Automated Detection and Mitigation of Inefficient Visual Searching Using Electroencephalography and Machine Learning

THESIS

Joshua P. Gallaher, Second Lieutenant, USAF

AFIT-ENG-MS-20-M-022

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENG-MS-20-M-022

AUTOMATED DETECTION AND MITIGATION OF INEFFICIENT VISUAL SEARCHING USING ELECTROENCEPHALOGRAPHY AND MACHINE LEARNING

THESIS

Presented to the Faculty

Department of Electrical and Computer Engineering

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Computer Science

Joshua P. Gallaher, B.S.

Second Lieutenant, USAF

March 2020

AFIT-ENG-MS-20-M-022

AUTOMATED DETECTION AND MITIGATION OF INEFFICIENT VISUAL SEARCHING USING ELECTROENCEPHALOGRAPHY AND MACHINE LEARNING

THESIS

Joshua P. Gallaher, B.S.
Second Lieutenant, USAF

Committee Membership:

Dr. Brett Borghetti
Chair

Dr. Michael Miller
Member

Justin Estepp, M.S.
Member

AFIT-ENG-MS-20-M-022

Abstract

Cognitive biases negatively affect the human decision-making process and can result in a sub-optimal outcome. Decisions made during the high-stress and fast-paced operations of the military are extremely prone to cognitive biases. One cognitive bias is known as confirmation bias. Confirmation bias, or the inappropriate bolstering of an unknown hypothesis whose truth is in question, can gravely impact a key component of military operations: a visual search. A visual search is a type of search where the operator must perform a visual scan of an environment for a specific object or feature while ignoring other distracting objects or features. In order to perform a visual search quickly, operators will often fall back upon prior knowledge of a situation. However, falling back upon prior knowledge can have ill effects if that prior knowledge is biased. Thus, for military operators to safely and effectively perform their jobs, it's necessary to detect biases that could arise during a visual search. While there are currently successful mitigation techniques for a confirmation bias, there exists little research into successfully and consistently mitigating a confirmatory, or an inefficient, visual search. This work investigates possible ways both to detect and to mitigate confirmation biases in a visual search.

In this study, the Efficient Search Experiment (ESE) was completed by 16 participants. The ESE elicited inefficient Visual Search Patterns (VSPs) while both behavioral and physiological data was collected. Various mitigation techniques were employed throughout the experiment to encourage efficient VSPs. The effects of the mitigation techniques on the number of efficient searches performed were examined.

Efficient VSPs significantly increased the accuracy of a visual search and also decreased the time it took find the target. The mitigation techniques of a "nudge"

and a "hint" had the most impact on the number of efficient searches with each mitigation technique having a p-value of < 0.0001. To classify an inefficient search from brain activity, the relationship between Electroencephalography (EEG) signals and inefficient searches was modeled through machine learning. Five models were examined: a Linear Discriminant Analysis (LDA), a random forest classifier (RFC), an Artificial Neural Network (ANN), a Long Short-Term Memory (LSTM) Network, and a Temporal Convolutional Network (TCN).

The best models in terms of participants that achieved better than 50% balanced accuracy were the RFC models with $\frac{14}{16}$ participants achieving higher than 50%. The models with the highest mean balanced accuracy were the LDA models with an average balanced accuracy of 58.1%.

While certain participants' models performed well, overall the models for each dataset only marginally perform better than chance. However, these results do suggest that it is possible to classify an efficient or inefficient search from EEG signals.

Acknowledgements

I would like to thank my faculty advisor, Dr. Brett Borghetti, for his guidance and patience. I would also like to thank my committee members, Dr. Michael Miller, and Mr. Justin Estepp, for their support. Finally, I would like to thank my cats, Kaladin and Auri, for without their understanding and love, none of this would have been possible.

Joshua P. Gallaher

Table of Contents

Page

Page

List of Figures

List of Tables

List of Acronyms

ANN	Artificial Neural Network
AUC	area under the curve
AUROC	Area Under the Receiver Operating Characteristic Curve
CNN	Convolutional Neural Network
CSPC	Control Station PC
CSV	Comma Separated Value
DAQ	data acquisition unit
DM	decision maker
ECG	Electrocardiography
EDA	Electro Dermal Activity
EEG	Electroencephalography
EOG	Electrooculography
ESE	Efficient Search Experiment
FFT	fast Fourier transform
FNR	false negative rate
FPR	false positive rate
GSR	galvanic skin response
ICA	Independent Component Analysis
ICD	informed consent document
IR	infrared
LDA	Linear Discriminant Analysis
LSTM	Long Short-Term Memory
MAPPS	Multi-modal Analysis of Psychophysiological and Performance Signals
MCC	Mathew's Correlation Coefficient

ML machine learning

MSS Matching Subset Size

OPTICAL Optimized CSP and LSTM

QDA Quadratic Discriminant Analsysis

ReLU rectified linear unit

RF random forest

RFC random forest classifier

RNN Recurrent Neural Network

ROC Receiver Operating Characteristic

TCN Temporal Convolutional Network

TNR true negative rate

TPR true positive rate

UDP User Datagram Protocol

VSP Visual Search Pattern

AUTOMATED DETECTION AND MITIGATION OF
INEFFICIENT VISUAL SEARCHING USING
ELECTROENCEPHALOGRAPHY AND MACHINE LEARNING

I. Introduction

1.1 Background and Motivation

When humans must make a decision in an uncertain environment, cognitive biases can affect the decision making process and can result in sub-optimal outcomes [1]. These outcomes can range from a slight delay in the decision-making process to making a judgement based in error. When the decisions to be made are in a military context, the resulting decision can be disastrous. In 1988, the USS Vincennes mistakenly shot down an Iranian commercial airliner which resulted in the loss of the 290 passengers on-board the aircraft [2]. The accident was attributed in part to the USS Vincennes' Captain's over-reliance on incorrect information. Despite many clear signs that the airliner was not a military fighter jet, the Captain of the USS Vincennes chose to focus on the few mistaken signs that the approaching aircraft was an Iranian F-14. Thus, the Captain's cognitive biases were partly to blame in the accident.

Decisions made during the high stress and fast paced operations of the military are prone to cognitive biases. As the availability of information grows ever more prevalent, so too does the challenge of making decisions which effectively use all of the available information. With the rapidly increasing amount of information available for military operators, it grows ever more necessary to detect the sub-optimal decisions made from cognitive biases. With the ability to detect cognitive biases, disasters due to a poor

decision-making process can be averted.

A key aspect of many military operators' jobs involves a visual search. A visual search is a type of search where the operator must perform a visual scan of an environment for a specific object or feature while ignoring other distracting objects or features [3]. In order to perform a search quickly, operators will often fall back upon prior knowledge of the situation [4]. Falling back upon prior knowledge can have ill effects if that prior knowledge is also subject to a cognitive bias as the visual search process can then become biased. An example of this situation occurring would be a pilot scanning the instrument panel and not noticing a dangerous situation. The pilot believes that the plane is working properly, and thus will perform a visual search in order to confirm this hypothesis. Because of the pilot's confirmation bias, the pilot could either see the gauge and discount the gauge's information due to the pilot's belief that nothing is wrong, or the pilot could avoid looking at the gauge completely because of the pilot's belief that nothing is wrong. Thus for military operators to safely and effectively perform their jobs, it's important to be able to detect cognitive biases that arise during a visual search.

1.2 Problem Statement

Confirmation bias is the "inappropriate bolstering of hypotheses or beliefs whose truth is in question" [5]. A biased visual search is a search in which the user falls prey to biases such as a confirmation bias. The effect of a confirmation bias on a visual search in military operations is crucial. When an operator falsely believes a hypothesis and thus performs a visual search to confirm that hypothesis, the operator can miss critical information that would otherwise be obtained by performing a total, efficient visual search.

While there are currently successful mitigation techniques for confirmation bias,

there exists little research into successfully and consistently mitigating a confirmatory, or inefficient, visual search. This study intends to replicate a visual search experiment in which a confirmatory search can be induced. Once induced, the confirmatory search can be identified and then successfully and consistently mitigated.

However, because it is still possible to perform a confirmatory search while also being efficient and also because it is possible to perform a non-confirmatory search while being inefficient, this research focuses on encouraging efficient searches rather than discouraging confirmatory searches.

Although this experiment used gaze tracking to determine when an inefficient search occurred, in a real world context, gaze tracking is normally not a realistic option to indicate whether an inefficient search is occurring. Thus, it is necessary to also be able to detect an efficient or inefficient search through physiological signals.

1.3 Research Questions and Hypotheses

The objective of this research is to determine whether an inefficient visual search during a visual search can be detected and subsequently mitigated. To complete this objective, the following research questions are investigated.

1.3.1 Research Question 1 - Categorizing Natural Behavior

What visual search patterns do participants naturally use during a visual search task?

Hypothesis: The majority ($> 50\%$) of participants will naturally resort to an inefficient Visual Search Pattern (VSP).

1.3.2 Research Question 2 - Behavior Detection

Can physiological signals such as Electroencephalography (EEG), Electrooculography (EOG), and Electrocardiography (ECG) be associated with an efficient visual search?

Hypothesis: Physiological signals can differentiate a participant performing an efficient visual search from a participant performing an inefficient visual search.

Research Objective: Develop a machine learning model that receives physiological data and is able to determine an efficient visual search with an equal-class-weighted classification accuracy of greater than 50%.

1.3.3 Research Question 3 - Behavior Mitigation

For a participant who is performing an inefficient search, can mitigation techniques change the participant's search patterns to an efficient search pattern that will persist for the remainder of the search tasks?

Hypothesis: By applying the mitigation techniques of a nudge, a hint, and by teaching the participant how to perform an efficient search, a participant will perform an efficient search pattern for the remainder of the search tasks.

1.4 Methodology

An experiment was adapted from an existing experiment designed to induce a confirmation bias during a visual search [6]. The experiment was adapted to a new experiment, named the Efficient Search Experiment (ESE), that dynamically applied various mitigation techniques based on the individual participants' search patterns over the course of the experiment. In each block of trials, participants are presented with search stimuli consisting of 8 colored circles, arranged in a ring, with white letters in the center of each colored circle. The white letters were one of "p, q, b, d" which

were chosen to reduce the chance that the target letter was easily distinguishable amongst the other stimuli. There were only two colors per block. Participants were instructed to indicate whether a specific target letter's circle was a specific color, called the target color. There would only ever be one instance of a target letter present per trial, and the target color would not change for the duration of the block. During the block, various proportions of the target color and non-target color appeared. A block was marked as an efficient block if the participant searched the minimum required number of circles to determine what color the target letter's circle was, while the block was marked as inefficient if the participant searched more than the required minimum number. During part of the experiment, a mitigation technique known as the "nudge" was applied on the following block if the participant performed an inefficient search on more than half of the search trials of the previous block. The nudge consisted of hiding the letters on the colored circles unless the participant visually fixated upon a circle. Additionally, a "hint," an "explanation," and "instructions" mitigation techniques were given to the participant during the experiment. The hint consisted of showing the participant how to perform an efficient search, the explanation involved telling the participant why the nudge was occurring, and the instructions instructed the participant to perform an efficient search.

During the ESE, behavioral and physiological measures were collected. Behavioral measures included the participants' VSP and the effects of the mitigation. Physiological measures included EEG, EOG, and ECG. The collected behavioral data was investigated to determine which VSP most participants initially used as well as the effect of the mitigations on their VSPs. The physiological data was investigated to determine whether machine learning classification models could be trained to identify when a participant performed an efficient or inefficient search. The machine learning models were trained within-participant and cross-validation metrics are reported. Finally,

model feature salience was evaluated to determine important features for estimating an efficient or inefficient search.

1.5 Assumptions

To answer the proposed research questions, the following assumptions about the experiment design were made:

- Participants had no knowledge of the ESE's research purpose or the mitigations present other than what was presented during training sessions.
- Participants would seek to perform a visual search that is both efficient and accurate.
- Participants would use a consistent VSP during the experiment.
- EEG activity is different when performing an efficient search versus when performing an inefficient search.
- The physiological recording equipment operates and records correctly.
- Participants would complete the ESE to the best of their ability.

1.6 Limitations

Each experimental session had a 2.5 hour time limit. To accomplish the ESE within this time, and to prevent participant fatigue, the experiment was designed with only 480 trials. With a dataset of this size there arises several opportunities for issues. First, the limited amount of data means that it may not be possible to split the data into the training, validation, and test sets that are normally used for machine learning and for metric reporting. Datasets such as this one also can lead to overfitting. Overfitting can be reduced by reducing the model's complexity, however, this limits the classifiers that can be used. During experimentation, it was observed

that the system used to record the electrophysiological data was malfunctioning and not recording the signals as it should. Because of this, certain participants' data may not reflect the truth. The participants for which the recording errors occurred are annotated.

1.7 Contributions

This work contributes to the field of cognitive biases within visual search by determining the effectiveness of various mitigation techniques. Furthermore, it contributes to the field of visual search by establishing base patterns of a VSP. At the time of this work, dynamically mitigating an inefficient, or a confirmatorily biased, search had not been explored. Additionally, no research had been conducted into which VSP humans tend to initially favor during a visual search. This work builds a foundation for further research into efficient visual searches.

This work determined that performing an efficient search increased the accuracy of target detection by 2.41% ($t(15) = 5.59, p = 0.00005$). Efficient searching also decreased search times by an average of 0.30 seconds ($t(15) = 5.53, p = 0.00005$). Initially, participants overwhelmingly performed inefficient searches. In the first eight blocks, 73.68% of searches were inefficient, 19.14% were efficient, and 7.18% were circular. At the end of the experiment, because of the use of mitigation techniques, participants performed more efficient searches than inefficient searches. In the last seven blocks, 47.53% of searches were inefficient, 51.41% were efficient, and 1.06% were circular.

The most effective mitigation techniques were the addition of the nudge and the hint which informed the participants how to perform an efficient search. In regards to increasing the number of efficient searches, both mitigation techniques had a p-value of < 0.001. However, the nudge's log worth ($-log_{10}(p - value)$) was 10.664 and the

hint's log worth was 8.493 which indicates that the nudge had more of an effect on increasing the number of efficient searches.

Two datasets were considered in this work: a dataset with extracted features based on the five frequency bands of the alpha, beta, delta, gamma, and theta bands from each electrode; and raw time-series voltage values from each electrode.

Three models were examined for the frequency feature dataset: Linear Discriminant Analysis (LDA), random forest classifier (RFC), and Artificial Neural Network (ANN) models. The LDA models achieved greater than 50% balanced accuracy on $\frac{13}{16}$ participants, the RFC models achieved greater than 50% balanced accuracy on $\frac{14}{16}$ participants, and the ANN models achieved a greater than 50% balanced accuracy on $\frac{6}{16}$ participants. The highest balanced accuracies were 74.75%, 66.71%, and 64.45% for the LDA, RFC, and ANN models respectively.

Two models were examined for the time-series dataset: Long Short-Term Memory (LSTM) and Temporal Convolutional Network (TCN) models. The LSTM models achieved greater than 50% balanced accuracy on $\frac{13}{16}$ participants and the TCN models achieved a greater than 50% balanced accuracy on $\frac{7}{16}$ participants. The highest balanced accuracies were 65.12% and 61.18% for the LSTM and TCN models respectively.

While certain participants' models performed well, overall the models for each dataset only marginally perform better than chance. However, these results do suggest that it is possible to classify an efficient or inefficient search from EEG signals.

1.8 Structure of the Document

The remainder of this document is structured into four chapters. Chapter II provides an overview of the literature on confirmation bias and visual search. Additionally, it provides a review of various machine learning approaches for classifying EEG. Chapter III describes the details of the experiment that was conducted to collect

behavioral and physiological data as well as the machine learning pipeline used to analyze the data. Chapter IV presents and discusses the results of the analysis of the behavioral and physiological data. Finally, Chapter V concludes this work by summarizing the significant findings of this research and by discussing areas for future work.

II. Literature Review

2.1 Chapter Overview

This chapter provides an overview of decision-making research on confirmation bias, confirmation bias mitigation, and visual search. Confirmation bias definitions, measures, and task environments used in research are discussed. Additionally, confirmation bias mitigation techniques and their effects are discussed. A brief overview of visual search is provided. Lastly, current state-of-the-art machine learning models for use in classifying Electroencephalography (EEG) are discussed.

2.2 Cognitive Biases

Today, military operators have more access to real-time information than ever before. No longer can poor results be blamed on a lack of information - rather, the blame lies on the operators' failure to appropriately rely on the information at their disposal. A failure to appropriately rely on the information at hand is more likely to result in a poor decision-making process that ends in a sub-optimal outcome. To ensure an optimal outcome, the operators' recognition and prevention of errors due to cognitive biases is paramount. The consequences of allowing cognitive bias errors to occur leaves a bloody trail throughout history. In 1998, the USS Vincennes shot down Iran Air Flight 655 killing all 290 passengers aboard [2]. The cause of the accident is tragic: the captain of the USS Vincennes received conflicting information about the type of aircraft seen on the radar and mistakenly believed the approaching airliner to be an Iranian Air Force F-14 fighter jet. One cause of the accident that was cited in the incident report was the high tension of the situation coupled with recent incidents in the area that caused the captain to suffer from a confirmation bias. Because the captain suffered from a confirmation bias, he overvalued the information

that supported his hypothesis that the incoming airliner was a hostile enemy aircraft. Although the destruction of Iran Air Flight 655 is one of the most prominent examples of a cognitive bias that has led to a loss of life, not all of the results of confirmation bias are so deadly. Other effects of a confirmation bias can include the persistence of discredited beliefs, the preference for information discovered early in the decision process, and the tendency to see non-existent correlations in a set of data [5], [7], [8]. Although not deadly in and of itself, these effects can have drastic consequences in a military environment. From military pilots making timely decisions in a rapidly evolving environment, to intelligence analysts making decisions given a vast amount of information, to cyber operators deciding how to mount a cyber attack, all situations require that the decision maker perform objective assessments of information to make an unbiased decision.

When there is limited information, or when in an unknown situation, people use heuristics, or mental shortcuts, to help simplify complex decisions [1], [9]. One example of a heuristic is when judging the distance of an object. Objects that are closer to the viewer appears sharper and clearer when compared to an object that is far away. Because of this, when objects are clear we tend to estimate that the object is closer than it actually is [1]. There are situations in which using heuristics are useful and can dramatically simplify a decision making process, but overall, using heuristics can lead to errors [10], [11], [12], [13]. When the use of a heuristic results in a systematic error, it is known as a cognitive bias [1]. Cognitive biases are not just limited to arbitrary examples such as estimating distance, but are prevalent in many widespread real-world contexts such as national policy, intelligence analysis, medical practices, the judicial process, and science [5].

A believed source of cognitive biases comes from the interaction of System 1 and System 2 thinking. Many contemporary models of cognition conceive of the

mind as composed of two overarching yet interconnected sets of processes, known as System 1 and System 2 [14]. System 1 thinking is fast, effortless, emotional, and is unavailable to conscious introspection. System 1 thinking excels at pattern recognition and works by association, meaning that it can grasp the essence of a situation and identify appropriate responses. It is also essential for recognizing safe foods, avoiding dangerous animals, and behaving appropriately in social situations. On the other hand, System 2 thinking is slow, effortful, conscious, logical, and is only capable of processing information sequentially. The main function of System 2 is to monitor System 1 and to identify potentially incorrect responses and then to correct those responses. Generally, the busier that people are, the more they have on their minds, and the more time constraints they face, the more likely they are to rely on System 1 thinking. This is not always a bad thing: in many situations, System 1 thinking leads to superior decision making by improving efficiency without sacrificing quality [15], [16]. However, in situations where we know that cognitive biases are likely to occur, relying exclusively on System 1 thinking is likely to lead to costly errors.

There are many cognitive biases that can significantly impact the decision making process. Some of the most prominent biases include availability bias, anchoring, and confirmation bias. The availability bias occurs when people over-estimate the probability of an event because they are easily able to recall a similar event [1]. For example, because a person is able to recall specific instances of a plane crash, they may be more likely to believe that his or her chance of getting in a plane crash is higher.

Anchoring occurs when individuals use an initial piece of information to make an estimate and fail to properly adjust their estimate in light of new information prior to making their final decision[17], [1]. Once an anchor is set, all information is judged

against the anchor. A common instance of anchoring is when negotiating a price for a car: the first price offered sets the standard for the rest of the negotiations, so that prices lower than the initial price seem more reasonable even if they are still higher than what the car is actually worth.

Last of the most prominent cognitive biases is confirmation bias. Confirmation bias is the inappropriate bolstering of a believed hypothesis in the face of uncertainty [5]. In an intelligence analysis of a cyber-attack, confirmation bias occurs if the analyst forms an initial hypothesis on which country is responsible for mounting the attack and consequently only searches for or overvalues evidence which supports their hypothesis. As seen in this example, confirmation bias is especially damaging in an intelligence analysis because it could cause the analyst to completely disregard or to misinterpret information. For this reason, this work focuses exclusively on confirmation bias.

2.2.1 Confirmation Bias

Although known at first by other names, the concept of confirmation bias has long been known to decision makers. Over 400 years ago, Sir Francis Bacon expressed that "the human understanding, when any proposition has been once laid down... forces everything else to add fresh support and confirmation..." [18]. However, confirmation bias in its current form first began to be investigated in earnest in 1960 with Peter Wason and his abstract rule discovery experiment [19]. In the experiment, the test administrator gives participants a sequence of three numbers and states that the numbers follow an unspecified rule. The participants would then attempt to determine the rule by generating a sequence of three numbers they thought fit the rule. After writing down their numbers, the test administrator provided feedback on whether the participants' proposed sequence matched the rule. At any point during

the experiment, the participants were able to declare what they believed the rule to be. In the experiment, the sequence provided to the participants was "2, 4, 6" with the unstated rule being "three numbers in increasing order of magnitude." If a participant were to guess "8, 10, 12" they would be told that their sequence conformed to the rule, while a participant who guesses "3, 2, 1" would be told their sequence does not conform to the rule. A possible, but incorrect rule, that a participant could surmise would be "sequences of even integers in increasing order." During the study, Wason found that participants generally chose to test sequences that confirmed their hypotheses, as opposed to those that did not. A second experiment performed by Wason involved a selection task [20]. In this task, participants were shown a set of four cards placed on a table. The participants were given a rule and could choose a card to test to prove the validity of the rule. Each card had a number on one side and a letter on the other side. The faces of the cards visible to the participants could show an odd number (e.g., 3), an even number (e.g., 8), a vowel (e.g., E), or a consonant (e.g., X). The rule that was then given to the participants was "If a card has a vowel on one side, then it has an even number on the other." To solve this puzzle in the most logical and efficient way, a participant should check the card with the vowel on it and the card with the odd number on it, for if either are proven false, then the rule is proven not valid. The results of the experiment showed that most participants would check the card with the vowel on it, but relatively few would choose the card with the odd number on it. These results indicated that participants preferred to resolve uncertainties that could be congruent with the rule being evaluated, but not those that couldn't be congruent with the rule.

Another area in which confirmation bias has been heavily studied is social cognition. In a 1978 experiment, Snyder and Swann performed an experiment in which participants were asked to determine if someone was an extrovert [21]. The results of the experiment

showed that participants were more likely to ask questions that, if answered in the affirmative, would confirm that the person being asked was an extrovert. In contrast, confirmation bias can be reduced by reframing the problem as a "cheater detection" question in which a social norm may be violated [22]. In a task where participants are charged with determining if a person is underage drinking, the participants were far more likely to select the positive antecedent (modus ponens is drinking) and negative consequent (modus tollens: isn't at least the legal age of drinking). Modus ponens is a mode of reasoning from a hypothetical proposition according to which if the antecedent be affirmed the consequent is affirmed (as, if A is true, B is true; A is true; therefore, B is true) [23]. Modus tollens is a mode of reasoning from a hypothetical proposition according to which if the consequent be denied the antecedent is denied (as, if A is true, B is true; B is false; therefore A is false) [23]. These experiments have led some researchers to believe that human hypothesis testing is an evolutionary trait for tracking social dynamics [24]. This theory could account for hypothesis testing's apparently poor performance in non-social problems.

If the drawbacks of confirmation bias are so readily evident, then why would we as humans ever use it? Despite the negative aspects of confirmation bias, there has been substantial research into how biased approaches to hypothesis testing could be globally optimal. Klayman and Ha proposed the positive test strategy as a way in which biased search is in fact optimal [10]. The positive test strategy is articulated as: when testing a hypothesis, people are far more likely to seek cases that are believed to demonstrate the event rather than conditions that are thought to lack the event. The positive test strategy is optimal when hypotheses are "sparse," or when a given hypothesis makes fewer positive claims than negative claims. For example, the claim "if it is a yellow fruit, then it is a banana" has only one positive consequent (that given a yellow fruit, it is a banana), but many negative consequents (given a yellow

fruit, it is an apple, a lemon, or a pear), and so it is a sparse hypothesis. If most of the hypotheses that humans deal with are sparse, then confirmatory searching, or Klayman and Ha's positive test strategy, is much more efficient at testing hypotheses than attempting to prove every possible negative consequent. Oaksford and Chater make a similar argument in that if Wason's selection task is analyzed using conditional probabilities, then then expected information gain by performing a confirmatory test (checking to see if the card has a vowel) is greater than the falsification tests (checking to see if the card has an odd number) [11]. This property holds true as long as the probability of the positive antecedents and consqeuents are low. Another argument in favor of the positive testing strategy is from the viewpoint of the expected utility of hypothesis test. Friedrich argued that hypotheses are generally made about positive outcomes (e.g. when I eat food, I am no longer hungry) [12]. If a hypothesis is about a positive outcome, then testing that hypothesis in a confirmatory manner is optimal. For example, if I am hungry and am only interested in what makes me not hungry, then it is more important that I become not hungry than whether I learn that food truly does not make me hungry. Hypothesies testing, Friedrich says, not only is designed to seek the truth, but is also designed to avoid "costly errors." When observed from this viewpoint, rejecting hypotheses that are falsely true is of a low priority. What is more important is gaining the benefit of an object or an action if the relationship is in fact true. Because humans are living creatures who constantly need resources for survival, short-sighted opportunistic confirmatory hypothesis testing is rational.

