixation rates on manipulation for the HoloLens condition than the Sketchfab condition [Z =8.9195, p=0.0028].

5.2.2 Fixation Duration. We found no evidence that the fixation duration on the artifact [X^2(2)=2.2405, p=0.3262] or description [X^2(2)=14.6945, p=0.0191] were different between the three conditions. However, the t-test results showed that the duration of fixations on the manipulation AOI was significantly higher for the HoloLens condition that the Sketchfab condition [Z=5.4896, p=<0.0006]. Note that there is no separate manipulation AOI for the 3D printed condition.

5.3 Learning Outcomes

We evaluated learning outcomes by counting the number of content codes in the responses to the question asking participants to write down the details they noticed while interacting with the artifact. We found no evidence that the total number of content codes appearing in the responses were different between the conditions [F (2,27) = 1.552, p= 0.8570]. On performing ANOVA on the frequency of mentioning of the individual content codes, we observed that the facial feature and detail code of the visual observation category were significantly different among the conditions. Post hoc testing indicated facial feature was mentioned significantly more often with 3D prints than the other two conditions. Detail was also mentioned significantly less with HoloLens than in the 3D prints and Sketchfab conditions. Rest of the content codes did not show significant difference.

5.4 Perceived Workload and Spatial Presence

There was a significant difference in the perceived workload between conditions [F (2, 27) = 5.7838, p= 0.0081]. Post hoc test suggested that the participants in the 3D prints condition experienced significantly higher effort. Users in the HoloLens condition felt as if the original artifact was physically present in their environment significantly more than users in the other two conditions [X^2(2)=6.3786, p=0.0412]. Participants also claimed to think more intensely about the characteristics of the 3D printed artifacts [X^2(2)=6.7055, p=0.0350].

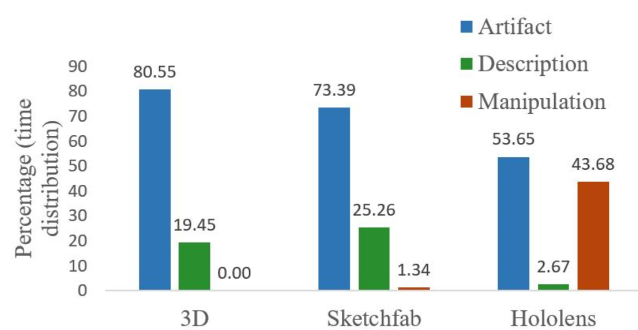


Figure 6: Time percentage distribution for users' gaze among the three areas of interest(AOI) for each condition.

6 DISCUSSION

6.1 Comparison with The Original Study

The time measured in the original study was the overall time spent on the task. Using eye tracking, we were able to classify the total time spent to multiple categories. This enabled us to draw a clear picture of how users distributed their time between analyzing the object, reading about it, and manipulating it. Figure 6 and 7 show this distribution for all three conditions. Our results confirm the results of the original paper in multiple aspects. Both studies found no significant difference in the complexity score of the open question responses about the artifact they viewed. Another common finding was that participants of both studies confirmed the immersive nature of AR by ranking it the highest when asked if they "felt as though the original ancient artifact was physically present in [their] environment". This is despite the fact that 3D printed artifacts were the only ones to include a sense of touch and to exist physically in

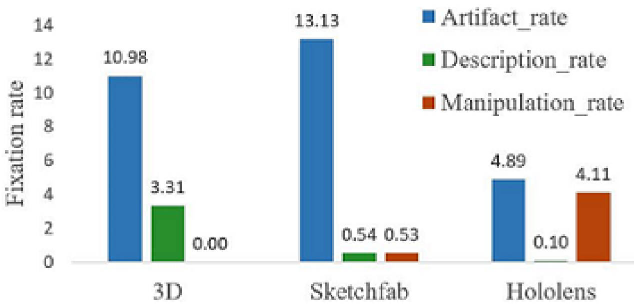


Figure 7: Fixations per minute for each condition.

the environment for the learners. This result suggests a new line of investigation, reconsidering Montessori's finding emphasizing the role of physical material in learning[24]. The total complexity score of the open question responses was comparable among conditions for both studies, However, there were differences with the original study in the particular thematic codes appeared in the responses. The original paper showed that the 3D printed condition was significantly lower in mentions of color, material, and context. Our study did not show such differences among conditions. The reason for this result is not immediately forthcoming; one possibility is difference between participant populations across the two studies, as the studies were conducted at different institutions. Fixations indicate maintaining gaze at a gaze target, during which almost all the visual information is collected [9, 35, 37]. The rate of fixations on the artifact was lowest for the HoloLens condition while the fixation duration was comparable among conditions. Both of the studies had significantly fewer mentions of facial features in the open question responses for the HoloLens condition. It is likely that both the fewer mentions of facial features, and the lower fixation rate are related to two innate features of the holograms: they have lower resolution compared to the Sketchfab condition and are not tactile like the 3D prints condition.

6.2 Physical Artifact vs. Virtual Artifact

Prior research indicates that users have different psychological responses to virtual and tangible objects[1]. However, studies comparing visual behavior for physical and virtual versions of the same task are surprisingly uncommon [22]. One of our contributions is making this comparison. Users in the HoloLens condition spent significantly more time completing the task, however, the total time they spent looking at the object and their object fixation duration was comparable to the other conditions. This is an indication that users' visual behavior towards virtual learning material is similar to tangible ones. This result supports users' claim about the immersive nature of the HoloLens condition; in other words, it seems that participants appreciated the virtual artifacts like the physical ones.

6.3 Interface and Design Implications

Participants reported comparable satisfaction for tangible and virtual interaction types despite technology limitations for current AR equipment. Such limitations like the low resolution, narrow filed of view, and the novelty of the equipment (even though they were trained on how to use HoloLens) posed interaction constraints which participants had to overcome in order to interact with the artifacts. Thus, we are likely to see higher satisfaction measures for the HoloLens condition if interaction becomes more seamless in future products. For example, we observed that the clicking gesture in the AR environment is easier for the users to perform than the dragging gesture. In fact, by visually inspecting the HoloLens videos, we observed that the higher gaze time and fixation duration for manipulation in the HoloLens condition were due to participants' difficulty performing the dragging gesture. Thus, one possible improvement to the interface could be to change how users rotate artifacts: instead of using the dragging gesture, they might prefer to click on a bar that controls object rotation.

Research has shown that interaction costs can lead to increased reflection on the material[9, 26, 36, 38]. Moreover, Marshall claims that the easy manipulation of concrete objects can result in decreased reflection on the learning material [22]. As discussed, the HoloLens condition introduced interaction constraint to the participants. However, in the 3D prints condition users reported that they "thought intensely about the characteristics of the ancient artifact" considerably higher than the other conditions. Based on Marshall's work[22], we expected that participants would report higher thought intensity in the HoloLens condition than in the 3D prints condition, where manipulation was the easiest. This implies that the manipulation effort required in the HoloLens condition might have been too high; this implication is also supported by the fact that participants spent the most time gazing at the manipulation AOI in this condition. Although manipulation for the Sketchfab condition was more complicated than handling the 3D printed artifacts, the Sketchfab condition did not provide the immersive experience for the users which might have been needed

7 LIMITATIONS

One limitation of our detection algorithm is that a minimum color contrast has to be maintained between the 3D printed object and background. However, in the future, we anticipate that 3D printed objects will include color, Hence, future work includes the use of machine learning algorithms to identify various AOI. Although a rare occurrence in our study, another challenge was eye trackers heating up after long usage, creating discomfort for users.

8 CONCLUSION AND FUTURE WORK

By replicating and extending the original study by Pollalis et al. we were able to gain a thorough understanding of users' visual behavior for the purpose of enhancing object-based learning. By adding eye tracking we found that users' visual behavior towards virtual learning material is similar to tangible ones. As mentioned above, users spent a significant amount of time looking at the manipulation features of the HoloLens interface. This highlights a need to further explore the role of physical manipulation in learning. We plan to extend this study to analyze collaborative object-based learning and how different interaction styles facilitate collaborative object-based learning.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation through grant OISE-1658594.

REFERENCES

[1] Ayman Alzayat, Mark Hancock, and Miguel Nacenta. 2014. Quantitative Measurement of Virtual vs. Physical Object Embodiment Through Kinesthetic Figural After Effects. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2903–2912. https://doi.org/10.1145/2556288.2557282

[2] Jorge Bacca, Silvia Baldiris, RamÃşn Fabregat, Sabine Graf, and Dr Kinshuk. 2014. Augmented Reality Trends in Education: A Systematic Review of Research and Applications. *Educational Technology and Society* 17 (10 2014), 133–149.

[3] Kuo-En Chang, Chia-Tzu Chang, Huei-Tse Hou, Yao-Ting Sung, Huei-Lin Chao, and Cheng-Ming Lee. 2014. Development and Behavioral Pattern Analysis of a Mobile Guide System with Augmented Reality for Painting Appreciation Instruction in an Art Museum. *Comput. Educ.* 71 (Feb. 2014), 185–197. https://doi.org/10.1016/j.compedu.2013.09.022

[4] Jill A. Cochran, Zane Cochran, Kendra Laney, and Mandi Dean. 2016. Expanding Geometry Understanding with 3D Printing. *Mathematics Teaching in the Middle School* 21, 9 (2016), 534–542. http://www.jstor.org/stable/10.5951/mathteacmiddscho.21.9.0534

[5] Eman Daraghmi, Chih-Tien Fan, Wei-Jun Lin, Zih-You Peng, and Shyan-Ming Yuan. 2015. Eye-Tracking Assisted Self-Directed Learning System. 52–54. https://doi.org/10.1109/ICALT.2015.33

[6] Christian Greiffenhagen and Stuart Reeves. 2013. Is Replication important for HCI? *CEUR Workshop Proceedings* 976.

[7] R. Gu, J. Zhang, J. Zhou, and M. Tong. 2014. The Baroque music's influence on learning efficiency based on the research of eye movement. In *2014 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)*. 511–515. https://doi.org/10.1109/TALE.2014.7062619

[8] Martin Hachet, Benoit Bossavit, Aurélie Cohé, and Jean-Baptiste de la Rivière. 2011. Toucheo: Multitouch and Stereo Combined in a Seamless Workspace. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST '11)*. ACM, New York, NY, USA, 587–592. https://doi.org/10.1145/2047196.2047273

[9] S. S. Hacisalihzade, L. W. Stark, and J. S. Allen. 1992. Visual perception and sequences of eye movement fixations: a stochastic modeling approach. *IEEE Transactions on Systems, Man, and Cybernetics* 22, 3 (May 1992), 474–481. https://doi.org/10.1109/21.155948

[10] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139 – 183. https://doi.org/10.1016/S0166-4115(08)62386-9

[11] Kasper Hornbæk, Søren S. Sander, Javier Andrés Bargas-Avila, and Jakob Grue Simonsen. 2014. Is Once Enough?: On the Extent and Content of Replications in Human-computer Interaction. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 3523–3532. https://doi.org/10.1145/2556288.2557004

[12] David E. Irwin and Gregory J. Zelinsky. 2002. Eye movements and scene perception: Memory for things observed. *Perception & Psychophysics* 64, 6 (01 Aug 2002), 882–895. https://doi.org/10.3758/BF03196793

[13] Karnyupha Jittivadhna, Pintip Ruenwongsa, and Bhinyo Panijpan. 2010. Beyond Textbook Illustrations: Hand-held Models of Ordered DNA and Protein Structures as 3D Supplements to Enhance Student Learning of Helical Biopolymers. *Biochemistry and molecular biology education : a bimonthly publication of the International Union of Biochemistry and Molecular Biology* 38 (11 2010), 359–64. https://doi.org/10.1002/bmb.20427

[14] Larry Johnson, Samantha Adams Becker, Michele Cummins, Victoria Estrada, Alex Freeman, and Courtney Hall. 2016. *NMC Horizon Report: 2016 Higher Education Edition*. Technical Report. The New Media Consortium, Austin, Texas. https://www.learntechlib.org/p/171478

[15] Thomas Kador, Helen Chatterjee, and Leonie Hannan. 2017. *The materials of life: Making meaning through object-based learning in twenty-first century higher education*. UCL Press, 60–74. http://www.jstor.org/stable/j.ctt1xhr542.10

[16] Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication (UbiComp '14 Adjunct)*. ACM, New York, NY, USA, 1151–1160. https://doi.org/10.1145/2638728.2641695

[17] Andrew L. Kun. 2018. Human-Machine Interaction for Vehicles: Review and Outlook. *Foundations and TrendsÂő in Human-Computer Interaction* 11 (01 2018), 201–293. https://doi.org/10.1561/1100000069

[18] Andrew L Kun, Hidde van der Meulen, and Christian Janssen. 2017. Calling While Driving: An Initial Experiment with HoloLens. 200–206. https://doi.org/10.17077/drivingassessment.1636

[19] Christian Lander, Marco Speicher, Frederic Kerber, and Antonio Krüger. 2018. Towards Fixation Extraction in Corneal Imaging Based Eye Tracking Data. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18)*. ACM, New York, NY, USA, Article LBW558, 6 pages. https://doi.org/10.1145/3170427.3188597

[20] Dorothea Lasky. 2009. Learning from Objects: A Future for 21st Century Urban Arts Education. *Penn GSE Perspectives on Urban Education* 6, 2 (2009), 72–76.

[21] Jiajia Li, Grace Ngai, Hong Va Leong, and Stephen CF Chan. 2016. Your eye tells how well you comprehend. In *2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)*, Vol. 2. IEEE, 503–508.

[22] Paul Marshall. 2007. Do tangible interfaces enhance learning?. In *Proceedings of the 1st international conference on Tangible and embedded interaction*. ACM, 163–170.

[23] Stefan Mathe and Cristian Sminchisescu. 2015. Actions in the eye: Dynamic gaze datasets and learnt saliency models for visual recognition. *IEEE transactions on pattern analysis and machine intelligence* 37, 7 (2015), 1408–1424.

[24] Maria Montessori and Henry Wyman Holmes. 1912. *The Montessori Method: Scientific Pedagogy as Applied to Child Education in" The Children's Houses"*. Frederick A. Stokes Company.

[25] Susan M Munn, Leanne Stefano, and Jeff B Pelz. 2008. Fixation-identification in dynamic scenes: Comparing an automated algorithm to manual coding. In *Proceedings of the 5th symposium on Applied perception in graphics and visualization*. ACM, 33–42.

[26] Kenton P O'Hara and Stephen J Payne. 1998. The effects of operator implementation cost on planfulness of problem solving and learning. *Cognitive psychology* 35, 1 (1998), 34–70.

[27] Muzaffer Ozdemir, Cavus Sahin, Serdar Arcagok, and M Kaan Demir. 2018. The Effect of Augmented Reality Applications in the Learning Process: A Meta-Analysis Study. *Eurasian Journal of Educational Research* 74 (2018), 165–186.

[28] Samiullah Paracha, Toshiro Takahara, and Ayaka Inoue. 2016. Interactive screening of learning issues via eye-tracking technology. In *2016 13th Learning and Technology Conference (L&T)*. IEEE, 1–6.

[29] Christina Pollalis. 2017. HoloMuse Study. https://sketchfab.com/cpollali/collections/holomuse-study

[30] Christina Pollalis, Whitney Fahnbulleh, Jordan Tynes, and Orit Shaer. 2017. HoloMuse: Enhancing engagement with archaeological artifacts through gesture-based interaction with holograms. In *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction*. ACM, 565–570.

[31] Christina Pollalis, Elizabeth Joanna Minor, Lauren Westendorf, Whitney Fahnbulleh, Isabella Virgilio, Andrew L Kun, and Orit Shaer. 2018. Evaluating learning with tangible and virtual representations of archaeological artifacts. In *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction*. ACM, 626–637.

[32] Miguel A Recarte and Luis M Nunes. 2000. Effects of verbal and spatial-imagery tasks on eye fixations while driving. *Journal of experimental psychology: Applied* 6, 1 (2000), 31.

[33] P. Reuter, G. Riviere, N. Couture, N. Sorraing, L. Espinasse, and R. Vergnieux. 2007. ArcheoTUI - a Tangible User Interface for the Virtual Reassembly of Fractured Archeological Objects. In *Proceedings of the 8th International Conference on Virtual Reality, Archaeology and Intelligent Cultural Heritage (VAST'07)*. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 15–22. https://doi.org/10.2312/VAST/VAST07/015-022

[34] Metod Rybar and Maria Bielikova. 2016. Automated detection of user deception in on-line questionnaires with focus on eye tracking use. In *2016 11th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)*. IEEE, 24–28.

[35] Dario D Salvucci and Joseph H Goldberg. 2000. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*. ACM, 71–78.

[36] Kamran Sedig, Maria Klawe, and Marvin Westrom. 2001. Role of interface manipulation style and scaffolding on cognition and concept learning in learnware. *ACM Transactions on Computer-Human Interaction (TOCHI)* 8, 1 (2001), 34–59.

[37] Ram Singh, Sushil Chandra, Kalyani Dhusia, and Greeshma Sharma. 2016. Capacitating surveillance and situational awareness with measure of visual engagement using eyetracker. In *2016 International Conference on Computing, Communication and Automation (ICCCA)*. IEEE, 1161–1165.

[38] Gunnvald B Svendsen. 1991. The influence of interface style on problem solving. *International Journal of Man-Machine Studies* 35, 3 (1991), 379–397.

[39] Hidde van der Meulen, Andrew L Kun, and Orit Shaer. 2017. What Are We Missing?: Adding Eye-Tracking to the HoloLens to Improve Gaze Estimation Accuracy. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*. ACM, 396–400.

[40] Hidde van der Meulen, Petra Varsanyi, Lauren Westendorf, Andrew L Kun, and Orit Shaer. 2016. Towards understanding collaboration around interactive surfaces: Exploring joint visual attention. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. ACM, 219–220.

[41] Peter Vorderer, Werner Wirth, Feliz R Gouveia, Frank Biocca, Timo Saari, Futz Jäncke, Saskia Böcking, Holger Schramm, Andre Gysbers, Tilo Hartmann, et al. 2004. MEC spatial presence questionnaire (MEC-SPQ): Short documentation and instructions for application. *Report to the European community, project presence: MEC (IST-2001-37661)* 3 (2004).

[42] Esra Yecan and Kursat Cagiltay. 2006. Cognitive styles and students' interaction with an instructional web-site: Tracing users through eye-gaze. In *Sixth IEEE International Conference on Advanced Learning Technologies (ICALT'06)*. IEEE, 340–342.

APPENDIX B

PUBLICATION [P2]

This research paper has been published in the long paper category of the Association for Computing Machinery (ACM) Symposium on Eye Tracking Research and Applications (ETRA ’20).

Eyes on URLs: Relating Visual Behavior to Safety Decisions

Niveta Ramkumar
nr1063@wildcats.unh.edu
Univ. of New Hampshire
Durham, NH, USA

Vijay Kothari
vijayk@cs.dartmouth.edu
Dartmouth College
Hanover, NH, USA

Caitlin Mills
caitlin.mills@unh.edu
Univ. of New Hampshire
Durham, NH, USA

Ross Koppel
rkoppel@sas.upenn.edu
Univ. of Pennsylvania
Philadelphia, PA, USA

Jim Blythe
blythe@isi.edu
Information Sciences Inst.
Marina del Rey, CA, USA

Sean Smith
sws@cs.dartmouth.edu
Dartmouth College
Hanover, NH, USA

Andrew L. Kun
andrew.kun@unh.edu
Univ. of New Hampshire
Durham, NH, USA

ABSTRACT

Individual and organizational computer security rests on how people interpret and use the security information they are presented. One challenge is determining whether a given URL is safe or not. This paper explores the visual behaviors that users employ to gauge URL safety. We conducted a user study on 20 participants wherein participants classified URLs as safe or unsafe while wearing an eye tracker that recorded eye gaze (where they look) and pupil dilation (a proxy for cognitive effort). Among other things, our findings suggest that: users have a cap on the amount of cognitive resources they are willing to expend on vetting a URL; they tend to believe that the presence of *www* in the domain name indicates that the URL is safe; and they do not carefully parse the URL beyond what they *perceive* as the domain name.

CCS CONCEPTS

• **Security and privacy** → **Social engineering attacks**; **Social aspects of security and privacy**; **Usability in security and privacy**; • **Human-centered computing** → *User studies*; • **Social and professional topics** → Spoofing attacks.

KEYWORDS

usable security, phishing, user study, eye tracking, cognitive psychology, reading

ACM Reference Format:
Niveta Ramkumar, Vijay Kothari, Caitlin Mills, Ross Koppel, Jim Blythe, Sean Smith, and Andrew L. Kun. 2020. Eyes on URLs: Relating Visual Behavior to Safety Decisions. In *Symposium on Eye Tracking Research and Applications (ETRA '20 Full Papers), June 2–5, 2020, Stuttgart, Germany.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3379155.3391328

1 INTRODUCTION

As people surf the web, check their email, and do other computer-related tasks, they interact with web addresses or Uniform Resource Locators (URLs) [Wikipedia contributors 2019c]. Unfortunately, URLs do not only serve legitimate content; bad actors may use URLs under their control to conduct attacks, e.g., to serve malware or steal credentials by masquerading as a legitimate service. Thus, users must be vigilant. Trusting an unsafe URL could present a security threat to the individual or their organization. Yet users don't want to ignore safe URLs either. This problem is compounded by user misperceptions of URL syntax, the sheer time required to vet URLs, and some practices of legitimate services (e.g., use of URL redirectors). These factors make it very difficult for users to vet URLs. Consequently, many attacks rely on the victim unwittingly clicking on a malicious URL.