Whether confirmation bias is rational or not, it stills presents a problem in that it can lead to the persistence of unsupported beliefs. Research has shown that confirmation bias is optimal in some situations, but why does it occur in non-optimal situations as well? First, evaluating evidence in relation to a hypothesis requires

comparing the evidence at hand to hypothetical data held in the mind. Such a setting is known as a conditional reference frame [25]. When evaluating evidence that lends itself to multiple varying hypotheses, it is necessary to represent these varying hypotheses in such a way that they can be updated upon the arrival of new information. Representing hypotheses in the mind in this way places considerable demand on any memory processes that are involved in representing the hypotheses. While humans have a great capacity for long-term memory storage, this long-term memory is not well-suited to deal with the dynamic updating that is required to properly update hypotheses during testing. Instead, to evaluate hypotheses, humans use short-term memory processes, known as working memory [26], [27]. This working memory has an extremely small capacity for information - most estimates range from 3 to 7 items [28], [29], [30]. Additionally, it's thought that although humans can store 3-7 items, really only one item can be "used" at a time [31]. For these reasons, reasoners tend to evaluate possible hypotheses in isolation. Because of this, an observation that is seen under two varying hypotheses may be taken to strengthen the hypotheses being considered [26].

By evaluating hypotheses individually, hypotheses are not able to compete against one another. Not only does this hinder one's ability to determine how to interpret a new observation, but it can lead to a selection bias in how one selects information. Examining a set of evidence with a specific hypothesis in mind leads to a selection of information that is "pseudodiagnostic", or information that allows one to know whether a given set of observations is likely or not under a pair of hypotheses, but does not indicate which of the hypotheses are more likely [32]. Pseudodiagnostic information can exacerbate the confirmation bias by allowing the collection of incomplete data. In fact, Fiedler argues that incomplete sampling of information by itself can lead to cognitive biases [33]. Thus, the selection of information is critically important

in determining the balance of hypothesis evaluation.

2.3 Confirmation Bias Mitigation

Because of the dangerous effects that confirmation bias has on the decision-making process, there has been a significant amount of research into techniques on how to mitigate these effects. These mitigation techniques fall into two main categories: modifying the decision maker (DM) and modifying the DM's environment [34].

Modifying the DM seeks to provide people with some combination of knowledge and tools to help them overcome their own limitations and dispositions. Techniques such as raising the accountability of the DM, having the DM consider an opposing viewpoint, improving the DM's education on confirmation bias, training the DM on how to deal with confirmation bias, providing warnings and feedback about confirmation bias, and having the DM go through a series of checklists and analysis have all been proven to be partially or fully effective at mitigating the negative effects of confirmation bias [35], [34], [36], [37], [38], [39], [40], [41], [42], [43], [44]. Despite the positive results of mitigation through modifying the DM, there are some issues with this technique. Most of these techniques rely on the DM to debias themself, but people naturally resist being biased for many reasons. Most people don't want to be told that they have been "doing it wrong" for their entire lives and they don't want to relinquish control over the decision process. Most importantly, people fail to understand the benefits of many debiasing techniques relative to their own abilities, not just because they are overconfident, but because the techniques themselves are alien and complex, and the benefits are noisy, delayed, or small [38], [45].

The second category of mitigation, modifying the environment, attempts to take the DM out of the mitigation equation. Modifying the environment seeks to alter the setting where judgments are made in a way that either encourages better strategies

or is a better match for the decision strategies that people naturally apply. It accepts that there is a bias, but it strives to create situations in which a bias is either irrelevant or may be helpful [34]. The techniques included in this category are conveying social norms, providing nudges, settings defaults, providing easy models to work through, forced breaks, and forced planning [46], [47], [48], [49], [50], [51], [52]. These techniques have all been proven effective, and because they provide an environment in which the mitigation technique is removed from the DM, they do not fall prey to the same issues as modifying the DM techniques.

In the modifying the environment category, nudges are perhaps the most effective debiasing technique. A nudge is a design choice that does not restrict a choice, but instead makes use of psychological principles to influence behavior for good [46], [49]. Nudges include a broad range of techniques and include methods such as setting a default option, kind representations of information, and nudges to induce reflection. One possible source of confirmation bias is from an under-reliance on System 2 thinking or from an over-reliance on System 1 thinking. By encouraging reflection, an under- or over-reliance can be reduced by requiring people to devote more time and attention to a decision. In this work, the debiasing technique used will be a nudge.

2.4 Visual Search: Attention

In cognitive psychology, the word attention is synonymous with information selection. Because there is far more information flowing through the senses than humans are able to handle, only a certain subset of available information can be analyzed at a time. Determining which subset gets chosen at any given time has long been an area of intense research [53], [54], [55]. The subset that gets selected for processing is often controlled by relatively simple aspects of its ability to stand out,

such as a sudden onset, movement, or its uniqueness [56], [57], [58]. Most importantly for confirmation bias, the subset can be selected by the intentions and goals of the searcher [59], [60]. This ability to adjust the processing of sensory information due to intentions and goals is a very important component of all goal-oriented behavior - if humans were not able to restrict the flow of information then we would simply be overwhelmed and would never be able to accomplish any tasks [61]. In order to properly control one's behavior while attempting to accomplish a goal, natural and necessary attentional mechanisms must ensure that the information needed to choose the correct actions is available for decision making. For example, when driving a car on the highway and attempting to change lanes, the driver must know where the other cars are and their trajectory; knowing the other cars' colors and passenger count is not important to accomplish the goal of changing lanes.

During goal-driven behavior, humans often seek very specific information as it relates to their goal. For example, an owner feeding his dog in the morning needs to be able to find the dog food. To find the dog food, he might putter about the kitchen, testing hypotheses for where the dog food is. His hypothesis of "is the dog food in the pantry?" would guide his subsequent information seeking behavior, leading him to open the pantry door and search the shelves until he found the food. Any object that he fixates upon is compared to his question of "is this dog food?" If it is not, it is unlikely that this object remains in his attention for very long and is most likely immediately forgotten. Seen this way, the controlled selection of information in vision, known as top-down visual attention, can be described as perception that is guided towards the goal of verifying or falsifying some hypothetical state of the perceiver's environment. In fact, the most often used laboratory task in the study of visual attention is visual search [62], [63], [64], [65].

In the usual visual search task, the participant is presented with an array of

discrete stimuli and is asked to determine the presence of one particular type of stimulus in the given display. Participants must then search the array of stimuli to determine whether the target is present or absent. Normally, the stimulus that is being searched for is known as the "target" while all other stimuli in the display are known as "distractors." Because the term "target" could mean the specific stimulus present in the display, or the more abstract description of a target, the term "target template" is helpful when referring to the description of the target during a visual search. This template can be used to select and process the aspects of the visual display that are most likely to yield information about a target [65]. While it is not common to refer to a template as such, in order to link a confirmation bias with visual search, it is useful to think of the template as a visual hypothesis.

Many theories of the top-down mechanisms used for visual selection in visual search propose that information processing is biased towards goal-relevant information. Similar to hypothesis testing, core information processing units have a limited capacity [26], [27]. The theories of Guided Search, Feature Integration, Visual Attention, the Boolean Map Theory of Visual Attention, and Biased competition all conclude that visual search is guided, biased, or is otherwise prioritized towards template matching stimuli [65], [66], [64], [67], [68], [69]. These theories do not make it clear whether the stimuli is prioritized through applying gain to template-matching stimuli or whether it is through suppression of template-mismatching stimuli. Again, similar to hypothesis testing, research on working memory during a visual search has led some researchers to conclude that visual selection can only be guided by a single template at a time [70], [42]. Visual search theories determine that a visual search performed in this way is more economical as it prioritizes aspects of the visual search that are task-relevant. However, prioritizing stimuli that are similar to a target template combined with the fact that only a single template can be maintained at any given time is theoretically

sufficient to produce a confirmation bias in a visual search. Looking at a target template as a hypothetical visual state, it will take long enough to gather evidence for an alternative visual state that by the time this information is gathered at all it will be incongruent with features in the template. This type of visual search is confirmatory in nature; any information that supports the presence of a goal is going to receive increased importance in working memory. Additionally, in the case where crucial visual information is brief in nature, the alternative states may never reach awareness in a top-down guided search. A search where this occurs has been demonstrated in studies of inattentional blindness, where conspicuous events go unnoticed because one is engaged in a demanding visual task [71], [72]. For example, observers of a game of basketball are easily able to focus on a template of people playing basketball while completely ignoring information that is grossly inconsistent with this template, such as a gorilla walking through in the middle of the game.

The current evidence supports the possibility that top-down visual selection mechanisms automatically lead to confirmatory searching. However, the design of most visual search tasks encourage confirmatory searching as a useful and efficient strategy. The typical visual search task is for a user to report whether a target is present or absent in a given array of stimuli, any one of which could be the target. In these tasks, where a target is either present or absent, the non-targets provide no information about the presence or absence of the target. Therefore, in order to find out if the target is present, the user must search the entire array of stimuli until they either find or do not find the target. This is a confirmatory search, as the proposition "there is a target" can be checked in less time than it can be proven wrong because the detection of the target can occur before the entire array has been searched. The falsification of the proposition can only occur once the entire array of stimuli has been checked and the target has not been found. Therefore, it is unclear if the task of visual selection

is confirmatory by its nature, or whether confirmatory selection is simply adopted as the most useful strategy for visual search.

Even if a confirmatory search is useful, it is often not the most strategic and efficient method of search. It is possible that humans use a confirmatory search pattern because they are unaware of other, more strategic, methods of search. An experiment designed to test whether participants, once taught a more strategic method of searching, used a more strategic search strategy in place of a confirmatory search strategy was undertook by Rajsic, Wilson, and Pratt [73]. Participants were given a circle of eight letters consisting of b's, d's, q's, and p's and told to search for one of the possible letters, with only one occurrence of the target appearing in each trial. The letters were one of two different possible colors. In each trial, participants were told to press a certain key if the letter matched the template color given and to press a separate key if the target did not match the template color. The results from their first experiment showed that most of the participants did in fact perform a confirmatory search by searching each possible template-matching color before moving onto the template-mismatching colored letters. This confirmatory searching occurred even when a confirmatory search was not efficient, for example, when 6/8 of the letters were template-matching and the other two were template-mismatching. In this example, an efficient and strategic searcher would search the 2 template-mismatching colored letters as they could determine the color of the target letter quicker than by searching the 6 template-matching letters. To test whether the participants were unaware of this more strategic searching strategy, the test administrators taught the participants the strategic method of search and then retested the participants. Even after being taught the strategic search method, the majority of participants still elected to perform a confirmatory search, showing that humans elect to perform confirmatory searches over more strategic and efficient search methods.

In a follow on experiment, Rajsic, Wilson, and Pratt hypothesized that humans use a confirmatory search because the cognitive cost of using a strategic search is too high while the cost of using a confirmatory search is low.[74]. They tested whether participants would still choose a confirmatory search pattern even if the cost of using such a search was raised. To implement this raised cost, eight colored circles were displayed arranged in a circle. Similar to the first experiment, the participants were told to search for a target-letter and to press different keys if the letter appeared on a template-matching circle versus a template-mismatching circle. However, using gaze-tracking equipment, the letter would only be displayed on the circle if the participant fixated upon a specific circle. Under these search conditions, the participants were able to prioritize their search towards the color with the smallest amount of circles present, thus reducing the confirmation bias present in visual search. Therefore, by raising the cost of performing a confirmatory search, confirmatory search can be reduced in a visual search.

2.5 Electrophysiological Measurements

Electrophysiology involves voltage or electric current changes as it occurs in the human body [75]. It includes measurements such as EEG, Electrocardiography (ECG), and galvanic skin response (GSR). Each of these measurements provides insight into both physiological and neurological activity inside the human body and can help in understanding confirmation bias. In addition to utilizing traditional measures, Minas et al. used electrophysiological signals such as EEG and Electro Dermal Activity (EDA) to detect and measure confirmation bias [76].

2.5.1 Electroencephalography

EEG is the measure of electrical activity in the brain. Due to its high sensitivity, EEG can be particularly useful over behavioral measures in distinguishing various cognitive processes [77]. In a research study in 2014, Minas et al. correlated the activation of the right frontal cluster of the brain with the presence of hypothesis confirming information when compared to disconfirming and irrelevant information [76]. In this study, participants were given the task of selecting three out of five applicants to admit to a university. During the experiment, participants wore an EEG headset. The participants were given incomplete information about the five applicants and were asked to make an initial decision about whom to admit based on the incomplete information that they had received. After the first initial decision, participants were informed that they would be working as part of a team and would use a text-based discussion tool to discuss the applicants. After this team discussion, the participants would then make a second decision on whom to admit. The participants were informed that each team member had received incomplete information and each team member had information that was unique only to that specific team member, thus it was important to share any unique information that the team member had and to carefully consider any new information. However, the "discussion" was in reality a team simulator that played a 12-minute prepared script. The results from the experiment indicated that the participants processed the information that they received from the other team members differently depending on whether it supported or challenged the decision that they had made previously. Receiving both supporting and challenging information activated the frontal, temporal, and occipital regions of the brain which suggested that significant cognitive resources were spent on figuring out what the information meant and how it related to the participant's pre-decision. However, only information that supported the participants pre-decision

activated the right frontal cluster which indicated that the working memory of the participant was working, and that the participant was comparing the new information to information already known. Additionally, preference-supporting information also triggered increased emotional arousal as indicated by increased skin conductance. These results suggest that once participants determined that the new information challenged their pre-decision, they found the new information less interesting and did not think about it further.

Before Minas et al. correlated the activation of the right frontal cluster when presented with preference-supporting information, previously the only way to determine a confirmation bias was through behavioral measures.

2.6 Machine Learning

Machine learning is where computers learn from data to achieve some task [78]. One classic example of machine learning is an email filter that marks incoming email as spam. A spam filter is a type of machine learning model that can take an email as an input and return whether the email is spam. For a model to successfully learn it requires data on which to learn. The data for this particular example takes the form of emails - both real and spam. The users then partition the data into training, validation, and test sets of data. The training set is the set of data from which the model learns patterns that are given certain labels. In the example of a spam filter, the model may learn to associate certain senders as being spam, or it could associate certain key phrases such as "free trial" or "sale" in an email as belonging to a spam email. After a model has been trained to identify spam emails, the model is then tested on a set of validation data to see how well the model is performing. In this case, the validation data would consist of previously unseen emails that are then marked as spam or not spam. The model's performance on the validation data can

then be used to tune hyper-parameters in the model so that it performs better in the future. After finishing training and validating, the model is tested for its true performance on the test set. This test set has not been previously seen or used by the model. The test set's purpose is to assess how well the model will perform in the real world.

Machine learning problems fall into two wide categories: supervised or unsupervised learning. In supervised learning, a label is available that details the "truth" about the data. An email spam filter is an example of supervised learning because an email is known to be either spam or not spam. Unsupervised learning contains no label or "truth" about the data. Additionally, supervised machine learning problems can be sorted into two more categories: regression or classification problems. The goal of a regression problem is to generate some type of numerical value. An example of a regression problem would be to predict the price of a stock after observing its behavior for the past week. A classification problem's goal is to predict to which category the data belongs. For example, predicting whether a stock's price will increase or decrease is a classification problem.

2.6.1 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is a classification method that linearly separates classes [79]. There are several reasons to use LDA over other linear models: it is stable when the differing classes are well-separated, it is stable when there are only a small number of observations, and it is able to be used for more than just two classes. LDA attempts to approximate a naive Bayes classifier by using the assumptions that every class has an approximate Gaussian distribution with a class-specific co-variance matrix. A naive Bayes classifier is a simple probabilistic classifier that assigns each observation to the most likely class, given its predictor values. It simply assigns a

test observation with predictor values x_0 to the class j for which $Pr(Y = j|X = x_0)$ [79]. The above assumptions lead to the discriminate function below:

$$\hat{\delta}_k(x) = x \cdot \frac{\hat{\mu}_k}{\hat{\sigma}^2} - \frac{\hat{\mu}_k^2}{2\hat{\sigma}^2} + log(\hat{\pi}_k)$$

$\hat{\mu}_k$: mean of class k

$\hat{\sigma}^2$: weighted average of sample variances for each of K classes

$\hat{pi}_k$: proportion of training observations that belong to class k.

LDA calculates the probability that an observation x belongs to each class and places x in the class of which it has the highest probability of being a member. An example of an LDA classifier is shown in Figure 1.

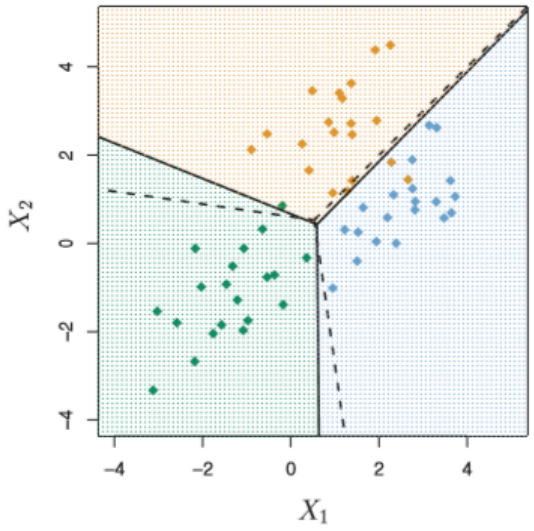

Figure 1. An example of an LDA classifier with three classes. The observations from each class are drawn from a multivariate Gaussian distribution with $p = 2$, with a class-specific mean vector and a common covariance matrix. 20 observations were generated from each class, and the corresponding LDA decision boundaries are indicated using solid black lines. The Bayes decision boundaries are shown as dashed lines [79].

LDA models are similar to Quadratic Discriminant Analsysis (QDA) models, however LDA models are much less flexible than QDA. LDA models make the assumption

that the predictor variables share a common variance across each response. This leads to a lower variance in LDA models, which can be beneficial if there are relatively few training observations.

LDA is generally not used for machine learning with EEG classication. However, it worked generally well in the work of Binias et al. In Binias et al.'s work, an LDA was used to determine whether EEG signals came from an airplane pilot's brain before or after an specific event occurred [80]. In classifying the pre- and post-event brain signals, LDA performed the second best out of all models with a mean accuracy of 73.01%. The LDA model outperformed all other models except for an artificial neural network's performance.

2.6.2 Random Forests

Decision trees are fairly common methods of making a decision. The branches of a decision tree split at a discrete point with the following path accessible by only following one side of the split. Random forest classifiers are an ensemble machine learning method that use large numbers of decision trees that operate as its ensemble. [78]. Ensemble methods are classifiers that classify using the outputs from multiple various algorithms, which in the case of a random forest are its decision trees. The theory behind the success of random forests is that a large number of uncorrelated models operating as a committee will outperform any of the individual models by themselves [81].

Random forest classifiers do not normally yield high-performance results when classifying EEG [82]. However, by their very nature random forests identify the features that are most important in a dataset and can identify features that allow for distinction between the classes of a dataset. The use of random forests in an EEG binary classification problem obtained an accuracy of 75% when classifying brain

activity during concentration and meditation [83]. This performance is not state of the art, however it is high enough to be able to identify the important features that differentiate the two classes.

2.6.3 Artificial Neural Networks

An Artificial Neural Network (ANN) is another approach to a machine learning task. These models were inspired by the biological makeup of the brain in which neural networks propagate signals and information [84]. ANNs are proficient at learning complex relationships amongst the data and are capable of expressing these complex relationships as more simple relationships.

2.6.3.1 Fully Connected Neural Networks

The first ANN that was developed is the fully connected neural network. In these networks, each neuron, or unit, is fully connected to every other neuron in the subsequent layers. Each layer in an ANN receives inputs, multiplies these inputs by a set weight, and then passes the weighted sum of the inputs through an activation layer. The output of the n^{th} layer is:

$$x_n = f(W_n^T x_{n-1} + b_n)$$

f: non-linear activation function

x_{n-1}: the input to the n^{th} layer

W_n: the matrix of weights that describes a mapping from x_{n-1} to x_n

b_n: vector of biases.

An ANN learns by modifying the parameters of a model until the network can correctly map an input to the desired output. An example of an ANN can be seen in Figure 2.

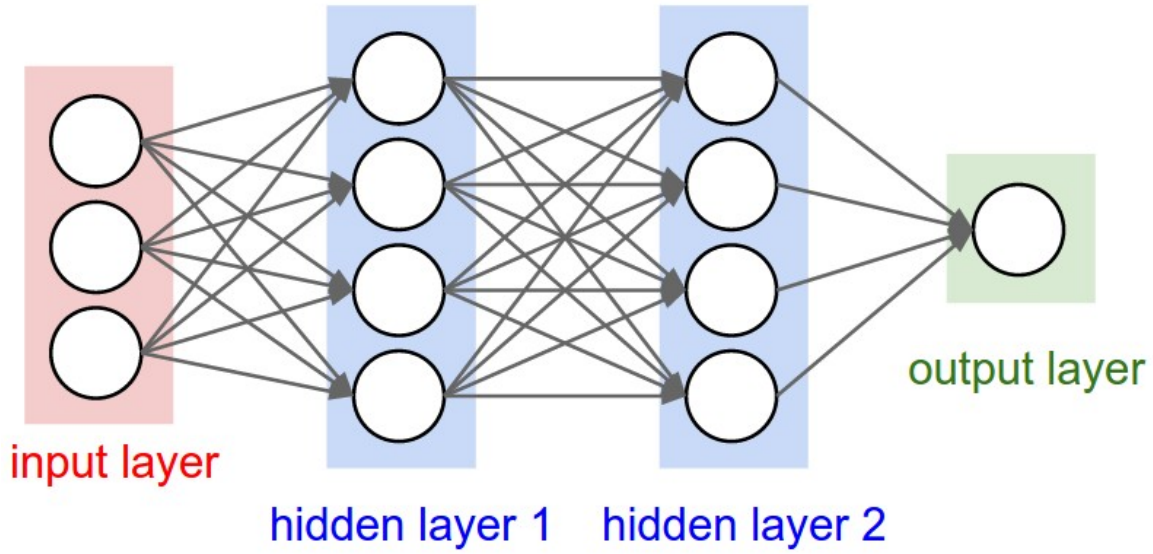

Figure 2. A simple example of an ANN. This fully connected ANN has two inputs, two hidden layers, and a single output.

ANNs generally perform well when classifying EEG data. In an attempt to classify operator workload via EEG data, Wilson et al. used a single, 43-node, fully-connected ANN [85]. In this task, eight participants performed NASA's Multi-Attribute Task Battery (MATB) at one of the three levels of workload: baseline, low, or high. EEG data was collected over three five-minute sessions during the course of a single day. Each session corresponded to one of the three levels of workload. Once collected, the EEG data was processed by using a fast Fourier transform (FFT) to transform it the frequency domain so that the average power could be computed. The EEG bands used included delta (1-3 Hz), theta (4-7 Hz), alpha (8-13 Hz), beta (14-30 Hz), and gamma (31-42 Hz). The network achieved a mean classification accuracy of 85.0% on the baseline level, 82.0% on the low workload level, and 86.0% on the high workload level. Further work in the same area by Christensen et al. showed that ANNs outperformed both LDAs and Support Vector Machines when classifying workload level with EEG [86].

2.6.3.2 Convolutional Neural Networks

A Convolutional Neural Network (CNN) is a neural network that is able to learn spatial patterns in the input data [84]. There are three main layer types used when building a CNN: the convolutional layer, the pooling layer, and the fully-connected layer. In the first layer, the convolutional layer, the input is convolved using a set of kernels. An activation map is then produced by applying an element-wise application function. The next layer, the pooling layer, downsamples the spatial dimensions of the activation map. The last layer, the fully-connected layer, calculates the scores of each class and classifies the input. An example of a CNN can be seen in Figure 3.