From a security standpoint, it is critical to safeguard users from malicious websites. And so, numerous solutions have been developed. Some companies specialize in security training for users (e.g., [KnowBe4 2019; Proofpoint 2019b]). Others focus on limiting user exposure to unsafe URLs: Products and services like Microsoft Office 365 APT Safelinks [Microsoft 2019] and Proofpoint URLDefense [Proofpoint 2019a] check for malicious content served by URLs before allowing users to visit them. Some browsers similarly warn the user when they detect unsafe URLs (e.g., [Mozilla 2019]). There is also abundant research on why users fall for URL-based phishing attacks (e.g., [Dhamija et al. 2006; Hong et al. 2013]), on training techniques (e.g., [Kumaraguru et al. 2009; Miyamoto et al. 2014; Wen et al. 2019]), and on defenses (e.g., [Fette et al. 2007; Maurer et al. 2011]), as well as other foci. However, to the best of our knowledge, this is the first study that solely focuses on understanding users' visual attention as they process URLs. Studying users' visual attention while processing URLs allows us to determine why certain attacks succeed, to measure the influence of URL characteristics on visual processing and cognition, and to determine the efficacy of countermeasures.

The work presented here serves as a first step toward developing a descriptive model of the relationship between URL characteristics and user visual behavior. We conducted a user study where users were asked to classify URLs as safe or unsafe while wearing an eye tracker. One key finding is that participants spent more time on processing URLs as URL length increased but only up to a point. Another is that participants relied more upon the `authority` component of URLs than any other component.

ETRA '20 Full Papers, June 2–5, 2020, Stuttgart, Germany

ACM ISBN 978-1-4503-7133-9/20/06…$15.00
https://doi.org/10.1145/3379155.3391328

2 RELATED WORK

2.1 Eye Tracking and Reading

Eye tracking is considered to be a window into users' cognitive states [König et al. 2016; Reichle et al. 2012]. It has been employed to assess cognitive load [Palinko et al. 2010; Pappusetty et al. 2017; Pomplun and Sunkara 2003; Zagermann et al. 2016], reading strategies [Beymer et al. 2008; Hyönä et al. 2002, 2003; Rayner et al. 2006], and design implications [Bergstrom and Schall 2014; Goldberg et al. 2002]. We study users' eyes as they process URLs.

Users assess the safety of a URL by reading. The amount of visual attention given while reading reflects moment-to-moment cognitive processing [Rayner 1998; Zagermann et al. 2016]. Researchers have sought to examine the relationships between reading and eye movements by using measures like fixations, saccades, regressions, and backtracks [Beymer and Russell 2005; Sibert et al. 2000]. Fixations are pauses in eye movements during which new information is acquired. Research has shown that users fixate longer while reading when "the processing load is greater" [Just and Carpenter 1980].

Reading and scanning text differs with respect to fixations and word skipping [Rayner and Fischer 1996]. When and where someone looks next while reading is influenced by the reader's ongoing mental processing [Rayner and Fischer 1996]. Six commonly used eye-tracking measures are: fixation count, fixation count on various areas of interest (AOIs), proportion of time spent on each AOI, average fixation duration, fixation rate (fixation count/second), and gaze duration mean on each AOI [Lai et al. 2013]. We used all these measures, as well as pupil dilation and backtrack fixation count.

2.2 Pupil Dilation and Cognitive Load

As users read and evaluate URLs, they use cognitive resources. A common measure of cognitive load is pupil dilation [Kun et al. 2013; Palinko et al. 2010; Poole and Ball 2006]. When users face challenging tasks, their pupils dilate on the order of 0.1 to 0.5 mm [Beatty 1982; Pfleging et al. 2016]. This task-evoked pupillary response (TEPR) indicates the cognitive load of the task. However, pupil dilation is also influenced by other factors like the amount of light entering the pupil (pupillary light reflex) [Palinko and Kun 2012; Pfleging et al. 2016] and one's emotional state [Bradley et al. 2008; Stanners et al. 1979; Xu et al. 2011]. To reduce these effects, we conducted the experiment in a windowless light-controlled room.

2.3 Neutral Mood Induction

Mood can affect a person's ability to comprehend text and their judgment [Bohn-Gettler and Rapp 2011; Forgas 1989]. Mood induction is used to understand and reduce the effect of mood [Mills et al. 2019]. Watching a film or a story is one of the most effective mood induction techniques [Westermann et al. 1996]. To reduce the effect of mood and improve replicability, we had participants watch a video chosen to induce a neutral mood.

2.4 URL Security and Phishing

Phishing is the act of masquerading as a legitimate entity to gather sensitive user information [Wikipedia contributors 2019b]. Adversaries often use URL obfuscation to carry out phishing attacks. In fact, URL security is primarily studied in relation to phishing.

Researchers have studied the efficacy of different phishing techniques and demographic factors affecting phishing susceptibility [Dhamija et al. 2006; Downs et al. 2007; Hong et al. 2013; Sheng et al. 2010; Wu et al. 2006]; the impact of psychological manipulation on phishing susceptibility [Goel et al. 2017]; and the effect of communication medium on phishing susceptibility [Benenson et al. 2017; Benenson et al. 2014]. Phishing and URL obfuscation techniques have been categorized, e.g., [Althobaiti et al. 2019; Drake et al. 2004; Ollmann 2004]. However, there are also (ostensibly) legitimate reasons to obfuscate or otherwise break user expectations of where URLs go, e.g., URL redirection [Wikipedia contributors 2019d], tracking links [Cyphers et al. 2018]. Researchers have developed and compared phishing training approaches and educational materials [Arachchilage et al. 2016; Kumaraguru et al. 2007; Sheng et al. 2010, 2007; Stockhardt et al. 2016; Wen et al. 2019]. Companies even provide security training [KnowBe4 2019; PhishingBox 2019; PhishLabs 2019; Proofpoint 2019b; SANS 2019a,b].

Many defenses have also been pursued. Researchers have: compared browser indicators and warnings [Dhamija et al. 2006; Egelman et al. 2008]; developed ways to effectively convey security information [Maurer et al. 2011; Schechter et al. 2007]; and studied ML approaches for email filtering and URL classification [Almomani et al. 2013; Bergholz et al. 2010; Blum et al. 2010; Fette et al. 2007]. Browsers [Mozilla 2019] and search engines [Whittaker et al. 2010] use blacklists and other techniques to protect users. Some products vet URLs in emails before allowing user access, e.g., [Microsoft 2019; Proofpoint 2019a]. However, these defenses are not always foolproof, e.g., [Nathaniel 2017].

Recently, there has been growing interest in using eye trackers for usable security. An eye-tracking based system was developed to train users to look at the status bar [Miyamoto et al. 2014]. Another study involved participants classifying websites, not just URLs, while wearing an eye tracker to examine how users gauge website legitimacy and evaluate security indicators [Alsharnouby et al. 2015]. Our study is similar in spirit. However, we exclusively focus on how users visually process URLs. This narrow focus lets us dissect URLs into smaller components and examine how people process them. We seek to understand which parts people pay attention to, when people give up, and how their eyes process different URLs, amongst other things.

2.5 A Brief Introduction to URL Structure

A uniform resource locator (URL) is a string of characters that specifies the location of a web resource and how to access it [Wikipedia contributors 2019c]. The original URL specification details URL structure [Berners-Lee et al. 1994]. Here, we present the bare essentials of URL structure at an appropriate level of granularity to understand our work.[1]

Each URL in our corpus has the form:

`<scheme>://<authority><rest>`

The `scheme` component [Berners-Lee et al. 1998, 1994; WHATWG 2019] corresponds to the scheme name, which specifies how to interpret the text following the colon. Common schemes are *http*, *ftp*, and *file*. Every URL in our corpus uses the *https* scheme.

[1] A more thorough treatment of URLs can be found in URL and URI specifications and standards [Berners-Lee et al. 1998, 1994; WHATWG 2019].

Table 1: Disaggregation of a URL into its three components.

`scheme`	*delims.*	`authority`	`rest`
https	://	www.google.com	/forms/about/

The `authority` component specifies a subset of the host, port, username, and password [Berners-Lee et al. 1998; WHATWG 2019]. For URLs in our corpus„ the `authority` component has either the form `host` or `user@host` where `host` represents the host and `user` represents the username. In this study, the host is always a fully qualified domain name (e.g., *www.wikipedia.org*) - "a sequence of domain labels separated by '.' " [Berners-Lee et al. 1994]. The last domain label is the top-level domain. For URLs in our corpus, the `authority` component comprises everything following the leading *https://* until either the next /, if present, or the end of the line.

We call the last component `rest`, a catch-all term that is *not* borrowed from any specification or standard. It captures everything following the `authority` component. The `rest` component includes the path [Berners-Lee et al. 1998, 1994; WHATWG 2019], which may be empty; it may also include queries, fragments, and accompanying delimiters [Berners-Lee et al. 1998, 1994; WHATWG 2019]. For every URL in our corpus, if the `rest` component is non-empty, it includes a path that "[identifies] the resource within the scope of [the] scheme and authority" [Berners-Lee et al. 1998], it begins at the first / character following the `authority` component, and it is the last part of the URL. Table 1 provides an example of a URL disaggregation into these three components. Please note the formatting style used for these components. Later, we define areas of interest of the same names but different formatting styles.

3 STUDY OUTLINE

Our long-term goal is to understand users' visual behaviors (and the underlying cognitive processes they manifest) as they process, interpret, and operationalize security information (including information embedded in URLs) when making security decisions. Identifying which factors affect visual behavior and how they affect it is vital in informing security solutions. Such information can be used to improve security awareness training or to better design user interfaces that aid in decision-making.

The work presented in this paper is one step towards this long-term goal. We aim to capture how some URL properties affect visual behaviors. We attempt to control for other factors, but we do not explore them in this initial study. We propose hypotheses pertaining to how various aspects of a URL affect visual processing of the URL, test these hypotheses, and observe trends in users' visual behaviors.

3.1 Hypotheses

We created hypotheses to examine how users visually process URLs and how URL features affect this processing:

H1: Total time spent on processing a URL is longer for complex URLs than it is for simple URLs.

H2: Total time spent on processing a URL, normalized by the URL length, is shorter for complex URLs than it is for simple URLs.

H3: There exists a URL length threshold over which increasing URL length does not result in more time being spent on processing URLs.