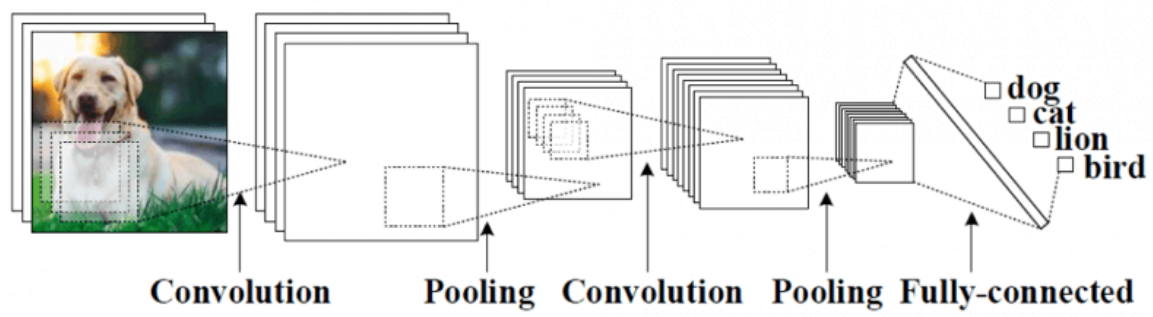

Figure 3. In this CNN, an image of an animal is taken as an input and the output is the type of animal [87].

Regarding classifying emotions from EEG data, CNNs have been shown to outperform other machine learning methods. Tripathi et al. used a CNN to classify human emotion using EEG data fom the DEAP dataset [88]. The DEAP dataset consists of 40-channel EEG data recorded from 32 participants as they watched 40 one-minute clips of music videos. The participants completed a self-assessment and scored themselves on arousal, valance, and dominance for each music video. The CNN model achieved an accuracy of 81.4% and 73.4% for the binary classification of the valence and arousal levels of the participants, and an accuracy of 66.8% and 57.6% for the

three-class classification of valence and arousal levels.

2.6.3.3 Recurrent Neural Networks

A Recurrent Neural Network (RNN) is a form of a neural network that is able to learn long sequences and their dependencies on each other by maintaining a state or a memory. This memory is maintained by a recurrent connection to itself which allows the model to process both the current input as well as previously seen inputs [84]. An example of a simple RNN can be seen in Figure 4. The major problems with simple RNNs is that they are a victim of the vanishing gradient problem. The vanishing gradient problem is a problem in which the gradients that are seen towards the end of the model become extremely small as they are back-propagated to the beginning of the model. The effect is that the model is unable to retain information about inputs seen "a long time ago" and is therefore unable to learn long-term dependencies. The Long Short-Term Memory (LSTM) model was created in order to solve this problem. LSTMs contain a separate channel in which important information is stored so that it is able to be used in learning long-term dependencies [84]. An example of a simple LSTM can be seen in Figure 5

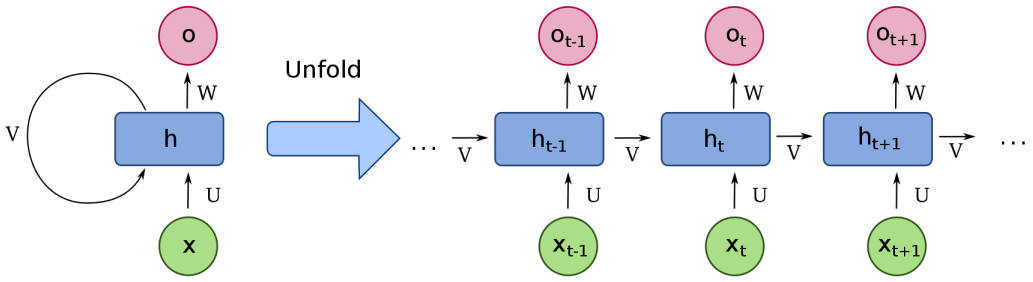

Figure 4. This is an example of a simple RNN.

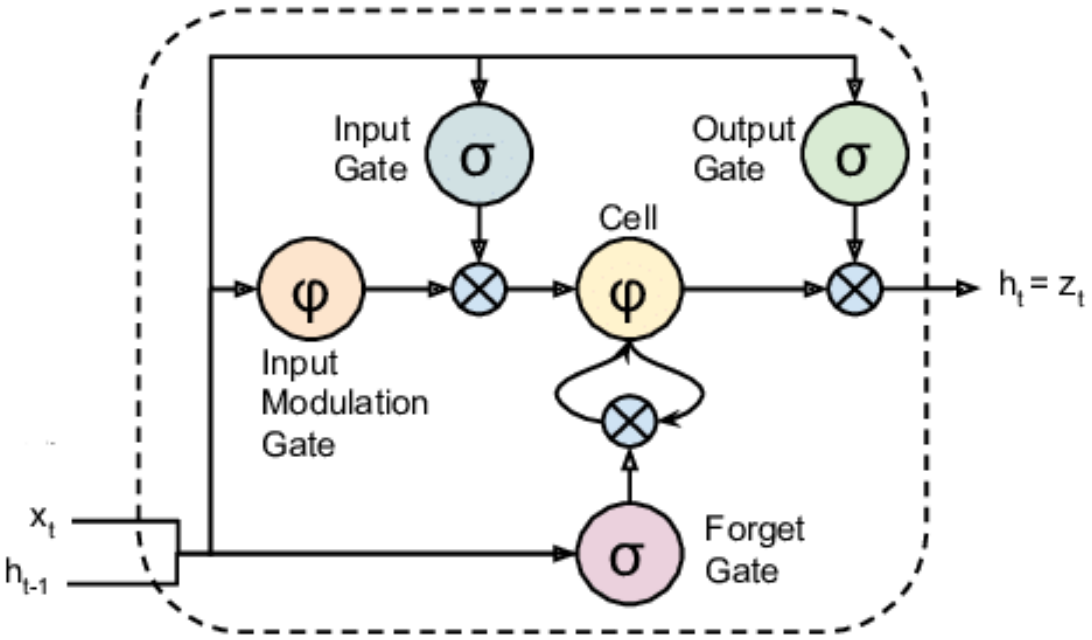

Figure 5. This is an example of a single LSTM [89].

RNNs are well-suited for machine learning problems in which time-ordered information is important or can lend clues as to the current state. Because EEG data is temporally organized, RNNs have been shown to outperform other machine learning models in classifying EEG signals. In classification of six different hand motions from a grasp-and-lift experiment, an RNN obtained an accuracy of 94.8% in classifying which motion was being performed [82]. This was an improvement of 23.5% over other machine learning methods. RNNs have also been shown to obtain the lowest test error of mental load classifications ; the addition of LSTM layers in a CNN reduced the test error of a four-class mental load classification by 21.5% [82].

Due to its ability to learn long-term dependencies, LSTMs are perhaps more powerful than simple RNNs when it comes to classification and EEG. Hefron et al. created a multi-path convolutional recurrent neural network (MPCRNN) that consisted of CNNs and LSTMs [90]. This network achieved a cross-participant accuracy

of 86.8% in classifying low and high workload from EEG data. This performance outperformed CNNs and LSTMs by themselves. These results suggest that RNNs may perform well in classifying EEG signals.

2.6.3.4 Temporal Convolutional Networks

A Temporal Convolutional Network (TCN) is a specific type of CNN that has faster training times and longer memory than traditional RNNs when modeling a sequence [91]. The general structure of a TCN can be seen in Figure 6. The first layer of a TCN is the dilated causal convolutional layer which is a standard convolutional layer with a dilated kernel. Using stacked, exponentially increasing, dilated convolutional layers causes the receptive field of the model to increase exponentially while the number of parameters increases linearly. The residual blocks used by a TCN consist of a dilated one-dimensional convolutional layer with a 'causal' padding, a weight normalization layer, a rectified linear unit (ReLU) activation layer, and a spatial dropout layer which repeats for as many blocks are present. The various layers of a residual block contained in a TCN can be seen in Figure 7.

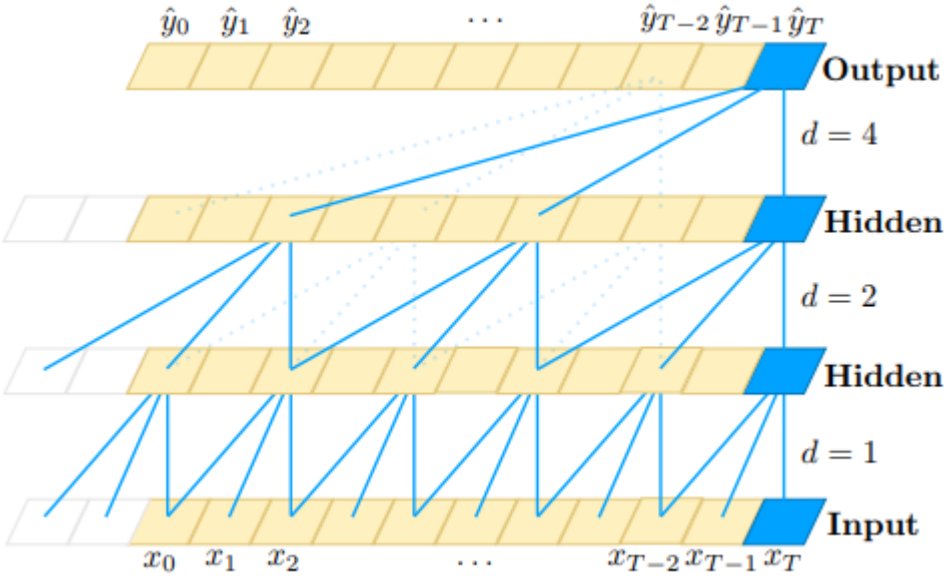

Figure 6. The architecture of a TCN.

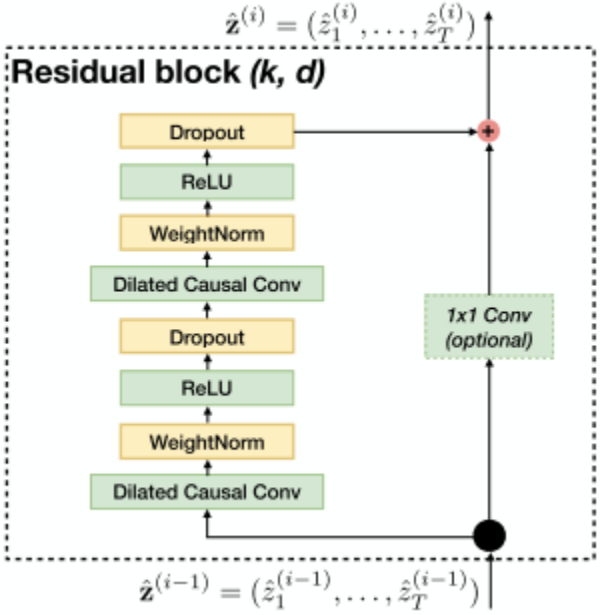

Figure 7. A residual block [91].

2.7 Summary

Confirmation bias is a prevalent cognitive bias in decision making. This bias results in errors in decision making by inappropriately bolstering a believed hypothesis. Confirmation bias can also affect visual search by restricting the searcher's attention to one visual hypothesis. Mitigating a confirmation bias during a visual search is possible but it is unknown how such a mitigation would affect the searcher's search patterns. Finally, current research indicates that confirmation bias can be detected through the use of EEG data.

III. Methodology

3.1 Chapter Overview

This chapter describes the outline of a human-participant visual search experiment and the process used to analyze the recorded data. First, the chapter discusses the research questions and hypotheses. Next, a description of the experiment which includes participant demographics as well as details about the various factors and variables present in the experiment is presented. This is followed by a description of how the results will be analyzed. This section will include details about the statistical tests to be performed as well as the machine learning approach used. Finally, a summary of this chapter is provided.

3.2 Background

Humans often default to a confirmatory method of searching because it generally requires less mental processing, it is a simple way to search for a single target, and it is often the most efficient way to search [10], [65], [73]. Because it is simple to perform, humans will perform a confirmatory search even when a confirmatory method of searching is not the most efficient way to search [74]. However, because it is still possible to perform a confirmatory search while also being efficient and because it is possible to perform a non-confirmatory search while being inefficient, this research focuses on encouraging efficient searches rather than discouraging confirmatory searches.

3.2.1 Previous Work

The experiment conducted to answer these research questions was an extension of Rajsic, Wilson, and Pratt's 2015 and 2017 experiments [73], [74]. As an overview, Rajsic's 2015 experiment was a visual search task in which participants were presented

with eight letters arranged in a ring. The letters on the screen could be one of two colors. Participants were instructed to search for a specific letter, called the target letter, and were given an example of what color the letter could be, or the template color. The target letter would appear only once amongst the other letters. The participants were to press a certain key if the target letter's color matched the template color and to press a separate key if the target letter's color did not match the template color. An example of Rajsic's 2015 experiment can be seen in Figure 8

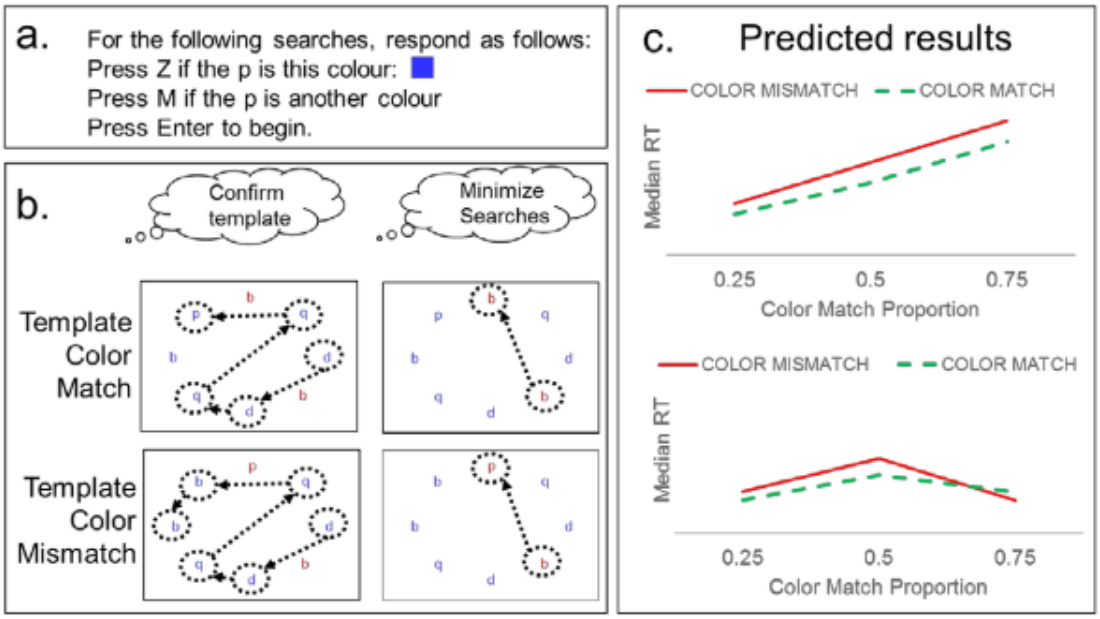

Figure 8. Rajsic's 2015 experiment. The instructions presented to the subject, along with a template color match and mismatch, and the predicted results can be seen.

The results of Rajsic's 2015 experiment showed that participants performed a confirmatory search by searching all of the template-matching colored letters first, even when there were more template-matching colored letters than template-mismatching colored letters.

In Rajsic's 2017 experiment, a cost was introduced to attempt to force the participants

to perform a non-confirmatory, or an efficient, search. This cost came in the form of increasing the time necessary to search for the target letter. The experiment was identical to the 2015 experiment except that instead of letters being present on the screen, there were now eight colored circles. To make a letter appear the participant would have to visually fixate on the circle. Under these search conditions the participants were able to prioritize their search towards the color with the smallest amount of circles present, thus reducing the confirmation bias present in the visual search. An example of Rajsic's 2017 experiment can be seen in Figure 9.

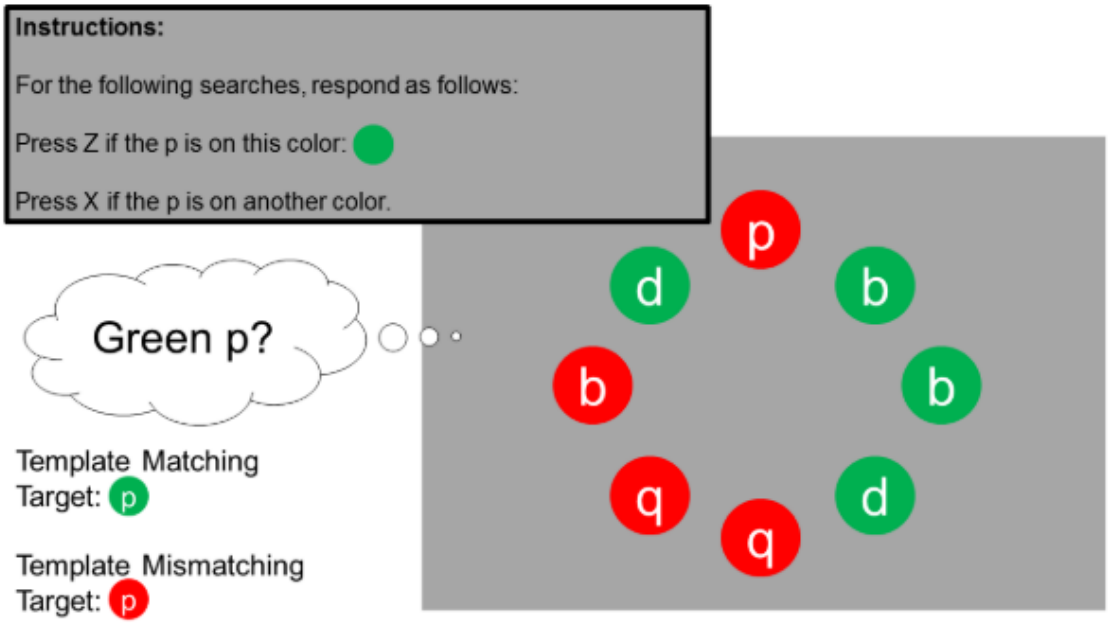

Figure 9. Rajsic's 2017 experiment.

3.3 Research Questions

The objective of this research is to determine whether an inefficient search during a visual search can be detected and subsequently mitigated. To complete this objective, the following research questions are investigated.

3.3.1 Research Question 1 - Categorizing Natural Behavior

What visual search patterns do participants naturally use during a visual search task?

Hypothesis: The majority ($> 50\%$) of participants will naturally resort to an inefficient Visual Search Pattern (VSP).

3.3.2 Research Question 2 - Behavior Detection

Can physiological signals such as Electroencephalography (EEG), Electrooculography (EOG), and Electrocardiography (ECG) be associated with an efficient visual search?

Hypothesis: Physiological signals can differentiate a participant performing an efficient visual search from a participant performing an inefficient visual search.

Research Objective: Develop a machine learning model that receives physiological data and is able to determine an efficient visual search with an equal-class-weighted classification accuracy of greater than 50%.

3.3.3 Research Question 3 - Behavior Mitigation

For a participant who is performing an inefficient search, can mitigation techniques change the participant's search patterns to an efficient search pattern that will persist for the remainder of the search tasks?

Hypothesis: By applying the mitigation techniques of a nudge, a hint, and by teaching the participant how to perform an efficient search, a participant will perform an efficient search pattern for the remainder of the search tasks.

3.4 Experiment

In Rajsic's experiment, the cost was present throughout the duration of the experiment. However, to answer the questions posed by this research, a more dynamic approach was needed. The experiment conducted in this research differs from Rajsic's experiment in that the added cost of displaying a letter when its circle is fixated upon, which this research calls a nudge, is only present if the participant is performing an inefficient search.

3.4.1 Stimuli

This experiment is known as the Efficient Search Experiment (ESE). In the ESE, a 1920 x 1080 pixel display monitor was used. Each trial consisted of eight circles that are spread evenly around the perimeter of an imaginary circle that is centered on a fixation cross in the center. The circles were 100 pixels in diameter, were positioned 360 pixels away from the fixation cross, and are separated by 45° of arc. Centered inside each circle is a white-colored letter in Arial font. The letter is always one of four similar lowercase letters (p, q, b, or d). These letters are rotation- and reflection-isomorphic of each other and were chosen to reduce the chance that the target letter was easily distinguishable amongst the other stimuli. The circles could be any of three pairs of colors: blue and orange (color set 1), green and red (color set 2), or purple and yellow (color set 3). For specific colors, see Table 1. In an effort to reduce the confusion of the participant, the two colors of a color combination would always appear together and the target color would always be the first color of the pair. All of the color combinations can be seen in Figure 10.

Table 1. RGB Values for Circle Colors

	Color 1	Color 2
Blue/Orange	(0, 18, 165)	(255, 148, 0)
Green/Red	(15, 148, 0)	(255, 0, 0)
Purple/Yellow	(99, 0, 165)	(244, 174, 52)

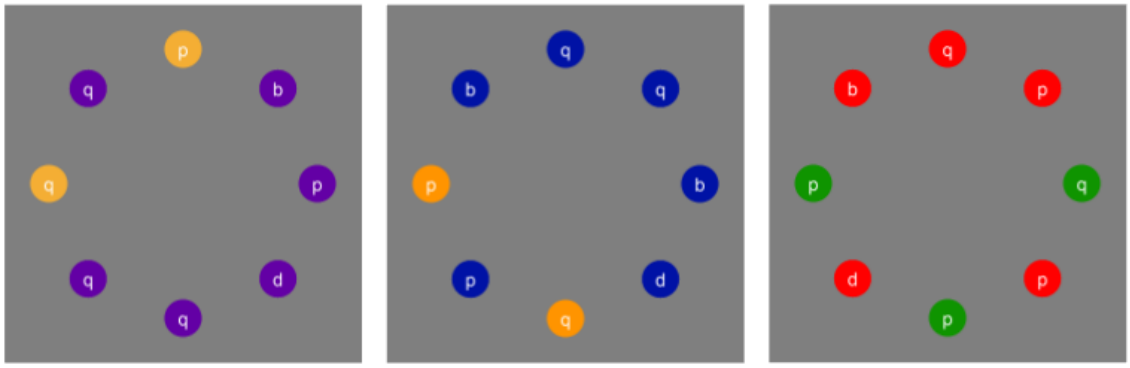

Figure 10. The color combinations used in the ESE: purple and yellow, blue and orange, and green and red.

Throughout the experiment, the participant's gaze was recorded using the Smart Eye Pro gaze tracking system. This gaze tracker updated at a rate of 60 Hz, the same refresh rate as the experiment stimulus screen. Before the experiment, participants were calibrated with the Smart Eye Pro system to ensure accurate gaze tracking. Additionally, the participants wore physiological recording devices that captured EEG, ECG, galvanic skin response (GSR), and EOG data. For more details about the equipment listed here, please see Section 3.4.7.1.

3.4.2 Procedure

The experiment began with a pre-experiment questionnaire (details in Section 3.4.8.2) intended to assess the participant's physical and mental state to ensure that the

participant was ready to undertake the experiment. Each experimental session consisted of 480 trials, partitioned into 24 blocks of 20 trials each. Before the experiment began, participants were shown the welcome screen that instructed the participant to press [space] when instructed. Immediately afterwards, a two-minute EEG baseline assessment was collected. During this time, a screen which simply displayed "Baseline" was displayed. During this baseline, participants were instructed to sit quietly with their eyes closed so that a baseline reading of their EEG signals could be recorded. At the beginning of each block, participants were shown an instruction screen that defined the target letter and the target color for that entire block. The target letter was chosen randomly from the available letters and only one instance of the target letter would appear in a trial. The colors of the circles rotated on a fixed cycle through color sets 1, 2, and 3. To help the participant remember the color that the participant was searching for, the target color would only ever be the first color of the color sets (blue, green, or purple). The target color would never be the second color of the color sets (orange, red, or yellow). An example of the instruction screen can be seen in Figure 11. In this example, the target letter is a "d" and the target color is blue. The participant is to type [c] if the circle that contains the target letter is blue and is to type [z] if the circle is another color. Once the participant has read the instruction screen for the block, the participant typed the target letter ("d" in this example) followed by typing the target color ("blue" in this example).

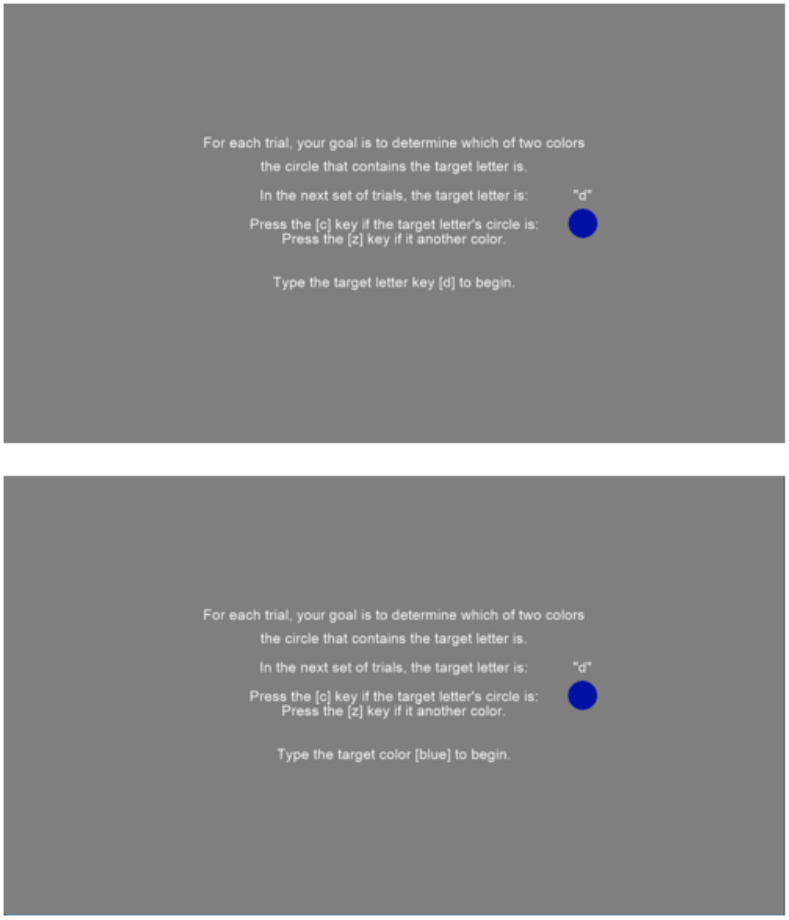

Figure 11. An example of an instruction screen displayed before a block of trials. In this example, the target letter is "d" and the target color is blue. The participant is to type the [c] key if the circle that contains the "d" is blue and is to type [z] if the circle is not blue. To continue on from the top screen, the participant must type the target letter "d." The participant now must type the word "blue" to continue.

After the block instruction screen has been displayed and the participant has typed in the correct target letter and target color, the trial instruction screen was displayed. The trial instruction screen reminded the participant that the circles and letters will not appear until the fixation cross has been fixated upon. The trial instruction screen also informed the participant of which trial they were on, and instructed the participant to press the "space" bar to start the trial.

Once the trial had started, only the fixation cross was present on the screen. The cross measured two cm by two cm and was centered in the screen. After 0.5 second, the participant's gaze was evaluated to ensure they were looking at the fixation cross. Once the participant had fixated on the fixation cross for three frames, or 80 milliseconds, the search stimuli was displayed and the fixation cross was hidden.

Trials in each block belonged to one of four conditions, as seen in Figure 12:

- Six target color matching circles
- Five target color matching circles
- Three target color matching circles
- Two target color matching circles

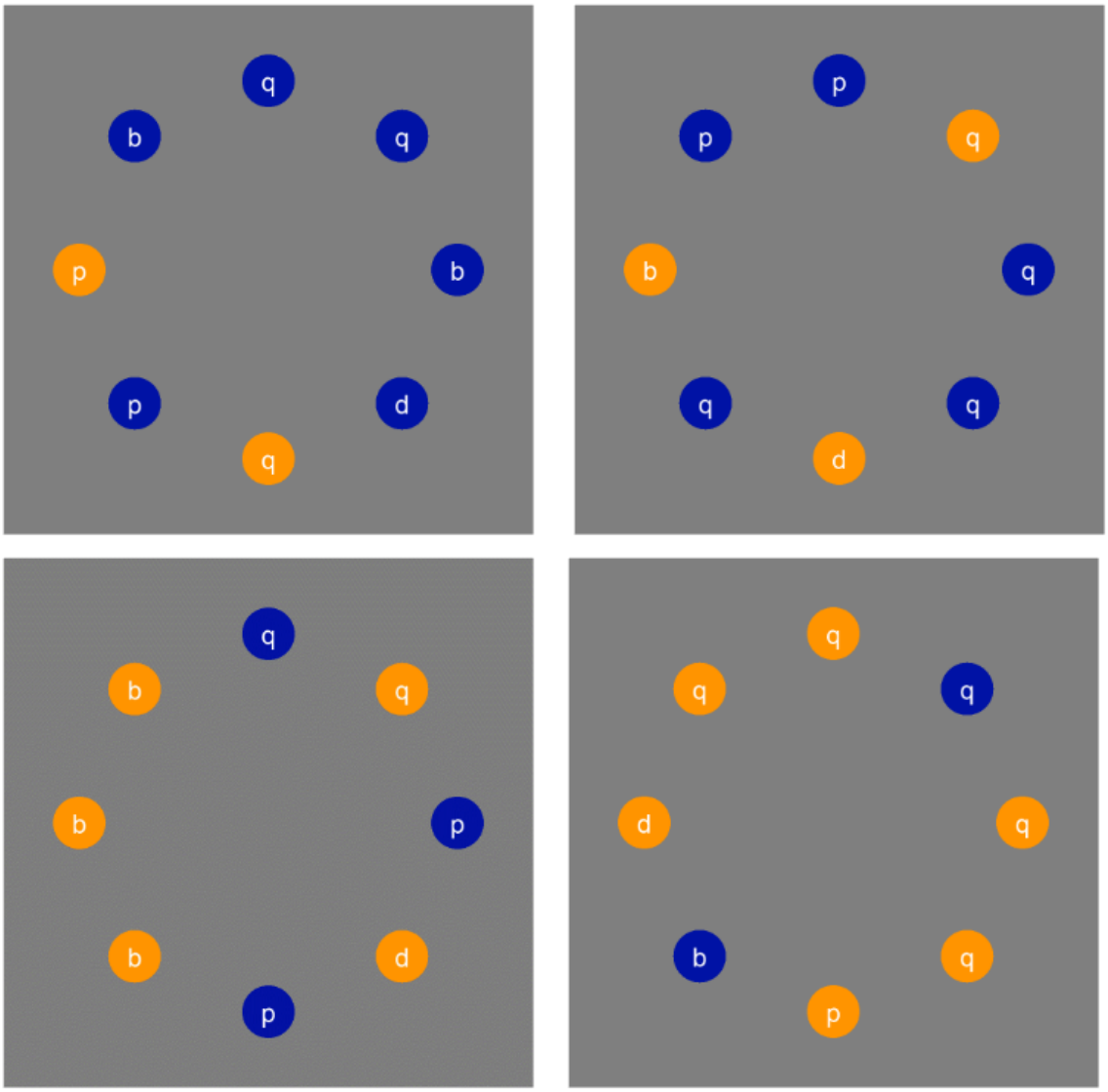

Figure 12. All of the possible trial conditions: six target color matching circles (top-left), five target color matching circles (top-right), three target color matching circles (bottom-left), and two target color matching circles (bottom-right).

Out of the 20 trials in a block, there were seven trials each of the six and five target color matching circles and there were three trials each of the three and two target color matching circles. The number of circles that match the target color is referred to as the Matching Subset Size (MSS). These trials were randomly arranged throughout

a block with the restriction that no more than two of the same MSS appeared in a row. The search stimuli was displayed on screen until a response was given. At this point, the search stimuli was removed from the screen and the response feedback was given as either "Correct" or "Incorrect." The feedback was displayed for two seconds before the next trial instruction screen was displayed. The participant was allowed to take a break after completing blocks 7 and 18 and was presented with the screen saying "You have reached a break point. Please notify the test administrator." At this point, the participant was able to take a break if they chose.

3.4.3 Human Responses

In general, an efficient search is the participant viewing the minimum amount of circles necessary to select the correct answer. Unless the participant finds the target letter before they searched all of the efficient circles, the minimum number of circles necessary to be viewed is two circles for the six and two MSS and is three circles for the five and three MSS.

In addition to determining whether the search was efficient, the participant's search was also analyzed to see if it is a "non-normal" search, if the participant missed the target, if the participant searched in a circular fashion, if the participant performed a multiple-minority-only search, and if the participant performed a majority-then-minority search. Each search type and its description can be seen in Table 2.

Table 2. Search Types

Search Type	Description
Efficient	Participant viewed the minimum amount of circles necessary to select the correct answer
Non-Normal	The participant first searched all of the minority-colored circles and then searched additional majority-colored circles
Miss	The participant gazed at the target letter and then continued searching additional circles
Circular	The participant searches the circles in a circular manner
Multiple-Minority-Only	The participant gazed at only the minority-colored circles but viewed one or more of the circles more than once
Majority-then-Minority	The participant performs a multiple-minority-only search except for the very first circle, which is a majority-colored circle

The fixation area for a letter was defined as triple the area of the circle stimuli, or 6°. For a fixation to occur, the participant's gaze must fall within this fixation area for ten or more frames, or about 160 milliseconds. A inefficient search is a search in which the participant does not search the color for which there are the fewest circles to determine whether the target letter's circle is the target color. For example, in Figure 13, there are six orange circles and only two blue circles. Because there is only ever one instance of the target letter, the participant can search the two blue circles and determine that the target letter "d" is not present amongst them. Therefore, the participant can conclude that the target letter's circle must be orange. In this example, an efficient search would be the participant viewing both of the orange circles and then selecting an answer choice. An example of an efficient search can be seen in Figure 13.

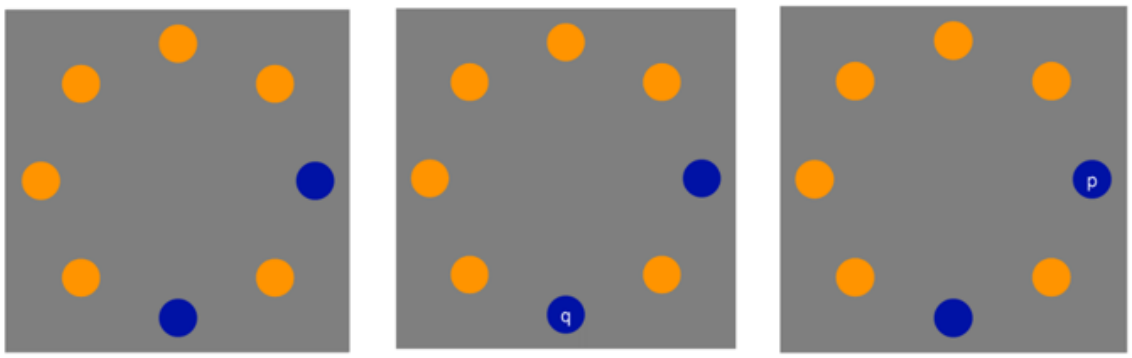

Figure 13. This trial has had a nudge applied. The participant is presented with the top image. The bottom-left image occurs if the participant fixates on the bottom circle, and the bottom-right image appears if the participant fixates on the right circle.

A trial is marked as inefficient if the participant does not only search the circles colored with the color that appears the least. A block is marked as inefficient if ten or more trials of that block are inefficient. If the previous block was a inefficient block, then the nudge was applied to all of the circles of the following block. However, if the previous block was efficient, then no nudge was applied to the following block.

3.4.4 Block Design

An overview of the block design can be seen in Figure 14.

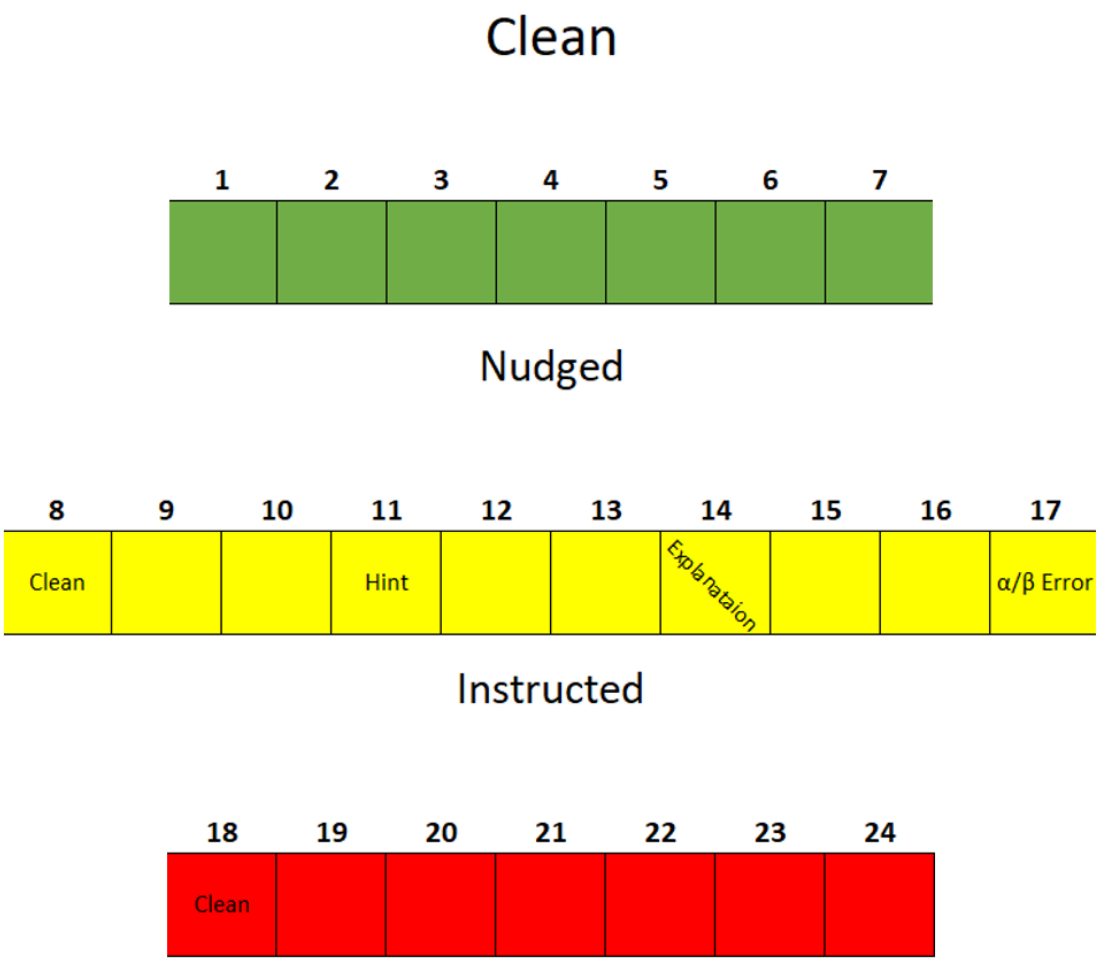

Figure 14. The block design of the experiment

The first seven blocks of the experiments proceeded as described above and are referred to as the "clean" blocks. However, the middle ten and the last seven blocks differed. During the middle ten blocks, a mitigation technique known as a "nudge" was applied to a block if the preceding block had been determined to consist of inefficient searches. Because of this, the middle ten blocks are referred to as the "nudge" blocks. The nudge that was used in this experiment is the same nudge that was used in Rajsic's experiment: the letters were not visible unless a participant visually fixated on an area around the letter. An example of a nudge can be seen in

Figure 13.

The first block of the nudged section never had the nudge applied so that the blocks could be kept separate. Thus the application of a nudge was wholly dependent on the search style of the participant during the trials. If the following block was to have a nudge applied, then the block instruction screen featured the additional notification informing the participant that they must look at a circle to make the letter inside appear. An example of a nudged block's instruction screen can be seen in Figure 15. Regardless of whether the preceding block was nudged, another mitigation technique known as the "hint" occurred before the 4^{th} block in the nudged blocks, or the 11^{th} block overall. This hint informed the participant that there is exactly one instance of the target letter and that there are only ever two colors present - thus an efficient way to determine the color of the target letter's circle is to look at the circles with the color that appears less on the screen. An example is then shown that demonstrates the method. The hint can be seen in Figure 16. Before the 7^{th} block in the nudged blocks, another mitigation technique known as the "explanation" occurred. The explanation informed the participant why the nudge was occurring and informed them that the computer would continue to apply a nudge if the computer detected a inefficient search. The explanation can be seen in Figure 17. Lastly, the 10^{th} and final block in the nudged block was inversely nudged - if the participant had performed a inefficient search on the 9^{th} block then no nudge would be applied, and conversely if the participant had performed an efficient search previously then the block would be nudged. This block was intended to capture the participant's reaction to receiving the nudge when it was not needed or to not receiving the nudge when it was actually needed.

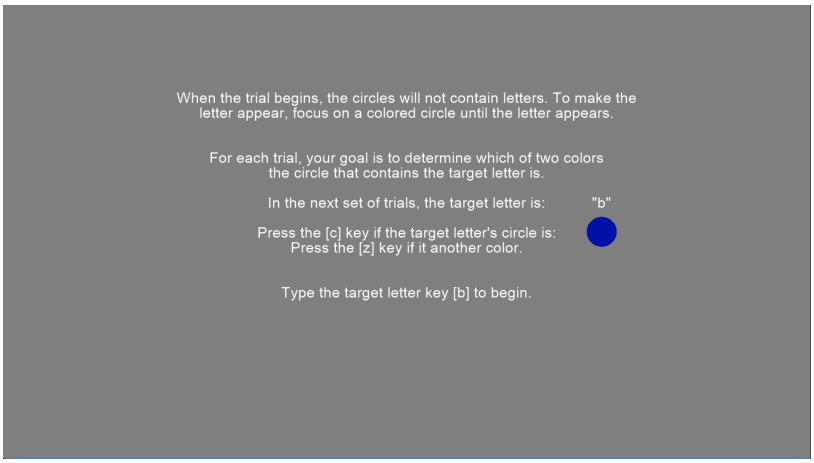

Figure 15. This screen appears before a nudged block only. It differs from the normal block instruction screen by the addition of the top line that instructs the subject to fixate upon a circle to make the letter inside appear.

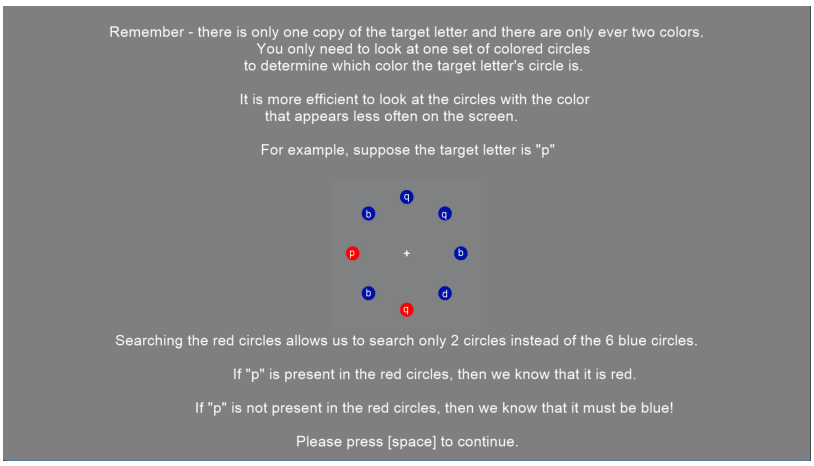

Figure 16. The hint screen is shown before the 5^{th} block in the nudged blocks. It instructs the subject on how to perform an efficent search for the target letter's color.

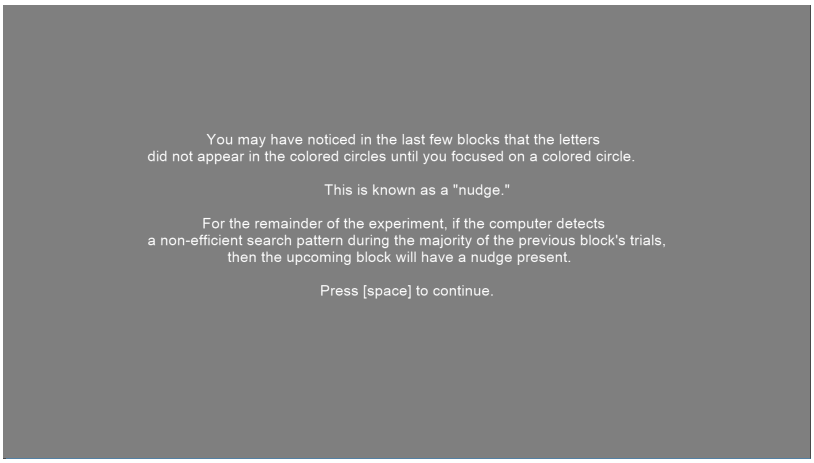

Figure 17. The explanation screen is shown before the 8^{th} block in the nudged blocks. It instructs the subject on how to perform an efficient search for the target letter's color.

The last seven blocks are referred to as the "instructed" blocks because the participant is instructed to perform an efficient search throughout the blocks. This instruction to perform an efficient search occurs at the beginning of the instructed blocks in block 18 with the screen seen in Figure 18. Immediately after this screen is the hint seen in Figure 16 which teaches the participant how to perform an efficient search. Following these two screens, the instructed blocks behave in the same manner as the nudged blocks in that the application of a nudge depends wholly on the preceding block and whether it is an efficient or non-confirmatory block. Similar to the nudged blocks, the first block of the instructed blocks did not have the nudge applied. However, the instructed blocks differed from the nudged blocks in that the instruction screen shown before the trial now featured the instruction to "perform an efficient search" as seen in Figure 19.

The purpose of including various mitigation techniques is to increase the number of efficient searches. The first seven blocks did not receive any mitigations so that the participant's default VSP could be captured. The nudge was introduced by itself in

the "nudged" blocks to see how the participant would react with just an additional cost to the search. However, the participant might not have realized that there existed a method to perform an efficient search, and so the hint was introduced at the 11^{th} block to teach them how to search efficiently. At this point, the participant now knew that there existed an efficient search method but still might not have known why the nudge was occurring. Thus, the explanation was used at block 14 to explain why the nudge occurred. Now the participant knows that there exists an efficient search method and that they will be penalized if they don't utilize this method. Yet, knowing all of the above, the participant could still choose to use an inefficient search method. Therefore, the instruction served to ensure that the participants used an efficient search method for the final seven blocks.

During the execution of the experiment, a change was made such that the last seven blocks would not have the nudge applied even if the previous blocks were inefficient. This was done in an effort to ensure that the EEG data was balanced so that it contained a similar number of efficient and inefficient searches from which to learn. Additionally, by removing one mitigation technique, the effect of the instructions on the participant's number of efficient searches can be determined.

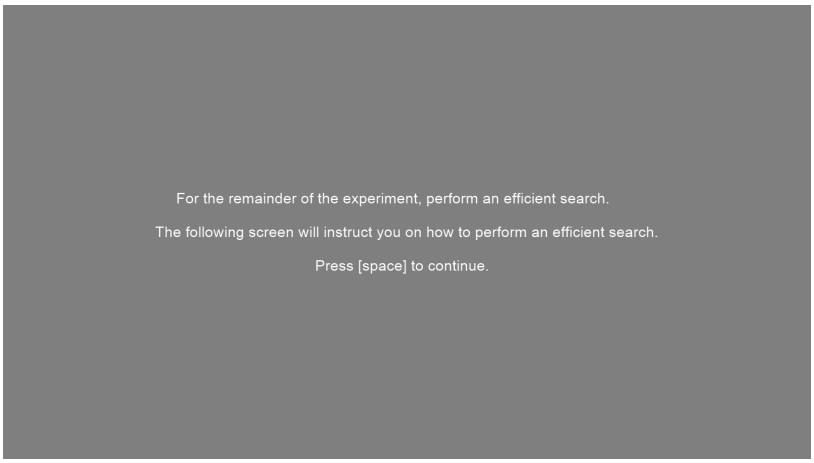

Figure 18. The instruction screen is shown at the beginning of the instructed blocks before block 18.

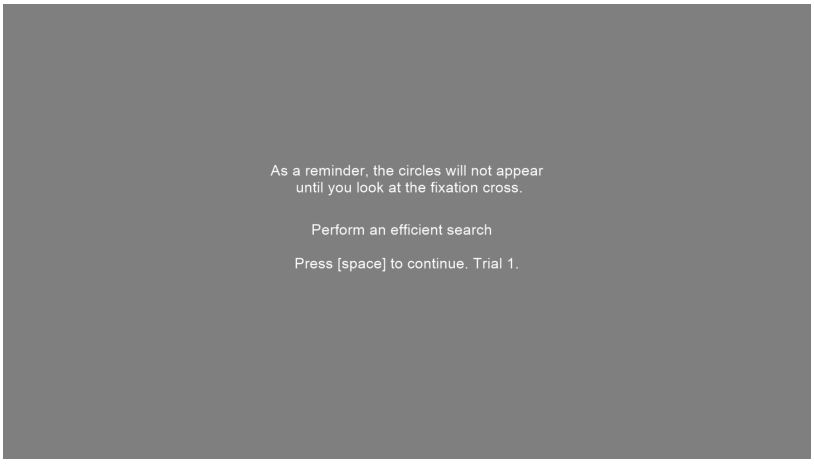

Figure 19. The instruction screen is shown during the instructed blocks. It differs from the standard instruction screen shown before trials in that it instructs the participant to perform an efficient search.

After the last seven blocks of the instructed blocks, the participant was informed that they had finished the experiment and was instructed to notify the test administrator. The participant underwent another two-minute baseline and completed a post-experiment questionnaire.

3.4.5 Variables

3.4.5.1 Independent Variables

The independent variables in this experiment are:

- The amount of target color matching circles present and thus the amount of target color mismatching circles
- The presence of a nudge
- Whether the participant has received the hint
- Whether the participant has received the explanation
- Whether the participant has been instructed to perform an efficient search

The independent variables with varying level in the experiment are listed in Table 3.