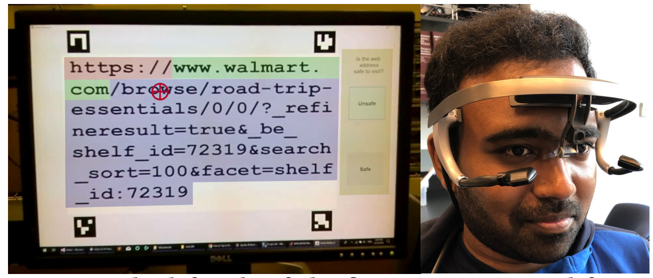


Figure 1: The left side of the figure is a processed frame from the eye tracker video (This is not the same as what the participant sees). The red cursor indicates gaze position and the four colored boxes represent four AOIs: the *scheme AOI* (red), the *authority AOI* (green), the *rest AOI* (blue), and the *response AOI* (yellow). The right side is an image of a participant performing the task wearing the eye tracker.

H4: Total time spent on the `scheme` per character is less than that of the `authority` and `rest` components.

H5: For URLs that have an `authority` component of form `user@host` where `user` ends with ".com", participants spend significantly more time per character looking at the `user` component than the `host` component.

4 METHOD

4.1 URL Corpus and Classification

We created a URL corpus comprising 64 URLs partitioned into 8 categories.[2] Categories are defined by features corresponding to (1) safety, (2) complexity, (3) a leading *www* in the `authority` component, and (4) the attack type for unsafe URLs. The corpus contains 8 URLs for each of the 8 categories. To reduce variability and maintain uniformity between categories, every URL uses *https* as the `scheme` component and *com* as the top-level domain.

The categories are defined by the following 4 features:

4.1.1 Safety: URLs that are *safe* use domain names associated with popular services within the USA, such as Facebook. We selected the fully qualified domain names used in these URLs primarily from the top 1,000 US websites in the Quantcast Top One Million list[3], although we consulted other lists as well. For the subset that were complex and included `rest` components, we chose the `rest` components by searching for legitimate content served by these domain names.

URLs that are *unsafe* have fully qualified domain names that, at the time of corpus construction, were eligible for purchase, did not have a domain name server record, or were spoofed websites. While many URLs with the *unsafe* feature were not actually unsafe to visit, it is exceedingly unlikely that participants would be knowledgeable about the status of the URLs tagged as *unsafe*, and, if an adversary wished to acquire the corresponding domains, they could do so. This decision allowed for greater control over the corpus.

[2]Materials used in this study can be found at https://drive.google.com/drive/folders/1ZNMLoXBxOU4R2nela-6d7MxsaQGrdyg4

[3]https://www.quantcast.com/top-sites

Table 2: Mean values and standard deviations of measurements for the eight URL categories (not normalized by length).

Category	Safety	Complexity	www	Attack Type	URL Length	Time Spent	Score	Fix. Ct.	Backtracking Fix. Ct.
C1	*safe*	*simple*	*www*	N/A	25.0 (4.8)	4.1 (2.3)	7.2 (1.1)	7.9 (4.9)	1.9 (1.8)
C2	*safe*	*simple*	*non-www*	N/A	19.8 (2.0)	4.0 (1.9)	3.8 (2.3)	7.1 (4.3)	1.6 (1.5)
C3	*safe*	*complex*	*www*	N/A	124.0 (13.2)	7.5 (3.8)	5.8 (1.5)	15.3 (8.2)	3.7 (3.1)
C4	*safe*	*complex*	*non-www*	N/A	105.3 (13.5)	7.9 (4.2)	4.4 (1.5)	15.9 (9.0)	4.1 (3.8)
C5	*unsafe*	*simple*	*www*	*positive*	28.5 (2.4)	5.4 (2.1)	4.4 (2.8)	9.5 (4.7)	2.4 (1.9)
C6	*unsafe*	*simple*	*www*	*negative*	29.3 (3.6)	4.8 (1.9)	5.9 (2.2)	9.2 (4.9)	2.4 (2.0)
C7	*unsafe*	*complex*	*non-www*	substring	96.0 (20.4)	7.4 (4.0)	5.5 (2.1)	14.5 (8.3)	3.7 (3.2)
C8	*unsafe*	*complex*	*www*	*user@host*	95.0 (17.4)	6.3 (3.4)	3.4 (2.4)	12.6 (7.4)	3.2 (3.2)

4.1.2 Complexity: URLs were grouped into two complexity classes: *simple* and *complex*. We define complexity in terms of (a) URL length and (b) URL features. A URL is *simple* if it is at most 36 characters long and does not contain a `rest` component. A URL is *complex* if it is at least 48 characters long and contains a non-empty path; it may also contain queries and fragments.

4.1.3 Presence of www: URLs with the *www* attribute begin with *https://www*. URLs with the *non-www* attribute do not.

4.1.4 Attack Type: We chose to explore four conditions for unsafe URLs. They are neither exhaustive nor fully representative of real-world attacks. Rather, our aim was to explore a variety of conditions that may affect visual behaviors and/or classification:

- ***positive*:** The fully qualified domain name contains positive or feel-good words or phrases, e.g., "happy", "bliss".
- ***negative*:** The fully qualified domain name contains words or phrases with a negative, technical, or a security connotation, e.g., "malware", "antivirus", "techsupport".
- ***substring*:** The fully qualified domain name has the form *https://X.Y.com* where *https://X.com* is a safe URL.
- ***user@host*:** The `authority` component has form *www.X.com@Y* where *https://www.X.com* is a legitimate URL. Moreover, some of the last four characters of *Y* are obfuscated using a hexadecimal representation, e.g., representing "*.com*" as "`.%63o%6D`".

The eight URL categories are presented in Table 2. In Section 4.5, we will discuss the measures in this table.

4.2 Experimental Design and Task

We conducted a within-subject experiment that was approved by the Institutional Review Board (IRB). Each of the 20 participants were shown the 64 URLs from the corpus over two sessions. The task was to classify each URL as safe or unsafe. Participants completed this task by viewing one URL at a time and clicking a button on the GUI to indicate whether they believed the URL was safe.

The URL corpus was split into two equal-sized sets presented over two sessions, such that four URLs from each category were represented in each set. For each session, the order in which URLs were presented was randomly determined but held fixed for all participants. However, session order alternated between participants.

4.3 Data Collection, Processing, & Analysis

We discuss the participant selection, the GUI, data collection, data processing, and data analysis:

4.3.1 Participants: We collected data from 20 participants (3 female, mean age = 22.68, SD = 2.65). All participants were students who participated in the user study as part of their coursework. We discarded data from 4 participants due to technical issues with the data extraction from the eye tracker. Hence, we report on the data from 16 participants (2 female).

4.3.2 User interface: The application was created using GUIs in MATLAB. It was presented to participants on a 24" monitor with a resolution of 1920x1200. Each URL image was created using bold monospace font [Wikipedia contributors 2019a] of size 64. The screen was made up of two panes. The first included the URL image, which was scaled and displayed on screen over 2-7 lines with a full line having approximate height of 20mm and width of 280mm. The second pane included the question "Is the web address safe to visit?", accompanied by two response buttons that read "Safe" and "Unsafe" (see Figure 1). Four markers were embedded in the application to identify the surface plane to mark various AOIs during post-processing of the eye-tracking data. Times of clicks and corresponding classifications/responses captured via button clicks were also recorded.

4.3.3 Eye Tracking: We used the head-mounted Dikablis eye tracker to collect gaze positions. It contains three cameras: two eye cameras sampling the eye at 60 Hz and a scene camera sampling at 30 Hz. Gaze positions are computed from the pupil movements and mapped onto the video from the scene camera. Establishing a mathematical mapping between the features of eye and the target being looked at is referred to as calibration. We used the four-point operator-controlled calibration method [Nyström et al. 2013].

4.3.4 Post-task questionnaire: Following the URL classification task, the participant filled in a questionnaire comprising: demographics questions; questions pertaining to security knowledge and behaviors, especially regarding URLs and phishing; and questions to help assess experimental validity.

4.3.5 Data Analysis: We used MATLAB for post-processing the eye-tracking data. We used JMP Pro 14 and R for statistical analysis. The Shapiro-Wilk test indicated that all of our data were non-normally distributed, thus we used non-parametric tests (Kruskal-Wallis test and Wilcoxon test) for analysis.

4.4 Procedure

After signing the consent form, the participant was given a brief introduction to the study and the user interface. They then saw a short neutral mood induction video to control for the effects of

Table 3: Disaggregation of a URL in accordance with the first three AOIs. This differs from Table 1 in that the scheme AOI includes the "://" following the scheme.

scheme AOI	*authority AOI*	*rest AOI*
https://	www.google.com	/forms/about/

mood. They then filled in a pre-task questionnaire to assess their mood [Schaefer et al. 2010], wore the eye tracker, and completed a practice trial to familiarize themselves with the task and the GUI.

Before calibration, we adjusted a nose pin and head band to reduce the movement of the eye tracker during the study; we did not use a chin rest. Next, we focused the eye and scene cameras and calibrated the eye tracker using the four-point operator-controlled calibration method. The participant then classified URLs for the first session and took a break. The calibration procedure was then repeated and the participant classified URLs for the second session. Last, they filled in the post-task questionnaire. The distance between the screen and the participant was kept at about 0.6 meters.

4.5 Measures

4.5.1 Mood: Each participant's mood was assessed along six emotional states: awake, pleasant, angry, fearful, happy, and sad [Mills et al. 2019]. The assessment used a 10-point scale, where 1 indicated that the participant's mood was not associated with the given emotional state, and 10 indicated that it was highly associated.

4.5.2 Score: The score represents the number of correctly classified URLs within a set with no penalty for incorrect classification.

4.5.3 Total Time Spent: The total time spent on classifying a URL is the time (seconds) from the presentation of the URL to the time when the user clicks on a button to classify it. This is a proxy for the cumulative effort and engagement in classifying the URL.

4.5.4 Time Spent on Areas of Interest: Using the UTC timestamps of each data point recorded by the eye tracker, we computed the percentage dwell time on five AOIs (Areas of Interest). These measures express the distribution of users' visual attention and help us understand which URL components users use to gauge URL safety. We examined five AOIs. Figure 1 captures the first four AOIs and Table 3 gives a disaggregation of a URL in accordance with the AOIs that correspond to the URL. We now present the five AOIs.

- The ***scheme AOI*** captures the `scheme` component and the delimiters immediately following it. As every URL in our corpus uses the *https* as the `scheme`, this AOI always corresponds to the leading *https://* in the URL.
- The ***authority AOI*** captures the `authority` component. For classes C1 through C7, the `authority` component is a fully qualified domain name, e.g., *www.google.com* is the `authority` component of *https://www.google.com.* For class C8, the `authority` component has form `user@host`, e.g., as in *www.google.com@evil.com.* To test ***H5***, the ***authority AOI*** was further split into two smaller AOIs, the ***user AOI*** and the ***host AOI*** corresponding to the `user` and `host` components.
- The ***rest AOI*** captures the `rest` component.
- The ***response AOI*** captures the response portion of the screen containing the "Safe" and "Unsafe" buttons.

Table 4: Probabilities of correctly classifying *safe* URLs given the participant knew of the service.

Probabilities	P[correct|known]	P[correct|unknown]
C1 (simple, www)	0.92	0.63
C2 (simple, non-www)	0.83	0.19
C3 (complex, www)	0.76	0.5
C4 (complex, non-www)	0.58	0.46

- The last AOI captured visual targets other than the previous four areas of interest.

4.5.5 Fixations and Backtracking Fixations: Fixating is the act of maintaining one's gaze at a particular target for a certain duration of time. It represents the time where new information is gathered [Ramkumar et al. 2019]. We extracted fixations of 100ms or more following prior research guidelines [Irwin and Zelinsky 2002; Munn et al. 2008; Salvucci and Goldberg 2000].

Backtracking is the process of revisiting information that was previously processed or skipped [Bruneau et al. 2002]. It usually occurs to re-establish previously processed information or it signifies a cognitive interest in an area with respect to the given task [Burton and Daneman 2007]. We measured the backtrack fixation count, i.e., the number of fixations involving backtracking.

4.5.6 Normalized Pupil Area: : The eye tracker records raw pupil area of both eyes in pixels. We used the right eye pupil area. We used the Hampel identifier technique to remove outliers [Foroughi et al.2017; Pearson et al. 2016]. Due to the non-uniform sampling rate, we interpolated the data to obtain a uniform sampling frequency of 60 Hz [Pfleging et al. 2016]. Then, we normalized the data to compare it between participants.

4.5.7 Accounting for Length Differences in URLs: URLs may differ in the number of characters in their `scheme`, `authority`, and `rest` components. Thus, for the corresponding AOIs, we calculated the time spent per character (total time spent on AOI divided by number of characters in AOI) and the fixation count per character (total number of fixations occurring on AOI divided by total number of characters in AOI). For the overall comparison, we computed overall time spent per character (total time spent/total URL length), overall fixation count per character (total fixation count/total URL length), and backtrack fixation count as a function of URL length (total backtrack fixations/total URL length).

5 RESULTS

5.1 Mood Induction Measures

On average participants were awake (ranking of M=7.50, SD=1.59), felt relatively pleasant (M=7.69, SD=1.40), and were mildly happy (M=6.75, SD=1.44). They did not feel angry (M=1.81, SD=0.83), fearful (M=1.56, SD=1.09), or sad (M=1.50, SD=0.82).

5.2 Scores

The average score was 40.44 out of 64. From the post-task questionnaire, we were able to identify whether the participants knew of the services associated with the *safe* URLs. Table 4 indicates the probabilities of participants correctly classifying the URL given that

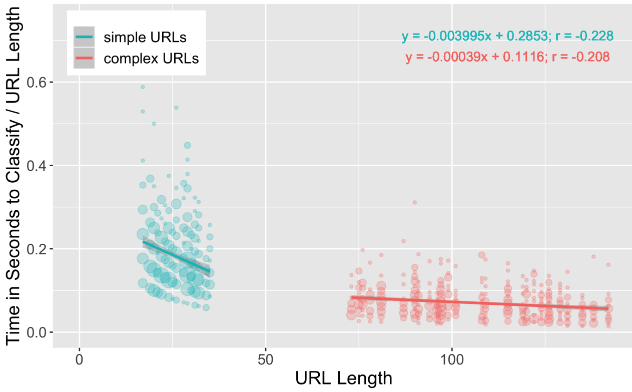


Figure 2: Time spent per character to classify URL vs. URL length with linear regression lines.

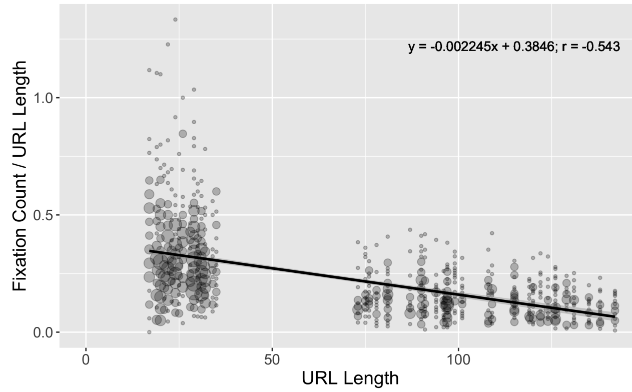


Figure 3: Fixation count per character vs. URL length with a linear regression line.

they knew the service. The Kruskal-Wallis test showed no significant difference between the four categories of *safe* URLs (C1-C4) in terms of the participant knowing the services associated with the domain names [$X^2(3)$=6.9674, p=0.0729].

5.3 Overview of Eye-Tracking Results

Table 2 presents some key results. The overall distribution of visual attention on the AOIs is shown in Figure 6. Using Kruskal-Wallis test, we found that the time spent per character was significantly different between the three AOIs corresponding to the URL [$X^2(2)$=30.4152, p<0.0001]. Post hoc analysis indicated time spent per character on the ***authority AOI*** was significantly higher than that of the ***scheme AOI*** and that of the ***rest AOI***. The fixation count per character was significantly different between the three AOIs [Kruskal-Wallis test: $X^2(2)$=23.9356, p<0.0001]. Post hoc analysis indicated that fixation count per character on the ***rest AOI*** was significantly lower than the other two. However, we found no evidence that fixation duration was significantly different between the three AOIs [Kruskal-Wallis test: $X^2(2)$=3.1692, p=0.0516].

The Kruskal-Wallis test indicated a significant difference in normalized pupil area [$X^2(2)$=8.7532, p=0.0126]. Post hoc analysis indicated a lower pupil area for the ***scheme AOI*** relative to other AOIs, suggesting less cognitive effort was expended on the ***scheme AOI***.

5.4 Complexity

We saw a significant difference in overall time spent (seconds) processing between *complex* and *simple* URLs [Wilcoxon test: Z=3.4865, p=0.0005]. More time was spent on *complex* URLs (M=7.26, SD=2.41) compared to *simple* URLs (M=4.58, SD=1.35). This can also be seen pictorially in Figure 4. Wilcoxon test indicated significant differences in overall time spent per character [Z=8.9998, p<0.0001], overall fixation count per character [Z=6.4883, p<0.0001], and backtrack fixation count as a function of URL length [Z=4.4399, p<0.0001].

People spent less time per character on *complex* URLs (M=0.06, SD=0.01) than *simple* URLs (M=0.13, SD=0.04). Figure 2 shows the time spent per character decreases as URL length increases. Also, the fixation count per character was smaller for *complex* URLs (M=0.12, SD=0.04) than for *simple* URLs (M=0.22, SD=0.10). Figure 3 shows a decrease in fixation count per character as URL length increases. But the backtrack fixation count was higher on *complex* URLs (M=3.68, SD=2.44) relative to *simple* ones (M=2.08, SD=1.18). We found no significant difference in the score between *complex* (M=4.76, SD=2.10) and *simple* URLs (M=5.34, SD=2.51). Examining *complex* URLs of different lengths tells a more nuanced story. Figure 5 suggests a peak in time spent per character that occurs near 100 characters. We observed similar trends with fixation count per character and backtrack fixation count as a function of URL length for *complex* URLs.

5.5 Existence of *www*

We compared *safe* URLs that have `authority` components that begin with *www* (C1&C3) to those that do not (C2&C4). Wilcoxon test results indicated a significant difference in time spent per character on the ***authority AOI*** between *www* URLs (M=0.16, SD=0.04) and *non-www* URLs (M=0.21, SD=0.04); [Z=4.2094, p<0.0001]. Also, there was a significant difference in the fixation count per character on the ***authority AOI*** between *www* URLs (M=0.24, SD=0.09) and *non-www* URLs (M=0.34, SD=0.12); [Wilcoxon test: Z=3.2292, p=0.0012]. The score obtained (maximum score: 8) was also significantly different between *www* URLs (M=6.50, SD=1.48) and *non-www* URLs (M=4.09, SD=1.90); [Wilcoxon test: Z=4.7020, p<0.001].

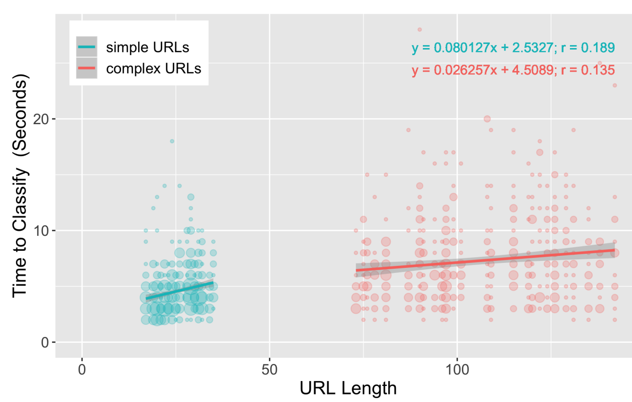


Figure 4: Time spent to classify URL vs. URL length with linear regression lines for simple and complex URLs.

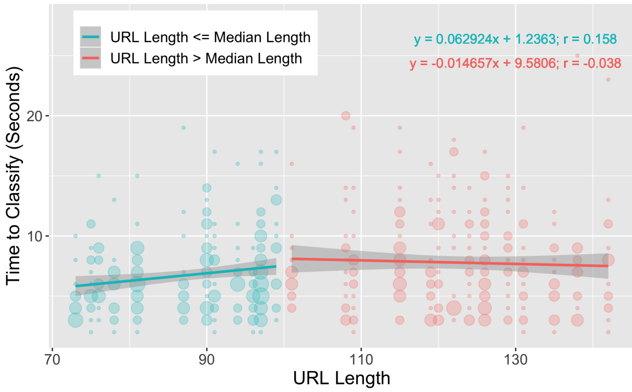


Figure 5: Time spent to classify URLs vs. URL length with two linear regression lines for data points separated by the median URL length (complex URLs).