Table 3. Independent Variables

Independent Variable	Type	Measurement	Predicted Effects
Number of Target Color Matching Circles	Numerical	[2, 3, 5, 6]	The more target color circles are present, the longer it will take for the subject to search
Subject has Received the Hint	Categorical	[Yes, No]	The subject will perform fewer inefficient searches
Subject has Received the Explanation	Categorical	[Yes, No]	The subject will perform fewer inefficient searches
Subject has Been Instructed to Perform an Efficient Search	Numerical	[Yes, No]	The subject will perform fewer inefficient searches

3.4.5.2 Response Variables

The participant's search pattern, and subsequently the classification of the search pattern as efficient or not, is a response variable for this experiment. Additionally, the association of physiological signals such as EEG, ECG, GSR, and EOG with an efficient visual search is an objective of this experiment. Correlations of an efficient

search with patterns in the physiological signals will be used to determine whether an efficient search can be detected via physiological signals. EEG consists of electrical signals recorded over the entirety of the head. ECG captures movements over the chest as well as heart rate. EOG accounts for eye blinks in the EEG data and can be used to measure the attention of the participant. All response variables are shown in Table 4.

Table 4. Response Variables

Response Variable	Type	Measurement
Participant's Search Pattern	Categorical	[Efficient, Inefficient, Non-Normal, Miss, Circular, Multiple-Minority-Only, Majority-then-Minority]
EEG	Numerical	64 Voltage Signals Over Time
ECG	Numerical	Two Voltage Signals Over Time
EOG	Numerical	Four Voltage Signals Over Time

3.4.5.3 Control Factors

Within the experiment, control factors were:

- The colors of the circle stimuli remained the same, with the same colors appearing in the same pairs
- The first color in the pair was always the target color
- The set of available letters were always the same letters and each letter in the set was used at least once
- Each participant completed the experiment in the same experimental station in the same laboratory
- Except for the nudge, each participant received the same mitigation factors at the same times throughout the experiment

Each of the 24 blocks in the experiment contained the same ratio of trial conditions, as seen in Figure 12:

- Six target color matching circles
- Five target color matching circles
- Three target color matching circles
- Two target color matching circles

However, both the blocks and the trials within the blocks were shuffled to create four separate sequences of blocks and trials. These sequences were created using the following conditions:

- 35% of the trials had six target color matching circles
- 35% of the trials had five target color matching circles
- 15% of the trials had three target color matching circles
- 15% of the trials had two target color matching circles
- No two of the same trials appears in a block
- The minority-colored circles are separated by at least one of the majority-colored circle

Trials with a higher color matching circle count were chosen to appear more because these trials provide more information than the other trials on whether the participant is searching efficiently. These sequences were used in a rotating order with the participants: participant 1 received sequence A, participant 2 received sequence B, and so on, with the 5th participant receiving sequence A again. All of the control variables can be seen in Table 5.

Table 5. Controlled Factors

Controlled Factor	Levels
Color Pairs	[Blue, Orange] [Green, Red] [Purple, Yellow]
Target Color	[Blue] [Green] [Purple]
Letters Present	[b, d, p, q]
Timing of Mitigation Attempts	[Nudges begins at block 8] [Hint occurs before block 11] [Explanation occurs before block 14] [Instructions to perform an efficient search begins at block 18]

3.4.5.4 Nuisance (Confounding Factors)

With an experiment consisting of human participants, there are many potential uncontrolled factors. The expected nuisance factors and the mitigation strategies are contained in Table 6.

Table 6. Nuisance Factors

Nuisance Factor	Strategy	Anticipated Effects
Learning Effect: Test progression may result in decreasing information search time, which would make information search time a poor behavioral measure.	Subjects attend a training session previous to the experiment day in which they perform a single block of trials to familiarize themselves with the search task.	A decrease in search time due to task familiarization, but due to the simple nature of the search tasks the time will be negligible.
Misinterpretation of instructions: Confusion on tasks could lead to undesired brain activity and false measure of efficient searches.	Subjects attend a training session previous to the experiment day in which they perform a single block of trials to familiarize themselves with the search task.	Subjects may not understand the instructions, but a majority will understand the instructions presented to them.
Unbalanced number of efficient and inefficient searches	The ESE has blocks dedicated to obtaining both efficient and inefficient searches through the use of mitigation techniques.	The subject will search inefficiently during the blocks with no mitigation attempts and will search efficiently when instructed to do so.

3.4.6 Participants

Sixteen United States Air Force personnel participated in this experiment. Participant age ranged from 22 to 37 with a mean age of 28.9, a standard deviation of 5.2, and a median of 27.5. All participants had, at minimum, a Bachelor's degree and all used computers daily in their job and personal lives. All participants had a sleep quality of "fair" or better, and had an average of 6.7 hours of sleep with a 0.9 hour standard deviation. Inclusion criteria included the ability to operate a computer, be at least 18 years of age, and be a U.S. citizen. Exclusion criteria included:

- Inability to use a keyboard
- Visual impairment causing an inability to view a computer screen

- Physical impairments causing an inability to use a computer
- Use of hair products which interfere with the EEG electrodes
- Thick hair which prevents a proper fitting of the EEG cap
- A head size that is unable to fit into an EEG cap

No potential participants met the exclusion criteria.

Before starting the training for the experiment, all participants read the informed consent document (ICD). Because of the placement of the electrodes, specifically the ECG electrodes, additional participant consent was obtained and the participants were able to apply the electrodes themselves if they so chose. Participants did not receive any form of compensation for this experiment.

3.4.7 Materials

Four computers were used in this experiment:

- A computer running the experiment via Pyshcopy v3.2, called the Control Station PC (CSPC) [92]
- A computer running the Smart Eye Pro gaze tracking software [93]
- A computer running the Multi-modal Analysis of Psychophysiological and Performance Signals (MAPPS) software for experiment screen and web camera recording [94]
- A computer running the Cognionics recording software to collect physiological signals [95]

The setup of these computers can be seen in Figure 20.

In addition to the four computers listed above, other necessary equipment includes the Smart Eye Pro gaze tracker and the three physiological sensors which are covered in further detail in Section 3.4.7.1. The experimental setup is shown in Figure 21.

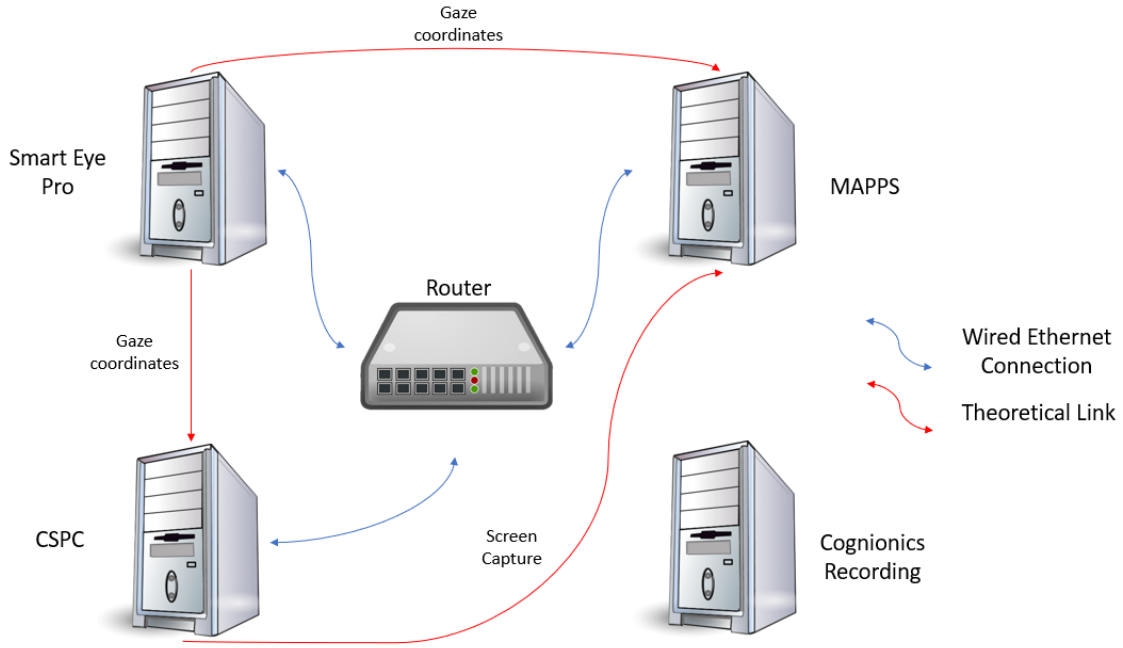

Figure 20. The layout of the computers used in the ESE. More information about the cognionics recording computer and its use can be seen in Section 3.4.7.1.

The six Smart Eye Pro cameras are visible along with the four infrared (IR) flashers. During the experiment, only the top monitor was on. The participant was seated at a chair centered in front of the monitor so that all six cameras obtained a clear image of the the participant's eyes. Once the participant was seated and adjusted properly, the participant was asked not to make large movements as it could impact the accuracy of the gaze tracking. Once the experiment began, the overhead lights were turned off and a dark blue backlight behind the top monitor was turned on to increase the light level of the screen.

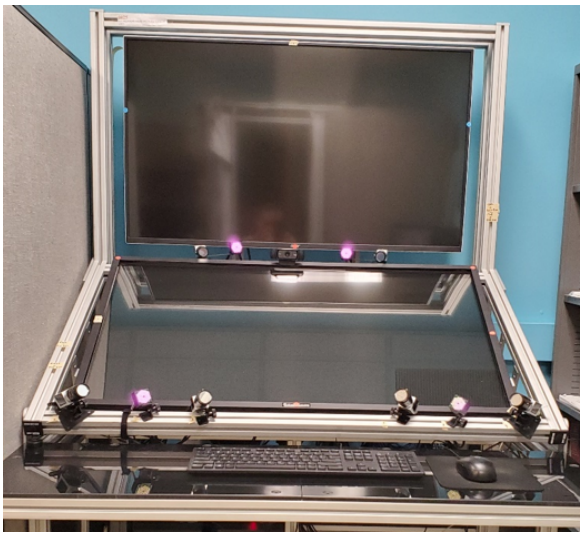

Figure 21. The subject was seated in front of the experimental station. Only the top monitor was used during the experiment.

3.4.7.1 Physiological Recording Devices

The collection of physiological signals required two computers: the computer upon which the experiment was running and a separate computer to record the physiological signals as seen in Figure 22. The CSPC was connected via USB to a piece of hardware known as the "trigger box." The trigger box communicated wirelessly with the data acquisition unit (DAQ) that was physically connected to the EEG cap. Triggers in byte format were transmitted first from the CSPC to the trigger box, then from the trigger box to the DAQ, and finally from the DAQ to the computer which was recording all of the physiological signal data. The triggers sent from the CSPC corresponded with events and data from the experiment itself, such as the block number that the participant is in or whether the participant just performed an inefficient search. These triggers create a marker in the Cognionics recording software so that certain events are able to be easily associated with physiological signals for post-processing. The physiological collection computer also received other signals

such as ECG and EOG.

Figure 22. The architecture of collecting and recording physiological signals.

This experiment utilized the Smart Eye Pro gaze tracking system. This gaze tracking system features a 60 Hz sampling rate and is capable of a gaze accuracy of 0.5° in ideal conditions. The Smart Eye Pro was capable of outputting data streams via User Datagram Protocol (UDP) to the PyschoPy program which was used to trigger stimuli.

For this experiment, the Cognionics Mobile-72 system was used to collect all physiological signals. In addition to capturing 64 EEG voltage channels, the Mobile-72 system also collects 8 additional channels such as ECG and EOG. To collect EEG signals, the participant wore the Cognionics EEG cap shown in Figure 23.

Figure 23. The Cognionics EEG Cap.

The 64 EEG electrodes present on the cap are located in positions based on the International EEG 10-20 Standard electrode placement, seen in Figure 24. To ensure proper collection and to minimize interference, a conductive gel was applied to each electrode until the electrode's impedance was below 100k-ohms.

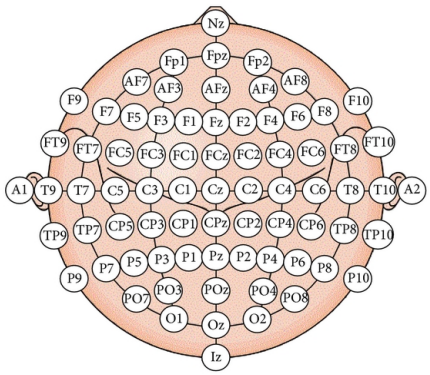

Figure 24. The International 10-20 Electrode Placement

The ECG electrode positions are shown in Figure 25. The locations for the horizontal and vertical EOG electrodes are shown in Figure 26. To ensure proper placement, the test administrators applied the EOG electrodes. For privacy reasons, the participant was instructed on where and how to place the ECG electrodes and was then shown to a private room to allow the participant to attach the electrodes onto themselves.

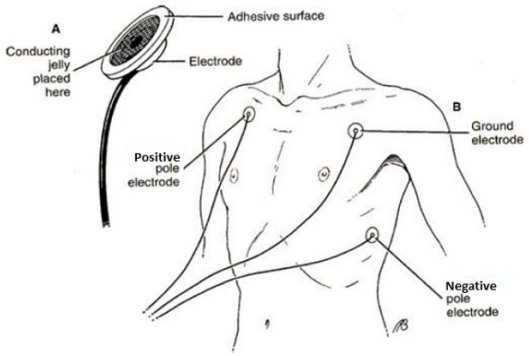

Figure 25. ECG Electrodes Placement

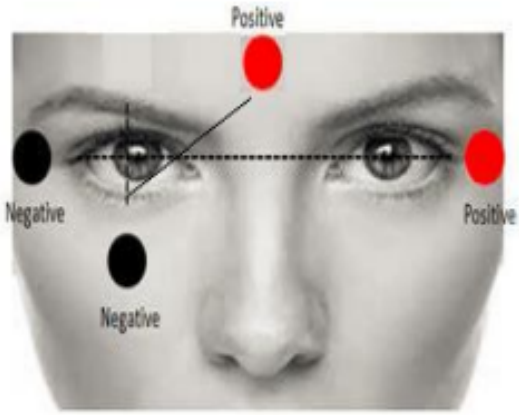

Figure 26. Vertical and Horizontal EOG Electrodes Placement

3.4.8 Procedures

3.4.8.1 Training

Prior to the day of the experiment, participants underwent a training session. During this training session, participants read the ICD and indicated their verbal consent to continue the experiment. To ensure that participants could correctly distinguish between the colors of the stimuli circles in the experiment, a color-blindness test was administered. The color-blindness test consisted of participants counting the number of circles of each color of an example trial displayed on the training screen. Participants also completed a single block of trials to familiarize themselves with the search task. Participants' heads were measured so that the correct EEG cap could be made ready for the experiment day, and a random participant number was assigned at this time.

3.4.8.2 Experiment Day

Upon arrival to the experiment, participants completed a pre-experiment questionnaire. This questionnaire can be seen in Appendix Chapter A.

Once the pre-experiment questionnaire had been completed, the electrodes for ECG, EOG, and GSR were attached. The EEG cap was fitted and a conductive gel was inserted into each electrode to ensure an impedance below 100k ohms.

The participant was seated in front of the experimental station and a gaze calibration with the Smart Eye Pro gaze tracker was performed until the participant had a calibrated accuracy below 3° in each eye. The participant then completed the experiment.

After completing the entire experiment, participants completed a post-experiment questionnaire. This questionnaire can be seen in Appendix Chapter A.

Once the post-experiment questionnaire had been completed, participants were instructed to not discuss the nature of the experiment with anyone who had not already completed the experiment.

The procedures completed before the experiment itself took, on average, around 45 minutes. The experiment itself lasted, on average, 1 hour. Participants were scheduled individually at different times for 2.5-hour blocks to ensure adequate time for preparation, the search trials, and cleanup.

3.4.9 Data Collection

Data from the ESE was collected in a Comma Separated Value (CSV) file generated trial-by-trial. This data includes information about each trial, including:

- The color and location of each circle
- The order in which the participant viewed each circle
- The response, response time, and whether the response was correct

- The target letter's location
- Whether the participant performed a confirmatory search, an efficient search, missed the target letter and kept searching, or searched in a circular manner

All physiological signals are collected by the Cognionics Data Acqusition software and are saved in the BrainVision .eeg file format. The triggers sent by the CSPC are captured and inserted into the EEG data. A list of all trigger values can be seen in Chapter B. An example of this can be seen in Figure 27.

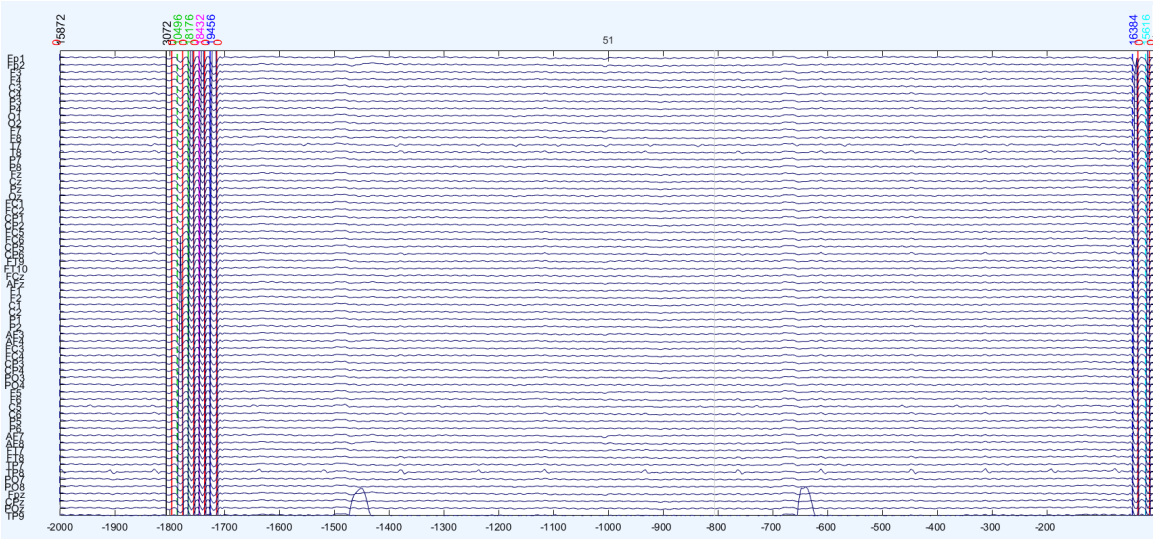

Figure 27. An example of a single efficient search epoch with triggers. The channel names are on the y-axis while the time is on the x-axis. This EEG sequence represents a single search from two seconds prior to the pressing of the answer key. In the trigger values at the -1800 second mark, the 3072 trigger value indicates this is a search in block 12, 10496 indicates it is trial 12, 18176 indicates there is no nudge present, 18432 indicates the participant has seen the hint, and 19456 indicates that the explanation has not been seen. In the trigger values at the end of the trial, the 16384 indicates the participant made a correct response and the 15616 indicates the trial was not confirmatory.

3.4.10 Analysis Strategy

The data collected during the ESE was analyzed using various statistical and machine learning packages in Python and JMP. The machine learning process is

covered in the machine learning pipeline section of this chapter in Section 3.5. The main objective for the analysis of the ESE data was to:

- Determine the initial VSP of the participant
- Determine whether the mitigation techniques used increased the number of efficient searches
- Asses whether physiological data can be linked with an efficient or inefficient search

To first determine the initial VSP used, the VSP that the participant used in the blocks which had not received any mitigation technique, or the first eight blocks, will be determined. This will be determined by classifying the VSP of each trial as either confirmatory, efficient, or circular. A confirmatory search is a search where the participant first searched a colored circle that matched the target color. An efficient search is where the participant searched the colored circles that appeared least and only the minimum required number of circles needed to determine what color the target letter's circle was. Given these conditions for an efficient search, it should be noted that a search could be both confirmatory and efficient. Lastly, a circular search is a search where the participant searched the circles in a consecutive circular fashion. Whichever VSP that the participant used during the majority of trials in the first eight blocks will be deemed that participant's VSP. This will serve to answer research question 1 (Section 3.3.1).

During the data exploration phase, the number of efficient searches per block will be plotted against the block numbers to determine if there is a trend. As the participant completes more blocks and receives more mitigations, the efficient searches per block should increase.

To first test whether the mitigation attempts had a significant impact on the number of efficient and inefficient searches, a linear regression model will be built

which will determine each mitigation technique's effect on the number of efficient searches. The levels and factors can be seen in Table 7.

Table 7. The linear regression's levels and factors

Factors	Levels
Participant has encountered the nudge	[No, Yes]
Participant has received the hint	[No, Yes]
Participant has received the explanation	[No, Yes]
Participant has been instructed to complete an efficient search	[No, Yes]

The statistical significance of the above test will determine the answer of research question 3 (Section 3.3.3).

Lastly, physiological signals both across and within participants will be compared to determine if a difference occurs in the participants while performing an efficient vs an inefficient search.

3.5 Machine Learning Pipeline

3.5.1 Data Pre-processing

Once the ESE had finished, the collected raw physiological signals were saved in a BrainVision .eeg file format. The EEG data was then processed following the PREP pipeline using the 2019 version of EEGLAB, an interactive Matlab toolbox for processing continuous and event-related EEG data [96]. A summary of the PREP pre-processing pipeline is given below:

1. The data was down-sampled from 512 Hz to 250 Hz to speed up computation and to cut off unnecessary high-frequency information.

2. A high-pass filter at 1 Hz was applied using a basic finite impulse response (FIR) filter. The purpose of this filter was to remove low frequency drift.
3. The International 10-20 channel system information was imported to allow for channel re-referencing.
4. A notch filter was applied at 60 Hz to remove electric line noise.
5. Bad channels were rejected using Automatic channel rejection using kurtosis with a Z-score threshold max of 5.
6. The information that was lost due to the removal of the bad channels was interpolated using spherical interpolation to prevent bias when re-referencing.
7. The data's reference was changed from the mastoid to the average of the channels.
8. Independent Component Analysis (ICA) was performed to identify the components that were associated with eye-blinks.
9. The components that were associated with eye-blinks were removed using ICA Blink Metrics with the vertical EOG channel used as an eye-blink reference [97].

After processing, the data was segmented in EEGLAB. The data was segmented by a window two-seconds in length, with the end of the window occurring when the participant pressed the key indicating which color they believed the target letter's circle was. Any epochs that overlapped, i.e. they contained data from another epoch, were discarded. This segmentation produced one file per participant. Due to an unknown error, the hardware used to transmit and receive the signals dropped certain epochs. Because of this, some participants did not have the full 480 epochs. Details about the epochs included for each participant can be seen in Table 25.

Table 8. Number of trials recorded per participant

Participant	Total Number of Trials Recorded
0271	475
1437	480
2070	478
2765	458
4030	479
4431	477
4613	480
5617	473
5669	480
5791	474
5952	337
6973	478
6969	480
7669	478
9138	480
9150	480

Following segmentation, any noisy data that was present was rejected by visual inspection in EEGLAB. Any epoch that had a large amount of noise when compared to the rest of the data was rejected.

Once the data has been segmented, for a single participant, a single file contains all of the EEG data of the searches that occurred during the ESE. The last two seconds of each search were captured, which means that there are at most 960 seconds worth of EEG data per file.

3.5.1.1 Time Series Feature Extraction

To prepare each segment for machine learning (ML), a sliding window generated smaller segments from each larger segment. The sliding window that was used was composed of two adjustable parameters: the window's size and step. The window's

size is the length of the sequence that is generated from the larger segment, while the step is the time points that the window moves to generate the next sequence. In this analysis, a sequence lasting two seconds at a sample rate of 256 Hz contains 512 frames. A sliding window with a window size of 0.5 seconds, or 128 frames, and a step size of 0.5, creates four 0.5-second non-overlapping epochs from the original two-second sequence. Using a sliding window creates multiple epochs from a single search. For a time series classification, each epoch is viewed as an observation and each observation has 64 features where each feature corresponds to the EEG electrodes.

3.5.1.2 Frequency Feature Extraction

All of the epochs present in the data were transformed to the time-frequency domain using MATLAB. A family of complex Morlet wavelets which spanned 30 frequencies over the logspace from 1 to 80 Hz was used to transform the data into the five traditional EEG bands: delta (1-6 Hz), theta (7-11 Hz), alpha (12-15 Hz), beta (16-22 Hz), and gamma (22-30 Hz) [77]. The mean of the power spectral density was then obtained. Using the mean power for each of the five frequency bands for a 64 electrode EEG cap produced 320 features for each participant. These features composed the inputs for a single observation for a ML model.

3.5.2 Datasets

One EEG recording per participant entered the data pre-processing step and two datasets emerged:

1. The Time Series Signal per Search dataset
2. The Frequency Features per Search dataset

The Time Series Signal per Search dataset contains a single participant's entire search, where individual epochs contain data from a single search. The label for each

epoch is either "efficient" (represented as a 1) or "inefficient" (represented as a 0) and is described in detail in Section 3.4.4. Because the entire search was labeled as either efficient or inefficient, all epochs from the same search will possess the same label. The ML problem for this dataset is a many-to-one binary classification of which the goal is to classify a time-series sequence of EEG data as originating from an efficient or an inefficient search. The ML input shape is (batch size, time steps, features), where batch size is variable and is the number of observations, time steps is the number of frames in the epoch and is the same as the size of the sliding window, and the features is 64 and is the same as the number of EEG channels.