5.6 User@Host Attack Type vs. Regular URLs

To examine user visual attention for the *user@host* URLs (C8), we considered two special AOIs at a finer granularity than the ***authority AOI***: the ***user AOI*** and ***host AOI***. We compared measurements on these two AOIs for the *user@host* URLs (C8) to those for the ***authority AOI*** for safe URLs of similar structure (C3). Using the Kruskal-Wallis test we found a significant difference on time spent per character between the ***authority AOI*** of C3, the ***user AOI*** of C8, and the ***host AOI*** of C8 [$X^2(2)$=32.1735, p<0.0001]. A significant difference was also observed with fixation count per character [Kruskal-Wallis test: $X^2(2)$=11.3323, p=0.0035]. Post hoc analysis indicated that both sets of measurements for the ***host AOI*** for C8 were lower than those of the ***user AOI*** for C8 and the ***authority AOI*** for C3; the measurements between the ***user AOI*** for C8 were comparable to those of the ***authority AOI*** for C3. These results suggests that users process the ***user AOI*** of C8 and the ***authority AOI*** of C3 similarly. Also, there was a significant difference in the score between the user@host attack type (M=3.37, SD=2.41) and safe URLs of similar structure (M=5.81, SD=1.51); [Wilcoxon test: Z=2.9176, p=0.0035].

6 DISCUSSION

First, participant responses to the pre-task questionnaire following the mood induction video [Schaefer et al. 2010] indicate they were awake and in a neutral mood. Responses to the post-task questionnaire reveal that participants did not fatigue, and, on average, correctly identified the safety of about 40 of the 64 URLs (63%).

Let us now turn to a detailed discussion of the results.

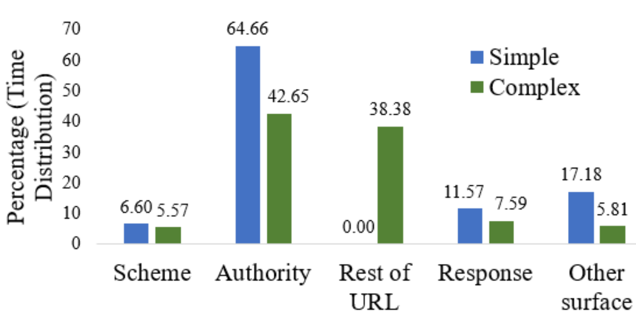


Figure 6: Percentage of time distribution on various AOIs.

6.1 URL Processing & Classification Factors

6.1.1 URL Length: The overall time spent on classifying *simple* (and shorter) URLs (C1, C2, C5, C6) was less than the total time spent on classifying *complex* (and longer) URLs (C3, C4, C7, C8). This weakly supports ***H1***, though follow-up work must be done to disentangle length from other complexity factors.

For *complex* URLs, we found URL length negatively correlated with time spent per character and fixation count per character. This supports ***H2***.

We did not observe a correlation between URL length and score. Also, while Figure 4 suggests participants spent more time parsing URLs as URL length increases, Figure 2 suggests time spent per character decreases as we increase URL length. Moreover, the positive correlation between URL length and time spent seems to cease at a point, which supports ***H3***. Specifically, Figure 5 suggests that at a threshold of approximately 100 characters, time spent stops increasing as we increase URL length. Similar trends were observed with fixation count per character and backtrack fixation count per character. We also observed no statistical difference between time spent on complex URLs under 100 characters and those above. One interpretation is captured by a notion similar to that of the compliance budget proposed by Beautement et al. [Beautement et al. 2008]: the user may only expend a finite budget of resources (here, time is a proxy for expended resources) to classify a URL, and, if the resources required to fully process a URL exceeds this budget, the user will not expend them. While the peculiarities of where that threshold is may depend on factors other than just URL length, we expect this notion of a finite budget applies more generally.

6.1.2 AOI:. We examine the influence of the AOIs:

- **Scheme AOI**: The decrease in the pupil area for the ***scheme AOI*** indicates reduced cognitive attention. Previous work found the frequency with which a user encounters a word affects the fixation duration and processing of that word [Rayner and Duffy 1986]. Users usually spend less time on frequently encountered words. Most legitimate websites use *https* nowadays, which is also used in each of the 64 URLs in our corpus. This explains the decrease in cognitive load for the ***scheme AOI***. We observed a statistically significant difference in time spent per character between the ***scheme AOI*** and the ***authority AOI*** (with the latter being higher); however, we did not observe such a difference for the ***scheme AOI*** and the ***rest AOI***. Therefore, we do not have evidence to support ***H4***.
- **Authority AOI:** The results indicate the time spent per character on the ***authority AOI*** is significantly higher than that of other AOIs. Time spent and fixation count per character on the ***authority AOI*** suggests users find *www* at the beginning of the domain name to be a strong indicator of URL safety.
- **Rest AOI:** Reduced fixation count while reading is characteristic of scanning text [Rayner and Fischer 1996]. The fixation count per character for the rest AOI is significantly

lower than it is for other AOIs, which suggests participants scanned the ***rest AOI***.

6.1.3 Attack Types: Participants classified *positive*, *unsafe* URLs (C5) correctly 55% of the time and they classified *negative*, *unsafe* URLs (C6) correctly 74% of the time. This suggests people are more inclined to trust URLs that use positive words or phrases, even if they have no familiarity with the domain name. Table 4 shows that participants, on average, correctly classified the URLs 77% of the time, given that they had heard of the associated services.

Results suggest users visually process the `user` component of URLs with the *user@host* attack type (C8) similar to how they process the `authority` of URLs without a `user` component. In general, the fixation count per character was low for the `rest` component relative to both the `scheme` and `authority` components. For C8, we observed a reduced fixation count per character and time spent per character on the `host` component, which suggests participants perceived the `host` component as part of the `rest` component. Visual evidence suggests participants misidentified the `user` component as the `host` for URLs in C8. Of the *unsafe* URL categories, participants scored worst on C8. Participants spent significantly more time per character on the `user` component than the `host` component for C8, in support of ***H5***.

We expect classification accuracies observed in this study are upper bounds on what users achieve in practice without additional safeguards in place. Sophisticated attacks that use URL features participants do not know about will likely be more effective. We also expect that attacks that use obfuscation in the `rest` component - or what users *perceive* as the `rest` component - are more likely to succeed given that participants spent less time on the `rest` component than the `authority` component in our study.

6.2 Improving Security in Practice

The study suggests a sort of ceiling effect: as URL length increases, participants spent more time vetting the URL until it capped out at around 100 characters. It also provides visual evidence of user misperceptions regarding URL structure. These insights into how users process and perceive URLs suggest concrete steps and best practices for services to improve the perceived security - and, we argue, the *actual* security - associated with the URLs they serve. For example, from a purely technical standpoint, there is no intrinsic security benefit to serving a URL that is short, has a domain name that begins with *www*, and has few special characters. But if those URLs match users' safety expectations, users would be better at classifying both safe URLs served by the service and unsafe, obfuscated URLs served by adversaries.

Some *unsafe* URLs from our corpus were classified as safe because they exploited uncommon URL features that users rarely encounter in practice with legitimate services. Ironically, this makes such URLs easy for a computer to classify as risky. Surprisingly, we found that some web browsers offer no user protection against such URLs, even though simple-to-write parsers could easily detect them. This provides an opportunity to improve security at minimal cost.

Last, our findings can improve the quality of security awareness training programs. Our study identifies various misperceptions held by users. It also provides concrete evidence of where users look as they process URLs. This study's methods and data may help in assessing, comparing, and improving training modules that aim to help users correctly identify URLs.

7 LIMITATIONS

Several considerations may have affected study generalizability: Participants were predominantly male college students pursuing electrical engineering degrees. To ensure the eye tracker accurately picked up on AOIs, we used a large font and displayed URLs over multiple lines. URLs were presented in isolation; contextual factors (e.g., the device on which a URL is displayed, the application on which a URL is viewed, or beliefs regarding who sent it) may affect visual behaviors and responses. Also, repeatedly asking participants whether URLs were safe likely sensitized them to phishing attacks.

However, we took precautions to minimize unintended effects. We conducted pilot runs to ensure the interface was clear and user fatigue was minimized. We used the post-experiment questionnaire to evaluate experimental validity. And we used a neutral-mood-inducing video to reduce variability in mood.

The available indicators provide some evidence of the study's validity. The average participant score of 63% is within the ballpark of similar studies, e.g., [Dhamija et al. 2006; Sheng et al. 2010]. Post-task survey responses indicate most participants took the task seriously, exercised equal or only slightly more caution than they would in practice, and were not fatigued. Though we did not observe significant bias, we believe any bias would be in the direction of more caution and would be unlikely to invalidate our security recommendations as problems during the classification task would also be at play in the real world. We also note that applications and interfaces in the wild may vary regarding font properties so there is no one-size-fits-all approach for conducting such studies.

Last, the URLs may have had features we could not identify that affected participants' visual behaviors and responses. We attempted to mitigate these concerns by including eight URLs per category, but further work is needed. Also, we only considered a few flavors of URL-based attacks. Notably, no attacks made use of the `rest` component, which may have affected participants' visual behaviors.

8 CONCLUSION AND FUTURE WORK

Eye tracking provides a window to examine security behavior. This paper is a first step toward developing a model that captures how users visually process, derive meaning from, and operationalize URL security information to gauge URL safety. We conducted a user study in which participants saw URLs and then classified them while wearing an eye tracker. The findings suggest that participants relied on poor security indicators such as presence of *www* to gauge URL legitimacy, that they spent more time and cognitive resources to vet longer URLs but only up to a point, and that, for the unsafe, *user@host* URLs, participants perceived the `user` component to be the `host` component. In future work, we plan to study other contextual factors such as mood, additional flavors of URL obfuscation, and the effectiveness of training the user.

ACKNOWLEDGMENTS

The work of Ramkumar and Kun was supported in part by NSF grant OISE 1658594.

REFERENCES

A. Almomani, B. B. Gupta, S. Atawneh, A. Meulenberg, and E. Almomani. 2013. A Survey of Phishing Email Filtering Techniques. *IEEE Communications Surveys Tutorials* 15, 4 (Fourth 2013), 2070–2090. https://doi.org/10.1109/SURV.2013.030713.00020

Mohamed Alsharnouby, Furkan Alaca, and Sonia Chiasson. 2015. Why phishing still works: User strategies for combating phishing attacks. *International Journal of Human-Computer Studies* 82 (2015), 69 – 82. https://doi.org/10.1016/j.ijhcs.2015.05.005

K. Althobaiti, G. Rummani, and K. Vaniea. 2019. A Review of Human- and Computer-Facing URL Phishing Features. In *2019 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. 182–191. https://doi.org/10.1109/EuroSPW.2019.00027

Nalin Asanka Gamagedara Arachchilage, Steve Love, and Konstantin Beznosov. 2016. Phishing threat avoidance behaviour: An empirical investigation. *Computers in Human Behavior* 60 (2016), 185–197.

Jackson Beatty. 1982. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin* 91, 2 (1982), 276.