The Frequency Features per Search dataset is labeled in an identical manner to the method above in the Time Series Signal per Search dataset. The major difference between the two datasets is that in the Time Series Signal per Search dataset, the data is organized in time series sequences in each epoch, whereas in the Frequency Features per Search dataset the data is the mean power in each of the five traditional frequency bands of EEG. The time series sequence for each trial is converted into a single mean value by the time-frequency transformation. The ML problem for this dataset is a one-to-one binary classification to classify a search as efficient or inefficient. The ML input shape is (batch size, features) where batch size is variable and is the number of observations in the input, and the features consist of the mean power of each of the five frequency bands of the EEG channels. In this dataset, a single search consists of 320 features because there are 64 channels of EEG data and each channel has the mean power of each of the EEG bands.

3.5.2.1 Challenges of EEG

There are many important factors to consider when applying ML to EEG. EEG signals often have a very low signal-to-noise ratio, meaning that EEG signals are very

likely to have a large amount of both noise and outliers [98]. One of the purposes of pre-processing the EEG data was to reduce this noise, however, there are still significant amounts of noise present in the data.

Additionally, because features can be taken from many channels over many different time steps, EEG signals are of high dimensionality [98]. In the Frequency Features per Search dataset there are 320 features, while there are only 480 observations per participant. The curse of high dimensionality occurs in ML when there are a high number of features when compared to a low number of observations. Because of the high number of features, extra steps must be taken to prevent over-training when applying ML on the EEG classification problems.

3.5.3 Classification Models

Because of the two datasets that were generated by the pre-processing, there are two ML problems present: a one-to-one classification problem and a many-to-one classification problem. For the one-to-one classification problem Linear Discriminant Analysis (LDA), random forest classifier (RFC), and a fully-connected Artificial Neural Network (ANN) were investigated. The many-to-one problem deals with classifying a time series of the EEG data. Deep learning models are able to classify sequenced data well, thus deep learning models for solving this classification problem were investigated. The deep learning models that were evaluated were an Long Short-Term Memory (LSTM) model and a Temporal Convolutional Network (TCN) model. For each model, the data was split into training, validation, and test sets. The training set size was 60%, the validation set size was 20%, and the test set size was also 20%. The test set was sequestered from the model until the final testing occurred.

3.5.3.1 Linear Discriminant Analysis

An LDA was used as a baseline score because LDAs are stable even with a small number of observations [79]. LDAs have no hyperparameters, thus making tuning unnecessary. Despite not using feature selection, the high dimensionality of the dataset was accounted for by choosing the LDA's shrinkage parameter to be 'auto' and its solver to 'lsqr.' Using these settings helps improve the LDA's estimation of covariance matrices for datasets that have high dimensionality [99].

3.5.3.2 Random Forest Classification

Random forests naturally have the ability to select the best features so all features were used. The best parameters per participant's model were chosen based on the Mathew's Correlation Coefficient (MCC). MCC can be calculated directly from the confusion matrix using the formula:

$$\frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}}$$

The hyper-parameters tested were:

- The number of trees - Increments of 10 from 10 to 500
- The number of features considered - integers from 1 to 25
- Maximum depth of a tree - integers from 1 to 25

The recommended number of features for random forest classifier is the square root of the total number of features. $\sqrt{320} = 17.8$, so the recommended number of features is 18. A sweep to 25 features was performed to ensure that the optimal number of features was obtained.

For each participant, an RFC model was trained using the default parameters of SKLearn's Random Forest Classifier. This model's MCC was used to measure the improvement in performance of the tuned RFC. Next, the optimal number of trees was found by training an RFC using the square root of the max number of features. Using the optimal number of trees, a hyperparameter sweep was performed to find the optimal number of maximum features and the maximum depth. Finally, each model was tested and the MCC of the tuned model was determined. MCC was only used during the tuning phase for the RFC and was used to find the best combination of hyperparameters. The results of the tuning can be seen in Table 9.

Table 9. The RFC optimal hyperparameters for each participant.

Participant	Initial MCC	Tuned MCC	Number of Best Trees	Number of Best Features	Best Depth
0271	-0.021259	0.367983	10	16	10
1437	0.227690	0.376607	210	22	6
2070	-0.008973	0.314031	30	12	2
2765	0.038338	0.380592	60	21	7
4030	-0.050538	0.241594	100	5	3
4431	0.205798	0.420841	130	24	16
4613	-0.046089	0.423797	30	10	6
5617	0.019038	0.386374	40	8	3
5669	-0.036232	0.345439	300	9	6
5791	0.087689	0.361359	40	9	7
5952	0.157911	0.575537	50	12	6
6973	-0.208656	0.205556	180	4	13
6969	0.259356	0.378002	90	20	8
7669	0.085624	0.273787	10	1	10
9138	0.095168	0.317808	80	12	3
9150	-0.103280	0.460261	20	20	1

3.5.3.3 Fully-Connected Neural Network

ANNs were chosen because they are relatively simple when compared to other deep learning models. The ANN model parameters was chosen through hyper-parameter

sweeps. The model that was chosen as the final model was a model that could achieve a high validation accuracy with a low training time. The hyper-parameters tested were:

- Number of hidden layers - Integers from 2 to 8
- Number of hidden nodes per layer - 2, 4, 8, 16, 32, 64, 128, 256
- Learning rate - 0.01, 0.001

Model selection was completed using a validation-based early stopping method with a patience of 10 epochs and a delta of 0.001. Each network was trained using a batch size of 32.

The first layer in the ANN model consisted of a fully connected Dense layer of the optimized number of nodes with a rectified linear unit (ReLU) activation function. Each subsequent layer consisted of a fully connected Dense layer of the optimized number of nodes and a ReLU activation function followed by a dropout layer of 0.2 and ending in a Batch Normalization layer. This sequence repeats for the optimal number of hidden layers per participant. The final layer is a fully connected Dense layer with one unit and a sigmoid activation function which serves as the output of the model. The sigmoid output of the final layer represents the probability of the input being one, or efficient. The model uses an Adam optimizer with the optimal learning rate and a binary cross-entropy loss function. The architecture for this model can be seen in Figure 28.

The sigmoid function of the final Dense layer returns the probability that the observation belongs to the efficient class. The classification thresholds determine which predictions will be classified into each category, i.e., an classification threshold of 0.5 indicates that an observation will be classified as efficient if the model is only 50% certain that it belongs to that class. While not exactly a tunable parameter, it is still important to determine a classification threshold that is appropriate for the

specific problem. Thus, it is necessary to determine an appropriate threshold for this problem. For each combination of hyperparameters during the sweep, the following thresholds were considered:

- Threshold values from 0.0 to 1.0 in 0.1 increments

Analysis of the participant questionnaires determined that only $\frac{1}{16}$, or 6.25%, of participants found the nudge "annoying." Thus, it was determined that balanced accuracy should be the metric that the models should be judged. The threshold value with the highest balanced accuracy was determined to be the best threshold for that combination of hyperparameters.

A model for each participant was tuned using the parameters above. The results of the tuning can be seen in Table 10.

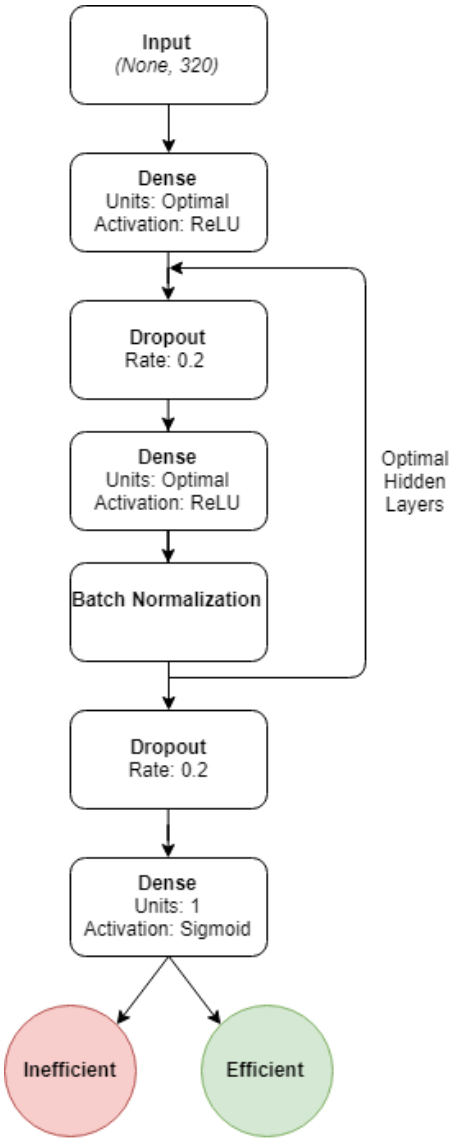

Figure 28. The ANN architecture.

Table 10. The ANN optimal hyperparameters for each participant

Participant	Hidden Layers	Width	Learning Rate	Threshold	Val Accuracy	Val Balanced Accuracy
0271	2.0	16.0	0.001	0.6	0.657895	0.660430
1437	2.0	16.0	0.010	0.5	0.558442	0.562078
2070	4.0	64.0	0.001	0.5	0.337662	0.527778
2765	4.0	8.0	0.010	0.5	0.608108	0.568267
4030	2.0	256.0	0.001	0.5	0.610390	0.632695
4431	3.0	128.0	0.010	0.5	0.662338	0.669048
4613	5.0	64.0	0.001	0.5	0.467532	0.575330
5617	5.0	64.0	0.010	0.5	0.631579	0.610994
5669	6.0	128.0	0.001	0.6	0.610390	0.634863
5791	6.0	8.0	0.001	0.5	0.618421	0.623512
5952	4.0	128.0	0.001	0.5	0.814815	0.796537
6973	3.0	2.0	0.010	0.5	0.532468	0.572917
6969	5.0	128.0	0.001	0.4	0.714286	0.695238
7669	3.0	16.0	0.010	0.5	0.584416	0.578716
9138	4.0	256.0	0.001	0.4	0.688312	0.670991
9150	2.0	128.0	0.010	0.5	0.714286	0.718182

3.5.3.4 Long Short-Term Memory

Unlike the other models which were fit using the frequency features of the EEG data, the time-series EEG data was used to fit an LSTM model. The LSTM model used was inspired by Kumar et al.'s model used in the Optimized CSP and LSTM (OPTICAL) predictor [100]. Kumar et al. demonstrated that a two-layer LSTM model, each with varying numbers of hidden units, performed better than other models when it came to classifying EEG data.

The hyperparameters tested were the number of hidden layers in each of the two LSTM layers. These hyperparameters were:

- Number of hidden units in the first LSTM layer: 10, 50, 100, 200
- Number of hidden units in the second LSTM layer: 10, 50, 100, 200

Similar to the ANN hyperparameter sweep, a threshold sweep was also performed.

The LSTM model consisted of an initial CuDNNLSTM layer containing the optimal number of hidden units, an input shape of (500, 64), and return_sequences was set to True. This layer was followed by a Batch Normalization layer. Next was the second LSTM layer which contained the optimal number of hidden units and an input shape of (500, 64). The final layer was a Dense layer with one unit and a sigmoid activation function. The architecture for the LSTM model can be seen in Figure 29.

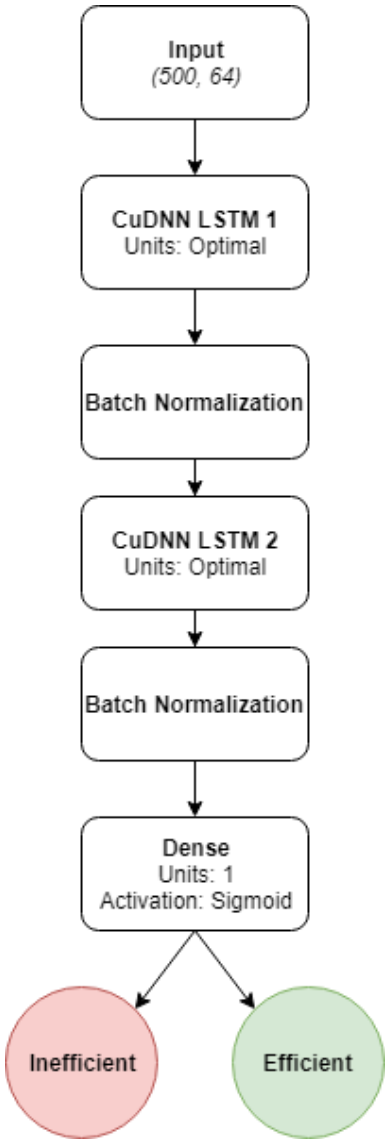

Figure 29. The LSTM model architecture.

A model for each participant was tuned using the parameters above. The results of the tuning can be seen in Table 11.

Table 11. The LSTM optimal hyperparameters for each participant.

Participant	LSTM 1 Hidden Units	LSTM 2 Hidden Units	Threshold	Val Accuracy	Val Balanced Accuracy
0271	100.0	50.0	0.5	0.618421	0.627555
1437	200.0	10.0	0.5	0.671053	0.659972
2070	100.0	200.0	0.4	0.565789	0.630008
2765	10.0	10.0	0.4	0.397260	0.568627
4030	200.0	200.0	0.5	0.631579	0.622827
4431	10.0	100.0	0.5	0.644737	0.646916
4613	50.0	100.0	0.3	0.631579	0.621429
5617	100.0	50.0	0.5	0.786667	0.666667
5669	200.0	50.0	0.5	0.710526	0.570196
5791	50.0	50.0	0.5	0.644737	0.626926
5952	200.0	200.0	0.5	0.685185	0.680556
6973	100.0	10.0	0.5	0.671053	0.639231
6969	100.0	50.0	0.5	0.657895	0.656965
7669	100.0	200.0	0.5	0.565789	0.571739
9138	100.0	10.0	0.5	0.723684	0.617749
9150	200.0	100.0	0.2	0.723684	0.669312

3.5.3.5 Temporal Convolutional Network

Unlike the other models which were fit using the frequency features of the EEG data, the time-series EEG data was used to fit a TCN. The TCN model used was inspired by Bai et al.'s model and was created through hyperparameter testing of kernel size, dilations, filters, and stacks [91]. The hyperparameters tested were:

- Number of filters - 5, 10, 15 and 20
- Kernel widths - 18, 32, 64
- Dilations - [2, 4, 8, 16], [4, 8, 16, 32], and [8, 16, 32, 64]
- Stacks - 4, 5, 6, and 7

These hyperparameters were chosen to ensure an adequate receptive field for the model. The equation for calculating the receptive field for a given convolutional layer,

l, and dilation rate, d is given in Equation (1).

$$receptiveField(l) = receptiveField(l-1) + [kernelSize - 1] * d \quad (1)$$

With the smallest kernel size of 18 and the smallest dilations of [2, 4, 8, 16], the network is designed to have at least a receptive field of 511 samples, or roughly 2 seconds, which allows it a sufficient receptive field to cover the entire input sample. Please see Table 12 for a complete list of all possible receptive field sizes.

Table 12. Receptive field sizes for all possible combinations of kernel widths and dilations.

Kernel Width	Dilations	Receptive Field
18	2, 4, 8, 16	511
	4, 8, 16, 32	1021
	8, 16, 32, 64	2041
32	2, 4, 8, 16	931
	4, 8, 16, 32	1861
	8, 16, 32, 64	3721
64	2, 4, 8, 16	1891
	4, 8, 16, 32	3781
	8, 16, 32, 64	7561

The kernel width denotes the width of the convolutional kernel size. The greater the width, the more data over which the kernel convolves. An increased kernel size could generate a better prediction due to this greater amount of data [101]. Likewise, the layers of dilations causes the effective receptive field of units to grow exponentially with layer depth even though the number of parameters grows only linearly [102]. Although some of the combinations of kernel widths and dilations overlap in their receptive field size, there is a chance that a different kernel width paired with the same dilations could capture data that another kernel width failed to capture.

The first layer in the TCN was a one-dimensional convolutional layer with 64

filters, a kernel size of 10, a dilation rate of 1, and a padding of 'causal.' This layer was added so that the architecture can take a sequence of any length and map it to an output sequence of the same length [91]. The TCN layers followed this first layer. The last layer of the model was a Dense layer with one output and an activation function of sigmoid. The TCN model architecture can be seen in Figure 30.

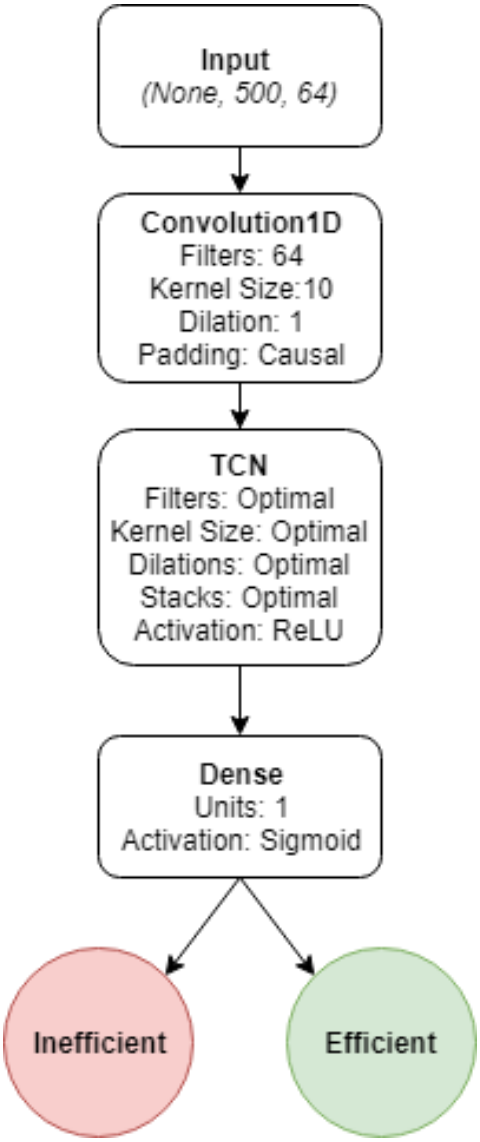

Figure 30. The TCN model architecture.

All models utilized a loss function of binary cross-entropy and used Adam as the optimizer. To regularize, batch normalization and a dropout of 25% were used. To select the best model, a validation-accuracy based early stopping with a patience of 5 epochs and a delta of 0.001 was used. All networks were trained with a batch size of 32.

Similar to the LSTM hyperparameter sweep, a threshold sweep was also performed.

A model for each participant was tuned using the parameters above. The results of the tuning can be seen in Table 13.

Table 13. The TCN optimal hyperparameters for each participant

Participant	Filters	Kernel Size	Dilations	Stacks	Threshold	Val Accuracy	Val Balanced Accuracy
0271	15	64	[2, 4, 8, 16]	4	0.7	0.697368	0.707893
1437	15	64	[8, 16, 32, 64]	6	0.5	0.710526	0.711648
2070	15	32	[8, 16, 32, 64]	6	0.8	0.750000	0.702862
2765	5	64	[2, 4, 8, 16]	4	1.0	0.739726	0.707680
4030	5	64	[4, 8, 16, 32]	5	0.6	0.671053	0.673943
4431	15	32	[2, 4, 8, 16]	5	0.7	0.697368	0.698611
4613	15	64	[2, 4, 8, 16]	5	0.8	0.750000	0.756734
5617	10	32	[2, 4, 8, 16]	5	0.6	0.706667	0.708929
5669	15	18	[4, 8, 16, 32]	5	0.1	0.750000	0.662458
5791	5	18	[4, 8, 16, 32]	5	1.0	0.671053	0.670290
5952	10	18	[2, 4, 8, 16]	4	0.1	0.740741	0.742069
6973	10	18	[2, 4, 8, 16]	4	1.0	0.750000	0.735119
6969	5	18	[4, 8, 16, 32]	5	0.4	0.723684	0.707971
7669	5	32	[2, 4, 8, 16]	6	0.1	0.684211	0.691667
9138	15	64	[2, 4, 8, 16]	4	0.3	0.776316	0.734848
9150	5	32	[2, 4, 8, 16]	6	0.6	0.736842	0.693603

3.5.4 Cross-Participant Models

Within-participant models generally perform well as a single participant's data is less variable than multiple participants' data. However, a cross-participant model is robust in that it is more genaralizable and can be applied to more than a single

participant. To create a cross-participant model, the highest-performing models from the within-participant models will be considered.

To train a cross-participant model, the data was split into train, validation, and test datasets. 13 participants were used as the train set, 2 were used as the validation set, and one's participant's data was used to test the model. The participants rotated which set they belonged to so that the model tested each participant's individual data. For the cross-participant models, the same hyper-parameters were tuned as in the within-participant models.

The results from the hyperparameter tuning for the RFC can be seen in Table 14. The hyperparameters from the model with the highest tuned MCC were chosen to be the final hyperparameters.

Table 14. The results of the cross-participant hyperparameter tuning for the RFC model.

Participant Left Out	Initial MCC	Tuned MCC	Number of Best Trees	Number of Best Features	Best Depth
0271	0.104060	0.209738	20	11	3
1437	0.103668	0.183697	130	20	9
2070	0.063555	0.188369	160	2	6
2765	0.111462	0.174800	100	1	5
4030	0.040205	0.186051	100	20	3
4431	0.100398	0.178894	70	2	4
4613	0.136622	0.208469	330	20	8
5617	0.098671	0.198390	100	10	4
5669	0.072923	0.230696	30	20	10
5791	0.099991	0.182309	210	23	2
5952	-0.000076	0.204645	180	25	1
6973	0.025042	0.191780	10	9	2
6969	0.070725	0.201183	430	1	4
7669	0.102238	0.198265	170	1	6
9138	0.141289	0.189731	300	1	5
9150	0.101242	0.206003	430	2	5

The results from the hyperparameter tuning for the ANN can be seen in Table 15.

The hyperparameters from the model with the highest balanced accuracy were chosen to be the final hyperparameters.

Table 15. The cross-participant hyperparameter tuning results for the ANN model.

Participant Left Out	Hidden Layers	Width	Learning Rate	Threshold	Validation Accuracy	Validation Balanced Accuracy
0271	8.0	256.0	0.001	0.7	0.577723	0.581051
1437	8.0	128.0	0.010	0.8	0.570379	0.576326
2070	2.0	64.0	0.010	0.5	0.597307	0.594459
2765	3.0	128.0	0.010	0.5	0.560588	0.573937
4030	3.0	128.0	0.001	0.5	0.578947	0.512744
4431	5.0	64.0	0.010	0.5	0.596083	0.591312
4613	2.0	256.0	0.001	0.5	0.620563	0.590035
5617	7.0	256.0	0.010	0.6	0.594859	0.582298
5669	6.0	128.0	0.001	0.5	0.591187	0.588737
5791	2.0	128.0	0.001	0.5	0.565483	0.561324
5952	7.0	256.0	0.001	0.5	0.575472	0.574480
6973	6.0	128.0	0.001	0.4	0.567086	0.563052
6969	3.0	256.0	0.001	0.5	0.567227	0.571324
7669	7.0	128.0	0.010	0.5	0.585084	0.588915
9138	6.0	256.0	0.001	0.4	0.563025	0.569215
9150	6.0	64.0	0.001	0.5	0.585084	0.587152

3.5.5 Performance Analysis

The ML models' performance will be analyzed using balanced accuracy. Because the number of inefficient and efficient searches are imbalanced, a useful metric is the balanced accuracy. Balanced accuracy accounts for this class imbalance and determines the average recall on each class. This metric is a better indicator of model performance when the classes are imbalanced because it better illustrates how the model performed when predicting both classes. To better illustrate classification errors, a confusion matrix will be used (Figure 31). Confusion matrices show the true positive rate (TPR), true negative rate (TNR), false positive rate (FPR), and false negative rate (FNR) and provide a closer examination into what kind of classification

mistakes the model is making.

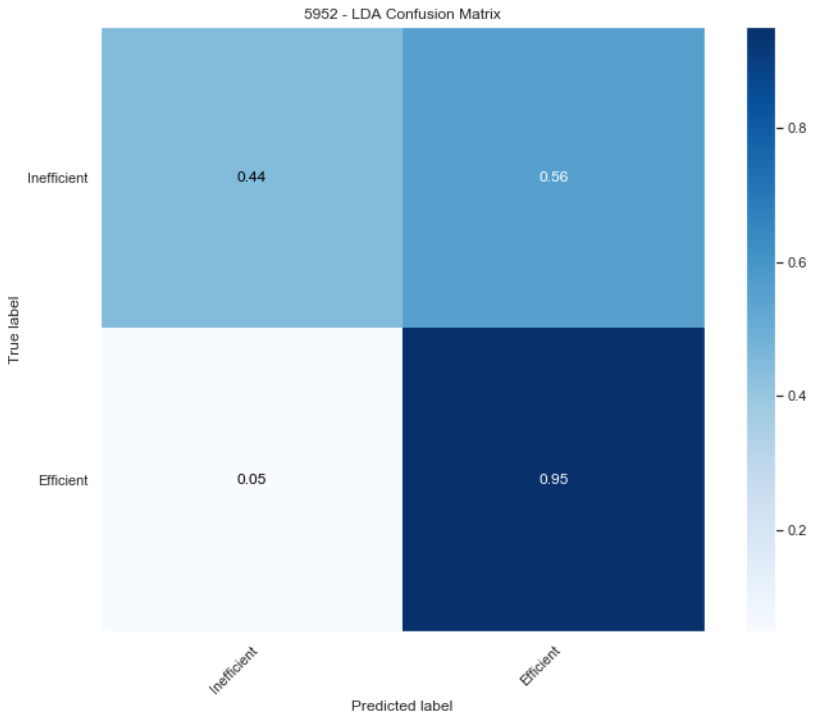

Figure 31. A confusion matrix.

Both a model's accuracy and its confusion matrix depend on tuning the thresholds that are used to classify a model's probability output. Therefore, the Area Under the Receiver Operating Characteristic Curve (AUROC) will be used to determine the overall best model. A Receiver Operating Characteristic (ROC) curve is made by plotting a model's TPR versus its FPR at various classification thresholds. An example of a ROC curve can be seen in Figure 32. An AUROC of 1.0 indicates a perfect model, one that correctly classifies all observations, while a score of 0.0 indicates a model that failed to properly classify any observations. Using an AUROC can indicate how well the data is being modeled.

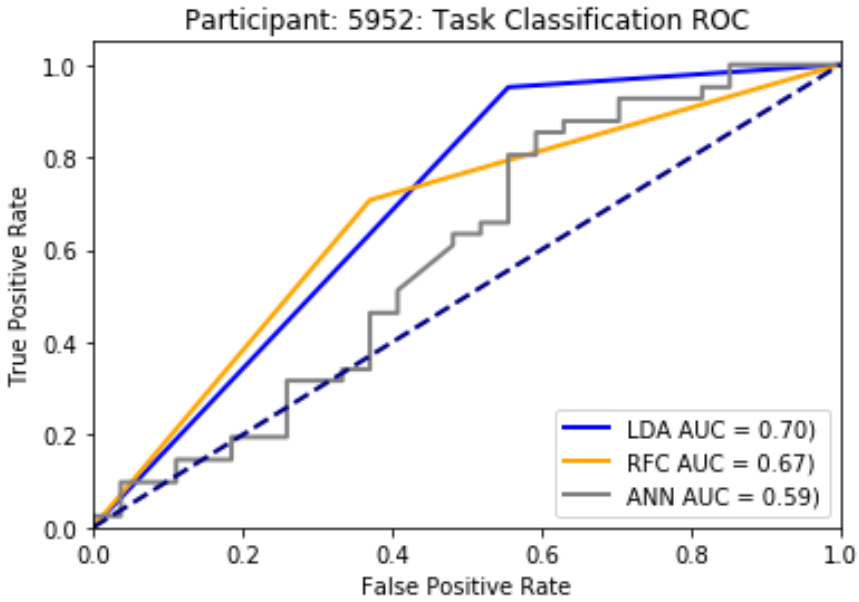

Figure 32. A ROC curve.

An important aspect of determining the performance of a model is determining where it does not perform well. The error analysis for the frequency dataset will focus on analyzing how classification errors are associated with each type of search. If efficient and inefficient searches are consistently producing classification errors, then the models may not be generalizing well across searches during the experiment. For the time series dataset, the error analysis will focus on the data segmentation. For this research, the data segmentation was chosen to be the two seconds before the participant selected an answer. However, there are many different ways to create epochs and to segment the data. Future work will be necessary to determine which is the best way to segment epochs from the time series data.

3.6 Summary

In summary, this chapter reviewed the previous work undertaken in identifying a confirmation bias during a visual search. It outlined a visual search experiment that was used to determine how effective various mitigation techniques are in mitigating an inefficient search pattern. The experiment consisted of participants participating in a visual search experiment in which they determine the color associated with a given letter. Next, information about the setup, equipment, and analysis of the experiment was reviewed. Finally, a machine learning pipeline was devised which discussed pre-processing the EEG data, the creation of various models to determine the relationship between efficient searches and EEG data, and the analysis of the data.

IV. Analysis and Results

4.1 Chapter Overview

This chapter provides an in-depth analysis of the results obtained from the experiment outlined in Chapter 3. The results include behavioral measures of efficient searches performed during the Efficient Search Experiment (ESE). Section 4.2 describes the subjective results of the ESE, which includes the participants' initial Visual Search Pattern (VSP) and the effect of the mitigations on these patterns. Section 4.3 describes non-subjective measures and covers results associated with the electrophysiological data that was collected during the ESE; the results in this section detail the machine learning performance metrics on classifying Electroencephalography (EEG) as well as the EEG time series analysis. The results in this chapter provide justification for answering all research questions in Section 3.3.

4.2 Behavioral Analysis and Results

Behavioral analysis contains all data recorded during the ESE that is not a physiological component. This includes:

- Accuracy Results
- Timing Results
- Initial visual search patterns
- Effects of mitigation on visual search patterns

Accuracy is an important metric to consider because it dictates how important is it that the participant uses an efficient search. If participants are using an efficient VSP but they are less accurate than when using an inefficient VSP, then it is not worthwhile to encourage the participants to use an efficient VSP.

Similarly, timing is a key metric to examine. If a participant's search times increased dramatically when using an efficient search pattern, then it may not be plausible to use an efficient search on a time-sensitive task.

The initial VSPs of participants are important to determine because it is necessary to decide whether a change has occurred after applying the mitigation techniques. A hypothesis proposed by previous work is that humans naturally will use a confirmatory search pattern because it is simple [74]. However, no research has yet been conducted that determines that humans resort to a confirmatory VSP instead of a circular VSP, a methodical scanning VSP, or any other type of VSP. Analysis of the participants' initial, unmitigated VSPs will help determine to which VSP humans naturally resort.

The main goal of this research is to determine whether mitigation techniques during a visual search can successfully change a participant's initial inefficient VSP to a more efficient VSP. Each mitigation technique (the nudge, the explanation, and the instructions) will be analyzed to determine its effect on the VSP of the participant.

4.2.1 Accuracy Results

Overall, participants had a high accuracy with an overall mean of 95.03% with a standard deviation of 2.15%. Each participant's accuracy score can be seen in Figure 33.

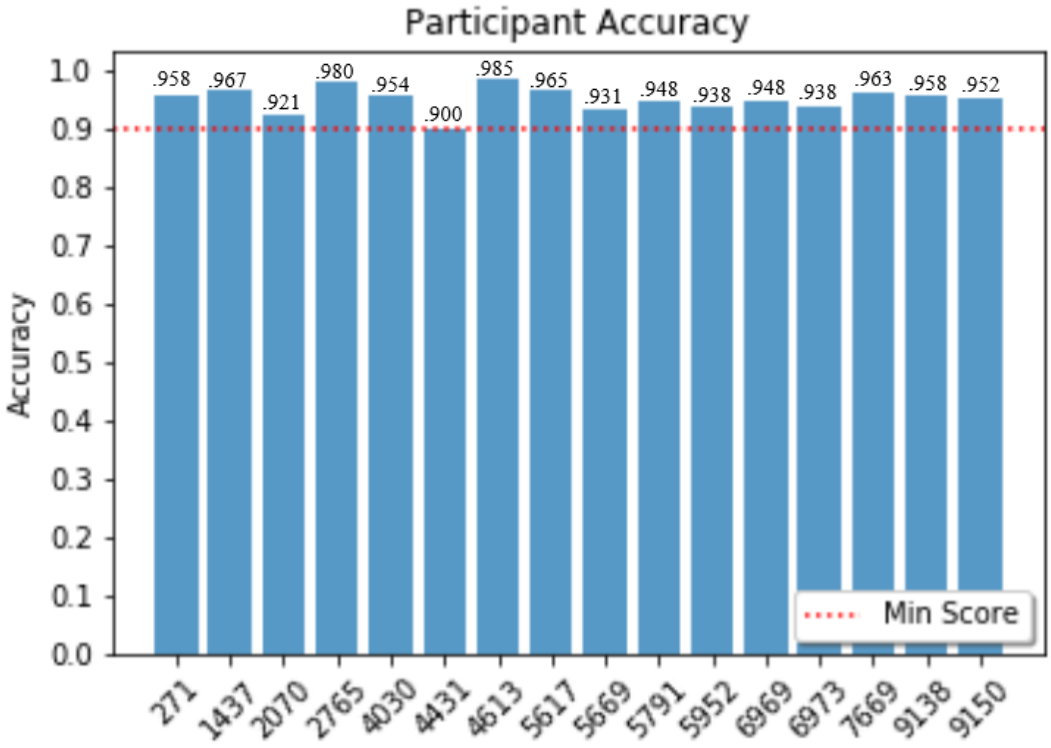

Figure 33. Participants' search accuracy.

Each participant's accuracy during efficient and inefficient searches is plotted in Figure 34 (with 95% confidence intervals) and can be seen in Table 16.

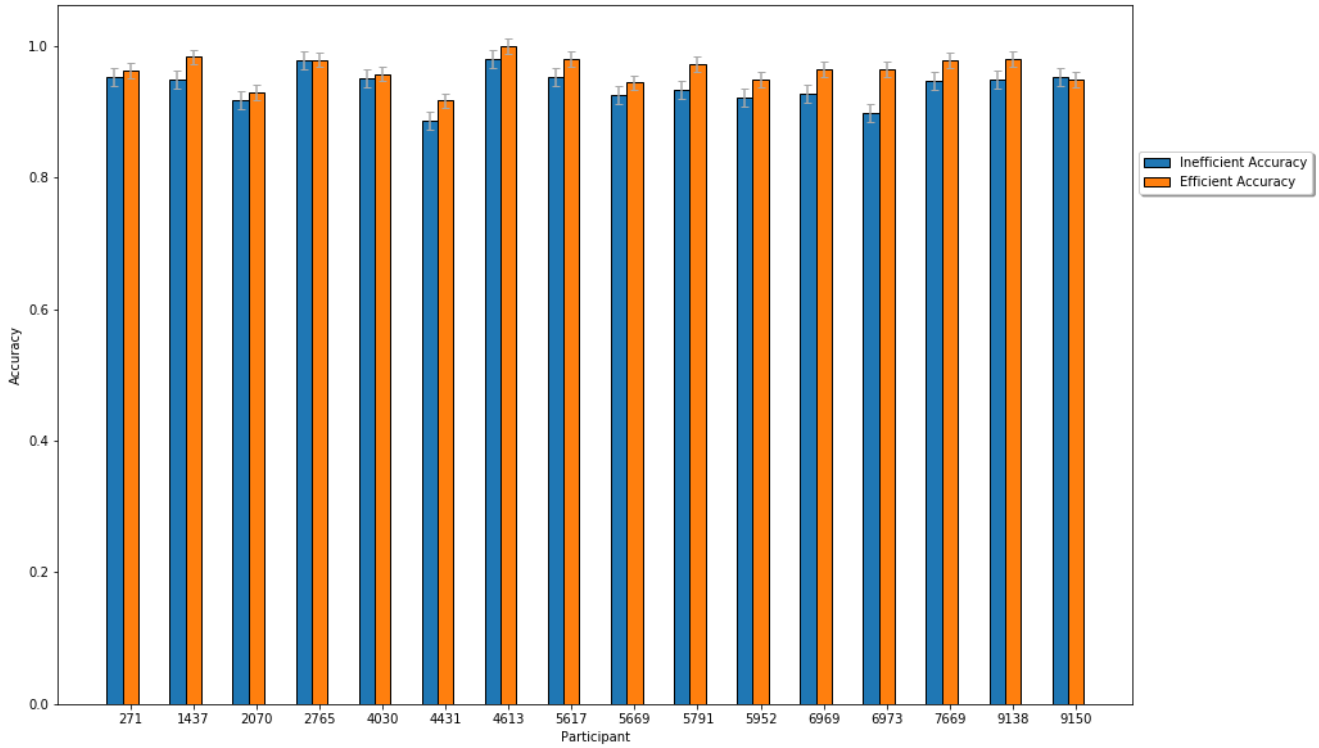

Figure 34. Participants' average accuracy.

Table 16. Search accuracy per search type

Participant	Efficient Correct	Inefficient Correct
271	96.30%	95.36%
1437	98.34%	94.98%
2070	92.91%	91.74%
2765	97.93%	97.91%
4030	95.77%	95.13%
4431	91.71%	88.59%
4613	100.00%	98.09%
5617	98.06%	95.26%
5669	94.44%	92.56%
5791	97.19%	93.38%
5952	94.93%	92.16%
6969	96.54%	92.73%
6973	96.45%	89.90%
7669	97.86%	94.72%
9138	97.97%	94.88%
9150	94.93%	95.32%

To determine whether there was a difference in the accuracies between efficient and inefficient searching, a two-sample paired t-test was performed. The hypotheses for this test were:

- Null hypothesis: The accuracy of efficient and inefficient searches are the same
- Alternate hypothesis: The accuracy of efficient and inefficient searches are different

The test had an α of 0.05 and the degrees of freedom were the number of participants minus one, or 15. A paired t-test has the assumptions that the data has no outliers and that the data is normally distributed. The boxplot in Figure 35 indicates that there are no outliers while the Q-Q plot in Figure 36 indicates that the data is normally distributed. Additionally, a Shapiro-Wilk test was performed to ensure the data was normally distributed. The test resulted in a p-value of 0.45 which fails to reject the

null hypothesis that the data is normally distributed. Thus, the assumptions of the t-test hold.

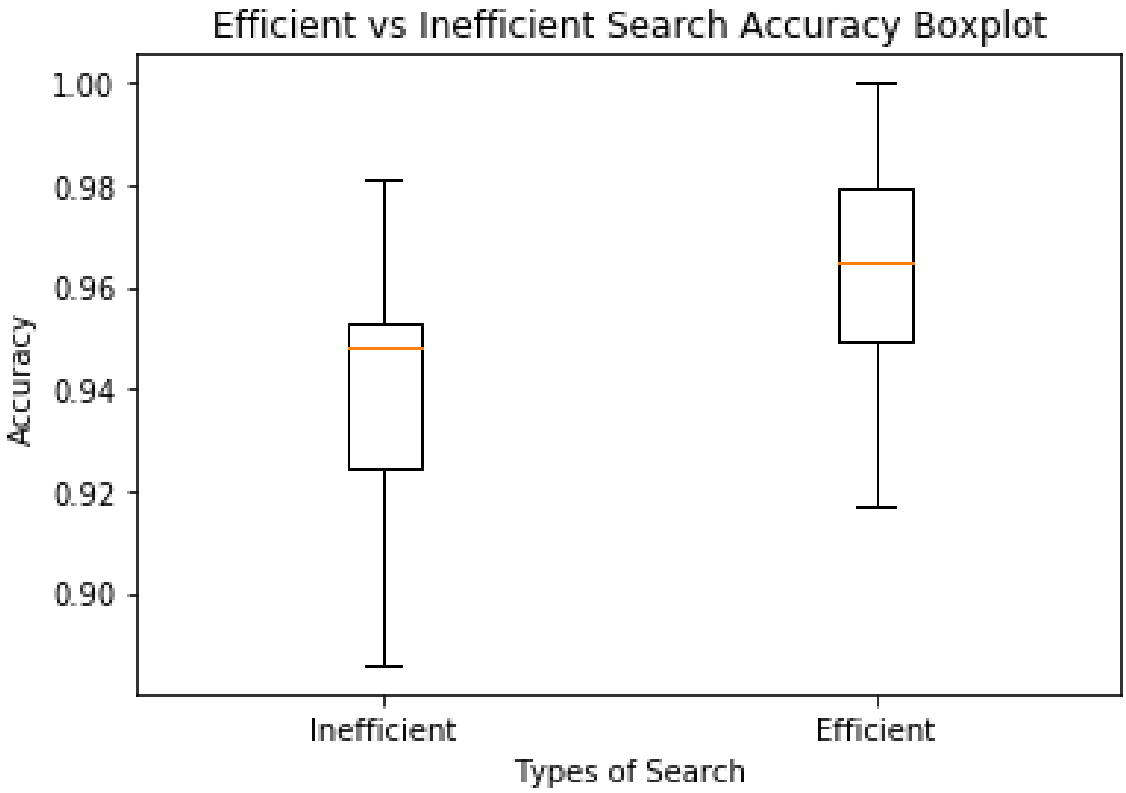

Figure 35. There does not appear to be any outliers in either the efficient accuracies or the inefficient accuracies.

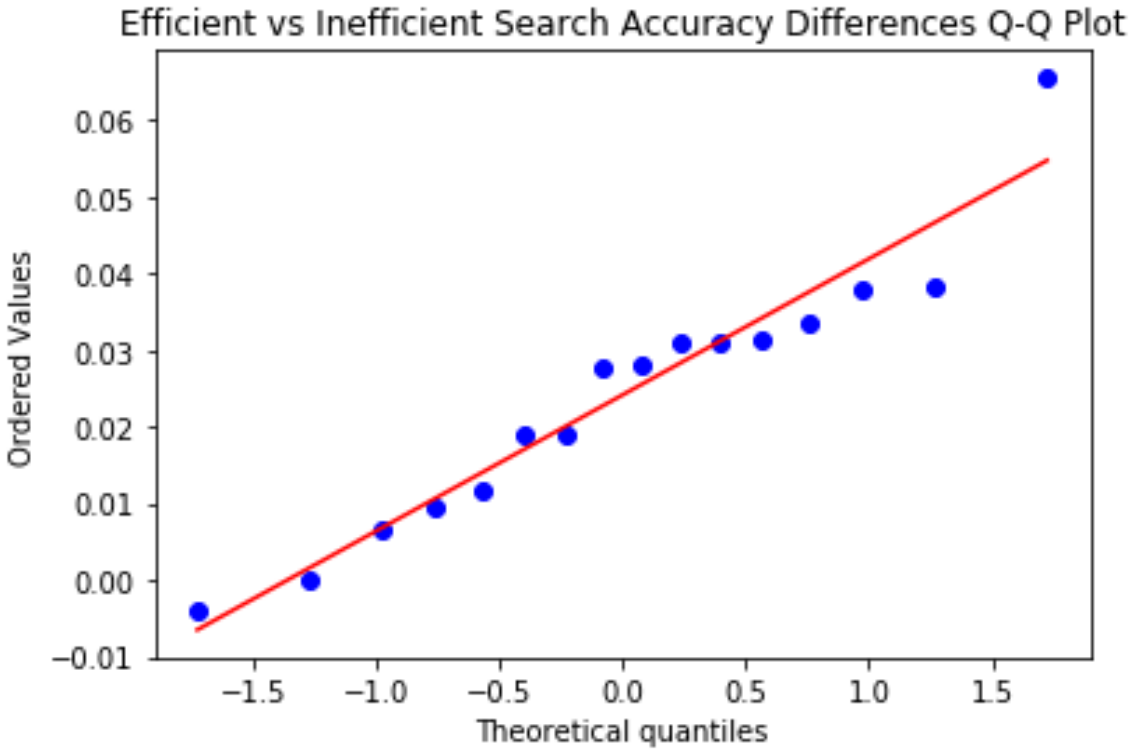

Figure 36. The assumption of normal distribution holds.

The accuracy during an efficient search was higher (96.33% +- 2.16%) compared to the accuracy during an inefficient search (93.92% +- 2.57%); there was a statistically significant increase in accuracy (t(15)=5.59, p= 0.00005) of 2.41%.

Thus, efficient searches are more accurate than inefficient searches.

4.2.2 Timing Results

Overall, participants' average search time was 2.17 seconds with a standard deviation of 1.17 seconds. The minimum search time was 0.60 seconds while the maximum search time was 17.81 seconds. High search times could be due to issues with the gaze tracking software. To minimize these issues, experiment administrators monitored participant performance. If it seemed that the gaze tracker was suffering

from accuracy issues, e.g. the participant was clearly attempting to fixate upon a specific stimuli and the gaze tracker was not registering the fixation attempt, then the test administrators performed a recalibration with the participant. Each participant's average search time can be seen plotted in Figure 37 and in Table 17.

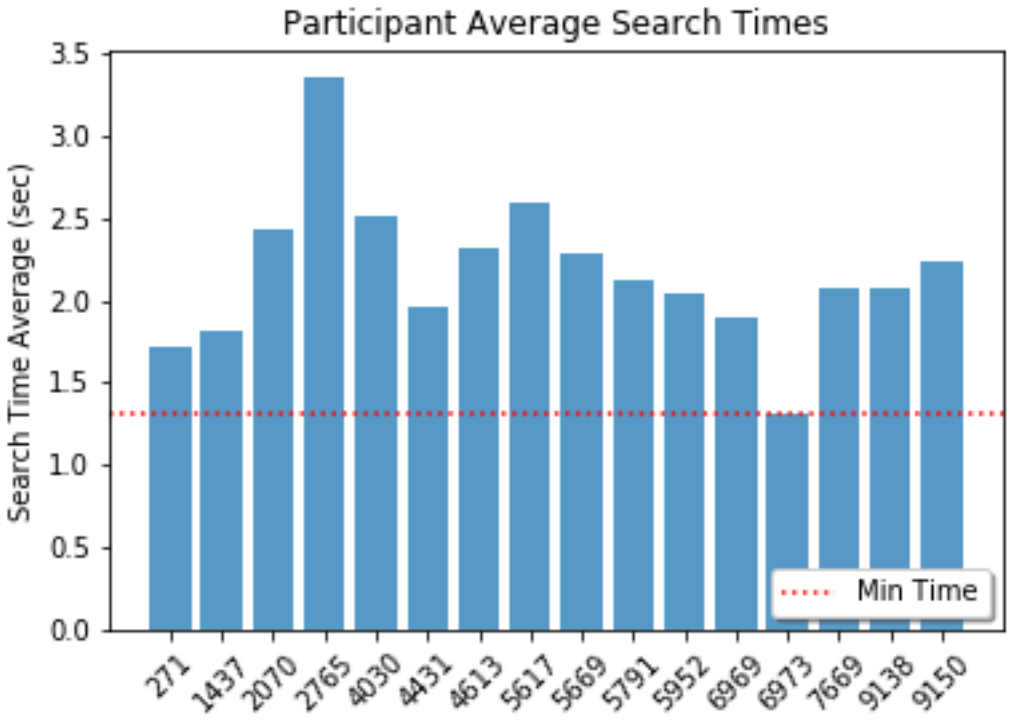

Figure 37. Participants' average search time.

Table 17. Average search times

Participant	Average Search Time (sec)
271	1.722709
1437	1.821991
2070	2.430740
2765	3.354313
4030	2.513310
4431	1.960416
4613	2.310181
5617	2.598590
5669	2.289290
5791	2.126426
5952	2.046704
6969	1.903119
6973	1.309763
7669	2.074877
9138	2.074976
9150	2.241542

Each participant's average search time during efficient and inefficient searches is plotted in Figure 38 and can be seen in Table 18.

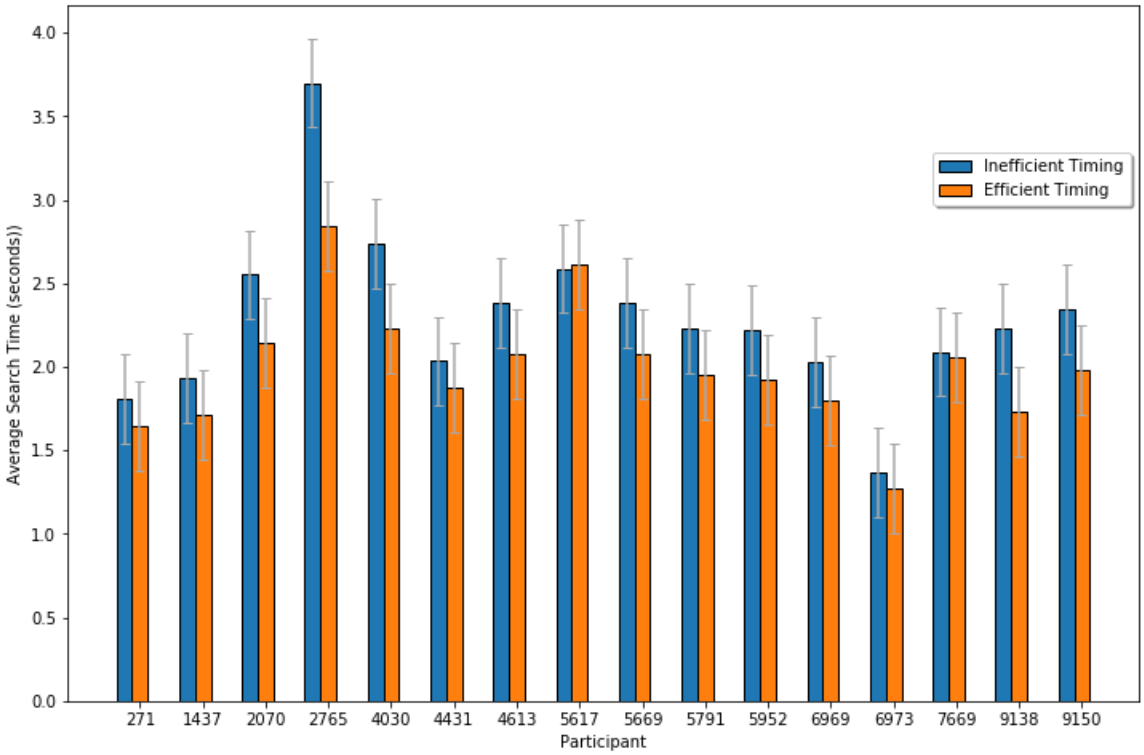

Figure 38. Participants' average search times per type of search.