Adam Beautement, M. Angela Sasse, and Mike Wonham. 2008. The Compliance Budget: Managing Security Behaviour in Organisations. In *Proceedings of the 2008 New Security Paradigms Workshop (NSPW '08)*. ACM, New York, NY, USA, 47–58. https://doi.org/10.1145/1595676.1595684

Zinaida Benenson, Freya Gassmann, and Robert Landwirth. 2017. Unpacking Spear Phishing Susceptibility. In *Financial Cryptography and Data Security*, Michael Brenner, Kurt Rohloff, Joseph Bonneau, Andrew Miller, Peter Y.A. Ryan, Vanessa Teague, Andrea Bracciali, Massimiliano Sala, Federico Pintore, and Markus Jakobsson (Eds.). Springer International Publishing, Cham, 610–627.

Z. Benenson, A. Girard, N. Hintz, and A. Luder. 2014. Susceptibility to URL-based Internet attacks: Facebook vs. email. In *2014 IEEE International Conference on Pervasive Computing and Communication Workshops (PERCOM WORKSHOPS)*. IEEE, 604–609. https://doi.org/10.1109/PerComW.2014.6815275

André Bergholz, Jan De Beer, Sebastian Glahn, Marie-Francine Moens, Gerhard Paaß, and Siehyun Strobel. 2010. New Filtering Approaches for Phishing Email. *Journal of Computer Security* 18, 1 (Jan. 2010), 7–35. http://dl.acm.org/citation.cfm?id=1734234.1734239

Jennifer Romano Bergstrom and Andrew Schall. 2014. *Eye Tracking in User Experience Design*. Elsevier.

Tim Berners-Lee, Roy Fielding, and Larry Masinter. 1998. *Uniform Resource Identifiers (URI): Generic Syntax*. RFC 3986. RFC Editor. 1–61 pages. https://tools.ietf.org/html/rfc3986

Tim Berners-Lee, Larry Masinter, Mark McCahill, et al. 1994. *Uniform Resource Locators (URL)*. RFC 1738. RFC Editor. 1–25 pages. https://tools.ietf.org/html/rfc1738

David Beymer, Daniel Russell, and Peter Orton. 2008. An Eye Tracking Study of How Font Size and Type Influence Online Reading. In *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction - Volume 2 (BCS-HCI '08)*. BCS Learning & Development Ltd., Swindon, UK, 15–18. http://dl.acm.org/citation.cfm?id=1531826.1531831

David Beymer and Daniel M. Russell. 2005. WebGazeAnalyzer: A System for Capturing and Analyzing Web Reading Behavior Using Eye Gaze. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems (CHI EA '05)*. ACM, New York, NY, USA, 1913–1916. https://doi.org/10.1145/1056808.1057055

Aaron Blum, Brad Wardman, Thamar Solorio, and Gary Warner. 2010. Lexical Feature Based Phishing URL Detection Using Online Learning. In *Proceedings of the 3rd ACM Workshop on Artificial Intelligence and Security (AISec '10)*. ACM, New York, NY, USA, 54–60. https://doi.org/10.1145/1866423.1866434

Catherine M Bohn-Gettler and David N Rapp. 2011. Depending on My Mood: Mood-Driven Influences on Text Comprehension. *Journal of Educational Psychology* 103, 3 (2011), 562.

Margaret M. Bradley, Laura Miccoli, Miguel A. Escrig, and Peter J. Lang. 2008. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology* 45, 4 (2008), 602–607. https://doi.org/10.1111/j.1469-8986.2008.00654.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-8986.2008.00654.x

Daniel Bruneau, M. Angela Sasse, and John Mccarthy. 2002. The Eyes Never Lie: The Use of Eye Tracking Data in HCI Research. In *In Proceedings of the CHI'02 Workshop on Physiological Computing*. ACM Press.

Christine Burton and Meredyth Daneman. 2007. Compensating for a Limited Working Memory Capacity During Reading: Evidence from Eye Movements. *Reading Psychology* 28, 2 (2007), 163–186. https://doi.org/10.1080/02702710601186407 arXiv:https://doi.org/10.1080/02702710601186407

Bennett Cyphers, Alexei Miagkov, and Andrés Arrieta. 2018. Privacy Badger Now Fights More Sneaky Google Tracking. https://www.eff.org/deeplinks/2018/10/privacy-badger-now-fights-more-sneaky-google-tracking.

Rachna Dhamija, J. D. Tygar, and Marti Hearst. 2006. Why Phishing Works. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, New York, NY, USA, 581–590. https://doi.org/10.1145/1124772.1124861

Julie S. Downs, Mandy Holbrook, and Lorrie Faith Cranor. 2007. Behavioral Response to Phishing Risk. In *Proceedings of the Anti-phishing Working Groups 2nd Annual eCrime Researchers Summit (eCrime '07)*. ACM, New York, NY, USA, 37–44. https://doi.org/10.1145/1299015.1299019

Christine E. Drake, Jonathan J. Oliver, and Eugene J. Koontz. 2004. Anatomy of a Phishing Email. In *CEAS 2004 - First Conference on Email and Anti-Spam, July 30-31, 2004, Mountain View, California, USA*.

Serge Egelman, Lorrie Faith Cranor, and Jason Hong. 2008. You've Been Warned: An Empirical Study of the Effectiveness of Web Browser Phishing Warnings. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. ACM, New York, NY, USA, 1065–1074. https://doi.org/10.1145/1357054.1357219

Ian Fette, Norman Sadeh, and Anthony Tomasic. 2007. Learning to Detect Phishing Emails. In *Proceedings of the 16th International Conference on World Wide Web (WWW '07)*. ACM, New York, NY, USA, 649–656. https://doi.org/10.1145/1242572.1242660

Joseph P Forgas. 1989. Mood effects on decision making strategies. *Australian Journal of Psychology* 41, 2 (1989), 197–214.

Sanjay Goel, Kevin Williams, and Ersin Dincelli. 2017. Got Phished? Internet Security and Human Vulnerability. *Journal of the Association for Information Systems* 18, 1 (2017), 2.

Joseph H. Goldberg, Mark J. Stimson, Marion Lewenstein, Neil Scott, and Anna M. Wichansky. 2002. Eye Tracking in Web Search Tasks: Design Implications. In *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications (ETRA '02)*. ACM, New York, NY, USA, 51–58. https://doi.org/10.1145/507072.507082

Kyung Wha Hong, Christopher M. Kelley, Rucha Tembe, Emerson Murphy-Hill, and Christopher B. Mayhorn. 2013. Keeping Up With The Joneses: Assessing Phishing Susceptibility in an Email Task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 57, 1, 1012–1016. https://doi.org/10.1177/1541931213571226 arXiv:https://doi.org/10.1177/1541931213571226

Jukka Hyönä, Robert F Lorch Jr, and Johanna K Kaakinen. 2002. Individual Differences in Reading to Summarize Expository Text: Evidence From Eye Fixation Patterns. *Journal of Educational Psychology* 94, 1 (2002), 44. https://doi.org/10.1037//0022-0663.94.1.44

Jukka Hyönä, Robert F Lorch Jr, and Mike Rinck. 2003. Chapter 16 - Eye Movement Measures to Study Global Text Processing. In *The Mind's Eye*, Jukka Hyönä, Ralph Radach, and Heiner Deubel (Eds.). North-Holland, Amsterdam, 313–334. https://doi.org/10.1016/B978-044451020-4/50018-9

David E Irwin and Gregory J Zelinsky. 2002. Eye movements and scene perception: Memory for things observed. *Perception & Psychophysics* 64, 6 (2002), 882–895.

Marcel A Just and Patricia A Carpenter. 1980. A Theory of Reading: From Eye Fixations to Comprehension. *Psychological review* 87, 4 (1980), 329.

KnowBe4. 2019. Kevin Mitnick Security Awareness Training. https://www.knowbe4.com/products/kevin-mitnick-security-awareness-training/.

Peter König, Niklas Wilming, Tim C Kietzmann, Jose P Ossandón, Selim Onat, Benedikt V Ehinger, Ricardo R Gameiro, and Kai Kaspar. 2016. Eye movements as a window to cognitive processes. *Journal of Eye Movement Research* 9, 5 (2016), 1–16.

Ponnurangam Kumaraguru, Justin Cranshaw, Alessandro Acquisti, Lorrie Cranor, Jason Hong, Mary Ann Blair, and Theodore Pham. 2009. School of Phish: A Real-world Evaluation of Anti-phishing Training. In *Proceedings of the 5th Symposium on Usable Privacy and Security (SOUPS '09)*. ACM, New York, NY, USA, Article 3, 12 pages. https://doi.org/10.1145/1572532.1572536

Ponnurangam Kumaraguru, Yong Rhee, Alessandro Acquisti, Lorrie Faith Cranor, Jason Hong, and Elizabeth Nunge. 2007. Protecting People from Phishing: The Design and Evaluation of an Embedded Training Email System. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 905–914. https://doi.org/10.1145/1240624.1240760

Andrew L. Kun, Oskar Palinko, Zeljko Medenica, and Peter Heeman. 2013. On the Feasibility of Using Pupil Diameter to Estimate Cognitive Load Changes for In-Vehicle Spoken Dialogues. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*. International Speech and Communication Association, 3766–3770.

Meng-Lung Lai, Meng-Jung Tsai, Fang-Ying Yang, Chung-Yuan Hsu, Tzu-Chien Liu, Silvia Wen-Yu Lee, Min-Hsien Lee, Guo-Li Chiou, Jyh-Chong Liang, and Chin-Chung Tsai. 2013. A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review* 10 (2013), 90 – 115. https://doi.org/10.1016/j.edurev.2013.10.001

Max-Emanuel Maurer, Alexander De Luca, and Sylvia Kempe. 2011. Using Data Type Based Security Alert Dialogs to Raise Online Security Awareness. In *Proceedings of the Seventh Symposium on Usable Privacy and Security (SOUPS '11)*. ACM, New York, NY, USA, Article 2, 13 pages. https://doi.org/10.1145/2078827.2078830

Microsoft. 2019. Office 365 ATP Safe Links. https://docs.microsoft.com/en-us/office365/securitycompliance/atp-safe-links.

Caitlin Mills, Jennifer Wu, and Sidney D'Mello. 2019. Being Sad Is Not Always Bad: The Influence of Affect on Expository Text Comprehension. *Discourse Processes* 56, 2 (2019), 99–116. https://doi.org/10.1080/0163853X.2017.1381059 arXiv:https://doi.org/10.1080/0163853X.2017.1381059

D. Miyamoto, T. Iimura, G. Blanc, H. Tazaki, and Y. Kadobayashi. 2014. EyeBit: Eye-Tracking Approach for Enforcing Phishing Prevention Habits. In *2014 Third International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS)*. 56–65. https://doi.org/10.1109/BADGERS.2014.14

Mozilla. 2019. How does built-in Phishing and Malware Protection work? https://support.mozilla.org/en-US/kb/how-does-phishing-and-malware-protection-work.

Susan M. Munn, Leanne Stefano, and Jeff B. Pelz. 2008. Fixation-identification in dynamic scenes: Comparing an automated algorithm to manual coding. In *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization (APGV '08)*. ACM, New York, NY, USA, 33–42. https://doi.org/10.1145/1394281.1394287

Yoav Nathaniel. 2017. Attack Report: Office 365 Security Hacked Using Google Redirect. https://www.avanan.com/resources/open-redirect-vulnerability.

Marcus Nyström, Richard Andersson, Kenneth Holmqvist, and Joost van de Weijer. 2013. The influence of calibration method and eye physiology on eyetracking data quality. *Behavior Research Methods* 45, 1 (01 Mar 2013), 272–288. https://doi.org/10.3758/s13428-012-0247-4

Gunter Ollmann. 2004. The Phishing Guide–Understanding & Preventing Phishing Attacks. *NGS Software Insight Security Research* (2004).

Oskar Palinko and Andrew L. Kun. 2012. Exploring the Effects of Visual Cognitive Load and Illumination on Pupil Diameter in Driving Simulators. In *Proceedings of the Symposium on Eye Tracking Research & Applications (ETRA '12)*. ACM, New York, NY, USA, 413–416. https://doi.org/10.1145/2168556.2168650

Oskar Palinko, Andrew L. Kun, Alexander Shyrokov, and Peter Heeman. 2010. Estimating Cognitive Load Using Remote Eye Tracking in a Driving Simulator. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications (ETRA '10)*. ACM, New York, NY, USA, 141–144. https://doi.org/10.1145/1743666.1743701

D. Pappusetty, V. V. R. Chinta, and H. Kalva. 2017. Using Pupillary Response to Assess Video Quality. In *2017 IEEE International Conference on Consumer Electronics (ICCE)*. 64–65. https://doi.org/10.1109/ICCE.2017.7889231

Bastian Pfleging, Drea K. Fekety, Albrecht Schmidt, and Andrew L. Kun. 2016. A Model Relating Pupil Diameter to Mental Workload and Lighting Conditions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 5776–5788. https://doi.org/10.1145/2858036.2858117

PhishingBox. 2019. phishingbox. https://www.phishingbox.com

PhishLabs. 2019. Security Awareness Training. https://www.phishlabs.com/security-awareness-training/.

Marc Pomplun and Sindhura Sunkara. 2003. Pupil Dilation as an Indicator of Cognitive Workload in Human-Computer Interaction. In *Proceedings of the International Conference on HCI*.

A Poole and Linden Ball. 2006. *Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects.* Idea Group, Inc, 211–219.

Proofpoint. 2019a. Proofpoint Essentials URL Defense: Advanced Protection with Proofpoint's Targeted Attack Protection. https://www.proofpoint.com/us/resources/data-sheets/essentials-url-defense.

Proofpoint. 2019b. Proofpoint Security Awareness Training. https://www.proofpoint.com/us/product-family/security-awareness-training.

Niveta Ramkumar, Nadia Fereydooni, Orit Shaer, and Andrew L. Kun. 2019. Visual Behavior During Engagement with Tangible and Virtual Representations of Archaeological Artifacts. In *Proceedings of the 8th ACM International Symposium on Pervasive Displays (PerDis '19)*. ACM, New York, NY, USA, Article 21, 7 pages. https://doi.org/10.1145/3321335.3324930

Keith Rayner. 1998. Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological bulletin* 124 3 (1998), 372–422.

Keith Rayner, Kathryn H Chace, Timothy J Slattery, and Jane Ashby. 2006. Eye Movements as Reflections of Comprehension Processes in Reading. *Scientific Studies of Reading* 10, 3 (2006), 241–255. https://doi.org/10.1207/s1532799xssr1003_3 arXiv:https://doi.org/10.1207/s1532799xssr1003_3

Keith Rayner and Susan A. Duffy. 1986. Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition* 14, 3 (01 May 1986), 191–201. https://doi.org/10.3758/BF03197692

Keith Rayner and Martin H. Fischer. 1996. Mindless reading revisited: Eye movements during reading and scanning are different. *Perception & Psychophysics* 58, 5 (01 Jul 1996), 734–747. https://doi.org/10.3758/BF03213106

Erik D Reichle, Alexander Pollatsek, and Keith Rayner. 2012. Using E-Z Reader to simulate eye movements in nonreading tasks: A unified framework for understanding the eye–mind link. *Psychological Review* 119, 1 (2012), 155–185.

Dario D. Salvucci and Joseph H. Goldberg. 2000. Identifying Fixations and Saccades in Eye-tracking Protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications (ETRA '00)*. ACM, New York, NY, USA, 71–78. https://doi.org/10.1145/355017.355028

SANS. 2019a. Robust Phishing Awareness Simulation Training that Changes Behavior. https://www.sans.org/security-awareness-training/products/phishing.

SANS. 2019b. The 2019 SANS EndUser Training Suite. https://www.sans.org/security-awareness-training/products/end-user.

Alexandre Schaefer, Frédéric Nils, Xavier Sanchez, and Pierre Philippot. 2010. Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition and Emotion* 24, 7 (2010), 1153–1172. https://doi.org/10.1080/02699930903274322 arXiv:https://doi.org/10.1080/02699930903274322

S. E. Schechter, R. Dhamija, A. Ozment, and I. Fischer. 2007. The Emperor's New Security Indicators. In *2007 IEEE Symposium on Security and Privacy (SP '07)*. 51–65. https://doi.org/10.1109/SP.2007.35

Steve Sheng, Mandy Holbrook, Ponnurangam Kumaraguru, Lorrie Faith Cranor, and Julie Downs. 2010. Who Falls for Phish?: A Demographic Analysis of Phishing Susceptibility and Effectiveness of Interventions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 373–382. https://doi.org/10.1145/1753326.1753383

Steve Sheng, Bryant Magnien, Ponnurangam Kumaraguru, Alessandro Acquisti, Lorrie Faith Cranor, Jason Hong, and Elizabeth Nunge. 2007. Anti-Phishing Phil: The Design and Evaluation of a Game That Teaches People Not to Fall for Phish. In *Proceedings of the 3rd Symposium on Usable Privacy and Security (SOUPS '07)*. ACM, New York, NY, USA, 88–99. https://doi.org/10.1145/1280680.1280692

John L. Sibert, Mehmet Gokturk, and Robert A. Lavine. 2000. The Reading Assistant: Eye Gaze Triggered Auditory Prompting for Reading Remediation. In *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology (UIST '00)*. ACM, New York, NY, USA, 101–107. https://doi.org/10.1145/354401.354418

Robert F. Stanners, Michelle Coulter, Allen W. Sweet, and Philip Murphy. 1979. The Pupillary Response as an Indicator of Arousal and Cognition. *Motivation and Emotion* 3, 4 (01 Dec 1979), 319–340. https://doi.org/10.1007/BF00994048

Simon Stockhardt, Benjamin Reinheimer, Melanie Volkamer, Peter Mayer, Alexandra Kunz, Philipp Rack, and Daniel Lehmann. 2016. Teaching Phishing-Security: Which Way is Best?. In *ICT Systems Security and Privacy Protection*, Jaap-Henk Hoepman and Stefan Katzenbeisser (Eds.). Springer International Publishing, Cham, 135–149.

Zikai Alex Wen, Zhiqiu Lin, Rowena Chen, and Erik Andersen. 2019. What.Hack: Engaging Anti-Phishing Training Through a Role-playing Phishing Simulation Game. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 108, 12 pages. https://doi.org/10.1145/3290605.3300338

Rainer Westermann, Kordelia Spies, Günter Stahl, and Friedrich W Hesse. 1996. Relative effectiveness and validity of mood induction procedures: A meta-analysis. *European Journal of Social Psychology* 26, 4 (1996), 557–580.

WHATWG. 2019. URL Living Standard. https://url.spec.whatwg.org.

Colin Whittaker, Brian Ryner, and Marria Nazif. 2010. Large-Scale Automatic Classification of Phishing Pages. In *NDSS '10*. http://www.isoc.org/isoc/conferences/ndss/10/pdf/08.pdf

Wikipedia contributors. 2019a. Monospace (typeface) — Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Monospace_(typeface)&oldid=884652181 [Online; accessed 20-February-2020].

Wikipedia contributors. 2019b. Phishing — Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Phishing&oldid=914413259. [Online; accessed 8-September-2019].

Wikipedia contributors. 2019c. URL — Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=URL&oldid=909233629 [Online; accessed 11-September-2019].

Wikipedia contributors. 2019d. URL redirection — Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=URL_redirection&oldid=916373985. [Online; accessed 19-September-2019].

Min Wu, Robert C. Miller, and Simson L. Garfinkel. 2006. Do Security Toolbars Actually Prevent Phishing Attacks?. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, New York, NY, USA, 601–610. https://doi.org/10.1145/1124772.1124863

Jie Xu, Yang Wang, Fang Chen, Ho Choi, Guanzhong Li, Siyuan Chen, and Sazzad Hussain. 2011. Pupillary Response Based Cognitive Workload Index Under Luminance and Emotional Changes. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems (CHI EA '11)*. ACM, New York, NY, USA, 1627–1632. https://doi.org/10.1145/1979742.1979819

Johannes Zagermann, Ulrike Pfeil, and Harald Reiterer. 2016. Measuring Cognitive Load Using Eye Tracking Technology in Visual Computing. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization (BELIV '16)*. ACM, New York, NY, USA, 78–85. https://doi.org/10.1145/2993901.2993908

APPENDIX C

INSTITUTIONAL REVIEW BOARD APPROVAL

University *of* New Hampshire

Research Integrity Services, Service Building
51 College Road, Durham, NH 03824-3585
Fax: 603-862-3564

18-Apr-2019

Kun, Andrew L
Electrical & Computer Eng Dept
Kingsbury Hall Rm W219
Durham, NH 03824-2619

IRB #: 6223
Study: Assessing Cognitive Load Using Remote Eye Tracking
Review Level: Expedited
Approval Expiration Date: 27-Apr-2020

The Institutional Review Board for the Protection of Human Subjects in Research (IRB) has reviewed and approved your request for time extension for this study. Approval for this study expires on the date indicated above. At the end of the approval period you will be asked to submit a report with regard to the involvement of human subjects. If your study is still active, you may apply for extension of IRB approval through this office.

Researchers who conduct studies involving human subjects have responsibilities as outlined in the document, *Responsibilities of Directors of Research Studies Involving Human Subjects.* This document is available at http://unh.edu/research/irb-application-resources or from me.

If you have questions or concerns about your study or this approval, please feel free to contact me at 603-862-2003 or Julie.simpson@unh.edu. Please refer to the IRB # above in all correspondence related to this study. The IRB wishes you success with your research.

For the IRB,

Julie F. Simpson
Director

cc: File