Table 18. Average search times per type of search

Participant	Efficient Average Search Time	Inefficient Average Search Time
271	1.641653	1.805817
1437	1.713109	1.931784
2070	2.142853	2.550480
2765	2.841759	3.698993
4030	2.229442	2.739765
4431	1.871430	2.033837
4613	2.077665	2.381773
5617	2.612755	2.587940
5669	2.073935	2.381585
5791	1.948505	2.231294
5952	1.919134	2.219299
6969	1.799368	2.025734
6973	1.271175	1.364721
7669	2.059425	2.089576
9138	1.734453	2.226776
9150	1.982148	2.346209

To determine whether there was a difference in the search times between efficient and inefficient searching, a two-sample paired t-test was performed. The hypotheses for this test were:

- Null hypothesis: The average search times of efficient and inefficient searches are the same
- Alternate hypothesis: The average search times of efficient and inefficient searches are different

The test had an α of 0.05 and the degrees of freedom were the number of participants minus one, or 15. A paired t-test has the assumptions that the data has no outliers and that the data is normally distributed. The boxplot in Figure 39 indicates that there are outliers. However, the outliers are balanced for both positive and negative and for both types of searches. Because the normality assumption holds, and the

outliers are not too severe, we can ignore this violation. The Q-Q plot in Figure 40 indicates that the data is normally distributed. Additionally, a Shapiro-Wilk test was performed to ensure the data was normally distributed. The test resulted in a p-value of 0.27 which confirms that the data is normally distributed. Thus, the assumptions of the t-test hold.

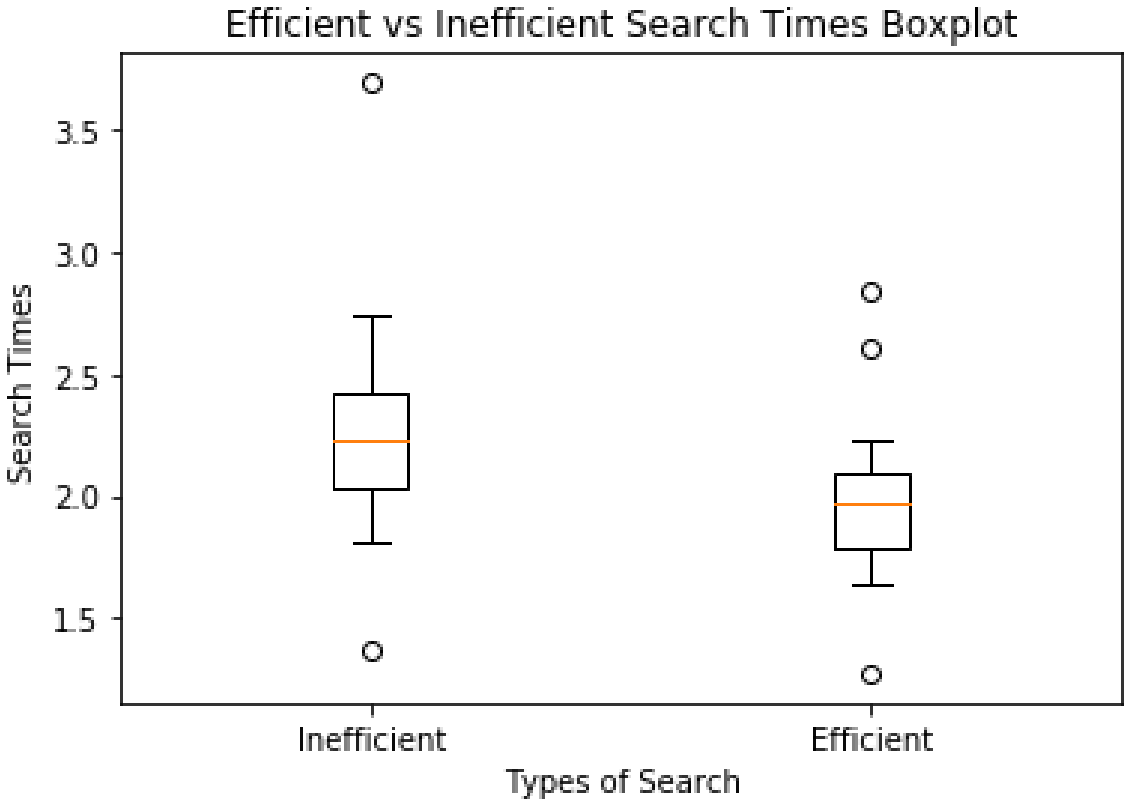

Figure 39. While there are outliers, the outliers are balanced on both searches and are both high and low. Because normality is not violated, we can ignore the violation of this assumption.

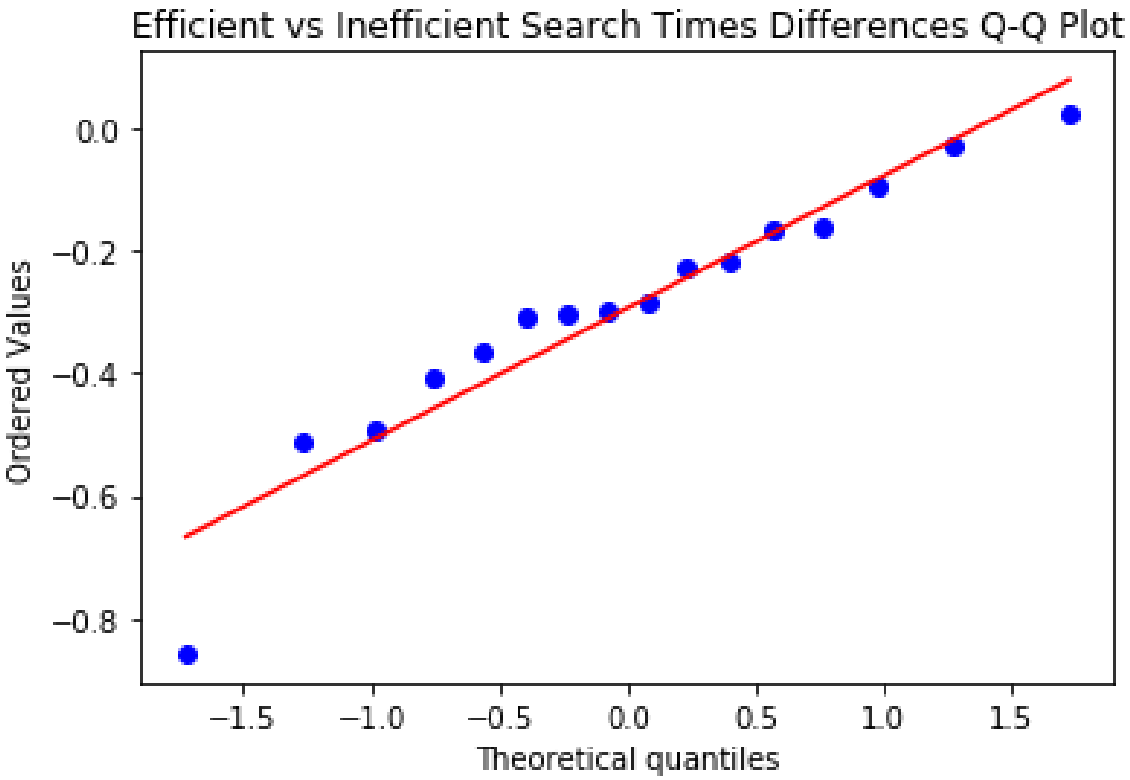

Figure 40. The assumption of normal distribution holds.

The average search time during an efficient search was faster (1.99 +- 0.37 seconds) compared to the average search time during an inefficient search (2.29 +- 0.50); there was a statistically significant increase in average search times (t(15)=5.53, p= 0.00005) of 0.30 seconds.

Thus, efficient searches are faster than inefficient searches.

4.2.3 Initial Visual Search Patterns

The initial VSPs for each participant were identified by analyzing the VSPs that the participant used during the first eight blocks. This includes all of the blocks in the "clean" blocks as well as the first block in the "nudged" blocks because up to this point the participant had not encountered any mitigation techniques. As such,

it can be assumed that the VSP that the participant used in the first eight blocks is the VSP to which the participant naturally resorts. The results of this section serve to answer research question 1 in Section 3.3.1.

A stacked bar graph was generated where each participant that displays the percentage of inefficient, efficient, and circular searches during the first eight blocks. This graph can be seen in Figure 41. Additionally, a table detailing individual search types can be seen in Table 19 and Table 20. In this experiment, a circular search is also an inefficient search - i.e., these two categories are not mutually exclusive. For details on how each search type was determined, please see Section 3.4.4.

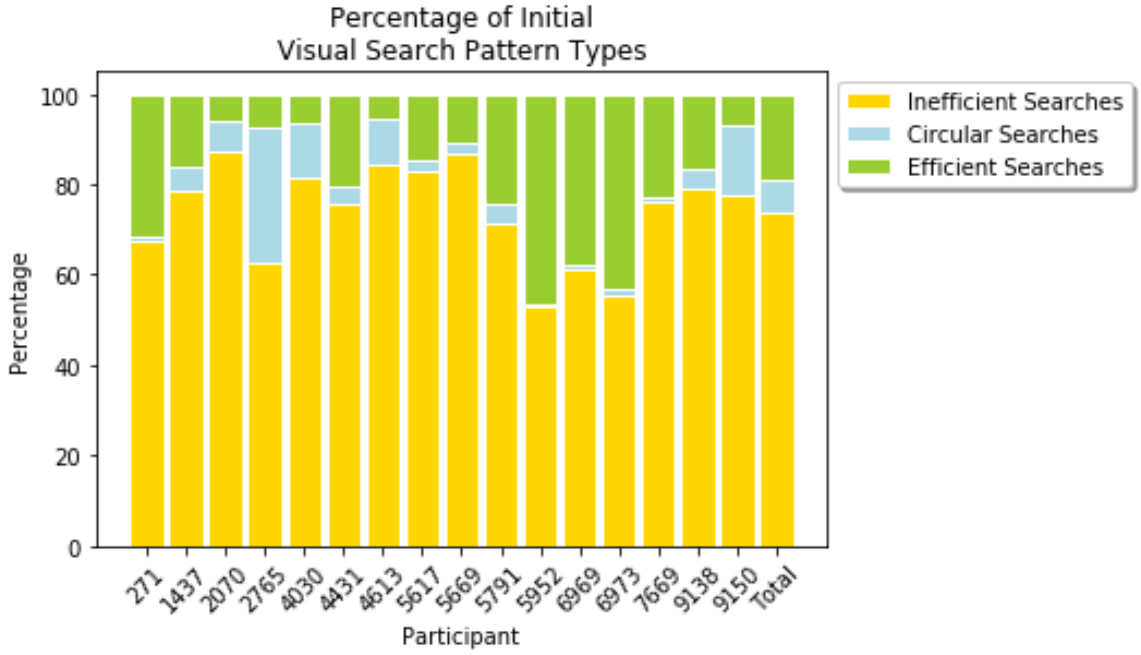

Figure 41. The initial VSPs for all participants.

Table 19. Visual Search Pattern types in the first eight blocks.

ID	Confirmatory Searches	Non-Confirmatory Searches	Efficient Searches	Inefficient Searches	Circular Searches
271	28	132	51	109	2
1437	140	20	27	133	9
2070	63	97	10	150	12
2765	52	108	17	143	69
4030	49	111	12	148	22
4431	140	20	34	126	7
4613	63	97	10	150	18
5617	138	22	24	136	4
5669	104	56	18	142	4
5791	39	121	41	119	7
5952	45	115	75	85	1
6969	52	108	61	99	2
6973	42	118	70	90	2
7669	43	117	37	123	2
9138	146	14	28	132	7
9150	55	105	13	147	30
Total	1199	1361	528	2032	198

Table 20. Visual Search Pattern types (percentages) in the first eight blocks

Participant	Percent Inefficient	Percent Efficient	Percent Circular
271	67.28%	31.48%	1.23%
1437	78.70%	15.98%	5.33%
2070	87.21%	5.81%	6.98%
2765	62.45%	7.42%	30.13%
4030	81.32%	6.59%	12.09%
4431	75.45%	20.36%	4.19%
4613	84.27%	5.62%	10.11%
5617	82.93%	14.63%	2.44%
5669	86.59%	10.98%	2.44%
5791	71.26%	24.55%	4.19%
5952	52.80%	46.58%	0.62%
6969	61.11%	37.65%	1.23%
6973	55.56%	43.21%	1.23%
7669	75.93%	22.84%	1.23%
9138	79.04%	16.77%	4.19%
9150	77.37%	6.84%	15.79%
Total	73.68%	19.14%	7.18%

The results indicate that all 16 (100%) participants initially used a primarily inefficient search pattern. Overall, 73.7% of the searches performed in the first eight blocks were inefficient, 19.1% were efficient, and 7.2% were circular.

Thus, the majority of participants initially used an inefficient search pattern.

4.2.4 Mitigation Effects on Visual Search Patterns

Once the VSP for each participant has been identified, the next step is to determine what effect each mitigation attempt had on the participant's VSP. The main purpose of this section is to answer research question 3 in Section 3.3.3. For details on all mitigations applied during the ESE, please see Section 3.3.

The data was first explored by creating a line graph that plotted the number of efficient searches against the block number. This graph was expected to show that as

the block number increased, so would the number of efficient searches due to increased mitigations. These graphs can be seen in Figure 42, and Figure 43 for all participants. The mitigated blocks are indicated as red dots and generally coincide with a higher efficient search number for that block. The locations at which the other mitigations are applied are also indicated. A trend line was fitted to the number of efficient searches and it is positive for all participants. This positive trend line indicates that all participants increased the number of efficient searches as the block number increased. However, this trend line alone can not detail why the number of efficient search per block increased. Additionally, due to changes in the experimental procedure, participants 4030, 2070, 2765, 271, 4613, 5791, 6969, and 1437 did not receive the nudge in the last seven blocks. For details on this, please reference Section 3.4.4.

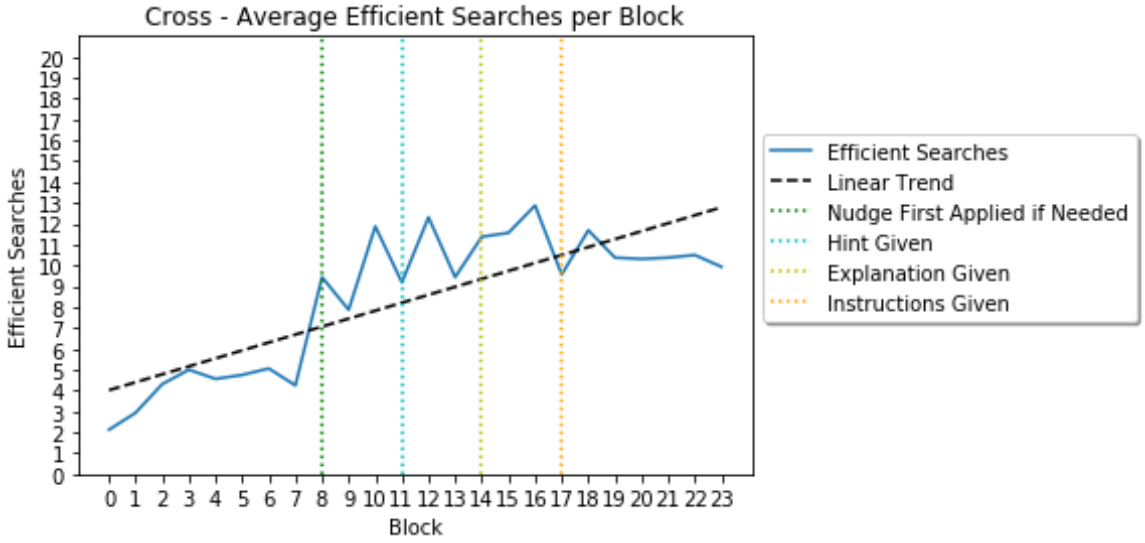

Figure 43. The number of efficient searches vs block number for all participants.

To find each mitigation technique's effect on the number of efficient searches, SAS's statistical software JMP was used to perform a linear regression for each participant. An example of a data table used for the linear regression can be seen in Figure 44. The

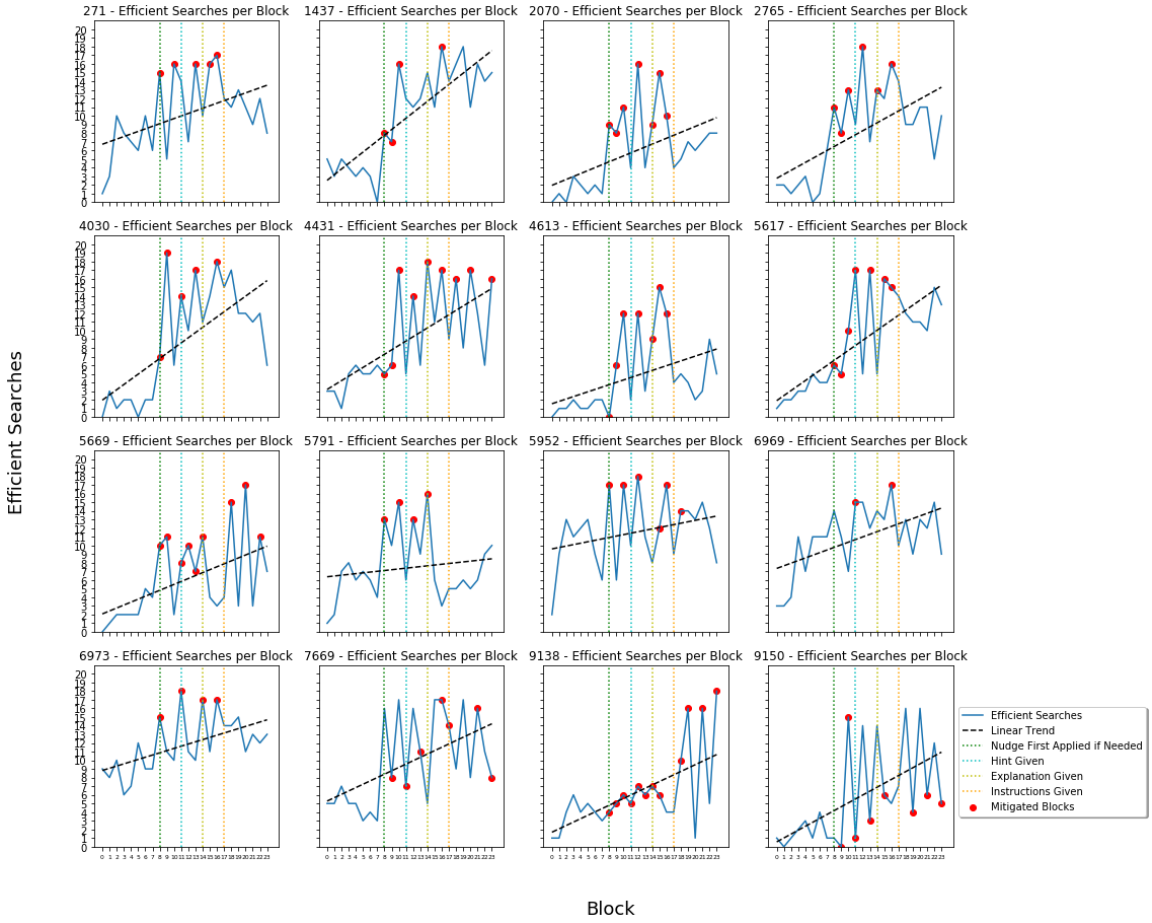

Figure 42. The number of efficient searches vs block number. Mitigation attempts and their locations are indicated on the graphs. The trend line of efficient searches is plotted.

number of efficient searches was the response variable, while the presence or absence of a nudge, hint, explanation, or instruction were the independent variables. The nudge was marked as present or absent for each participant depending on the actual blocks that the participant received the nudge, while for all participants the hint was marked as present for block 11 and subsequent blocks, the explanation was marked as present for block 14 and subsequent blocks, and the instructions were marked as present for block 18 and subsequent blocks. This was implemented because once the participant had encountered the hint, explanation, or the instructions, it impacted the participant for the remainder of the experiment.

Each question on the pre- and post-experiment questionnaires was assigned a numerical value. A linear regression was performed on each of these variables with the response variable being the number of efficient searches. No significant effect was found.

Nudge	Hint	Explanation	Instructions	Efficient
0	0	0	0	2
0	0	0	0	9
0	0	0	0	13
0	0	0	0	11
0	0	0	0	12
0	0	0	0	13
0	0	0	0	9
0	0	0	0	6
1	0	0	0	17
0	0	0	0	6
1	1	0	0	17
0	1	0	0	10
1	1	0	0	18
0	1	1	0	11
0	1	1	0	8
1	1	1	0	12
1	1	1	0	17
0	1	1	1	9
1	1	1	1	14
0	1	1	1	14
0	1	1	1	13
0	1	1	1	14
0	1	1	1	12
0	1	1	1	8

Figure 44. The data table used to perform a linear regression participant 5952. A "1" in a column indicates the presence of the mitigation technique while a "0" indicates its absence.

The method used for linear regression was standard least squares. Using this method, the effects of each individual mitigation method were determined. Additionally, the interactions between the mitigation methods were also determined. Due to confounding of variables, only three interactions were able to be determined: Nudge * Hint, Nudge * Explanation, and Nudge * Instructions. Thus, the full linear model used

was:

$$y = \beta_0 + \beta_1 * nudge + \beta_2 * hint + \beta_3 * explanation + \beta_4 * instructions + \beta_5 * nudge * hint + \beta_6 * nudge * explanation + \beta_7 * nudge * instructions \quad (2)$$

The overall effects of each mitigation technique can be seen in Figure 45 while the coefficients can be seen in Table 21. The log worth is the JMP software's version of the p-value of a split, and is calculated by taking the $-log_{10}$ of the p-value. Typically, values above two are considered significant. The higher the log worth, the more impact that the variable has on the response. The effects for each participant for each mitigation technique can be seen in Figure 46. Additionally, learning effect, or the likelihood that participants were learning how to search more efficiently simply as the experiment continued without any effect from the mitigation techniques, was investigated. The learning effect was investigated by performing the above linear regression with the block number as an additional independent variable. The results of this test determined that the block number was a significant effect on only one participant, 6969. For participant 6969, the log worth of the block number was 3.000. On all other participants, there was no evidence of learning effect.

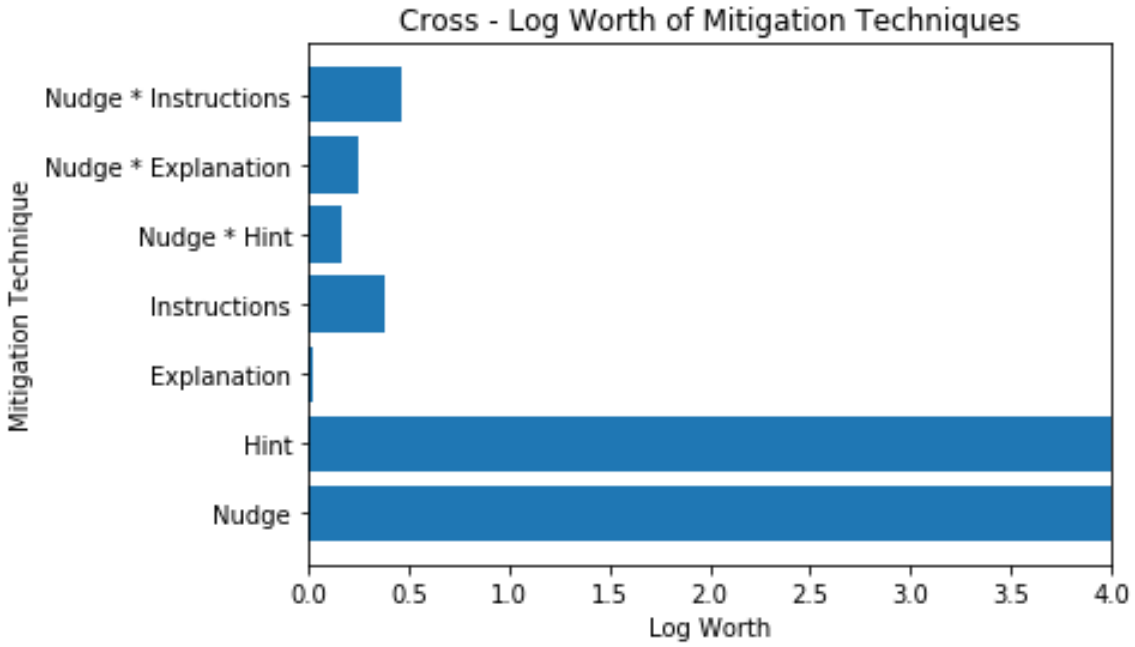

Figure 45. The log worth for all participants. The larger the blue bar, the more of an effect the mitigation technique had.

Table 21. Each mitigation technique's coefficient for the linear regression model

Mitigation Technique	**Coefficient**
Nudge	3.70
Hint	4.60
Explanation	0.05
Instructions	0.56
Nudge * Hint	-0.59
Nudge * Explanation	0.91
Nudge * Instructions	-1.41

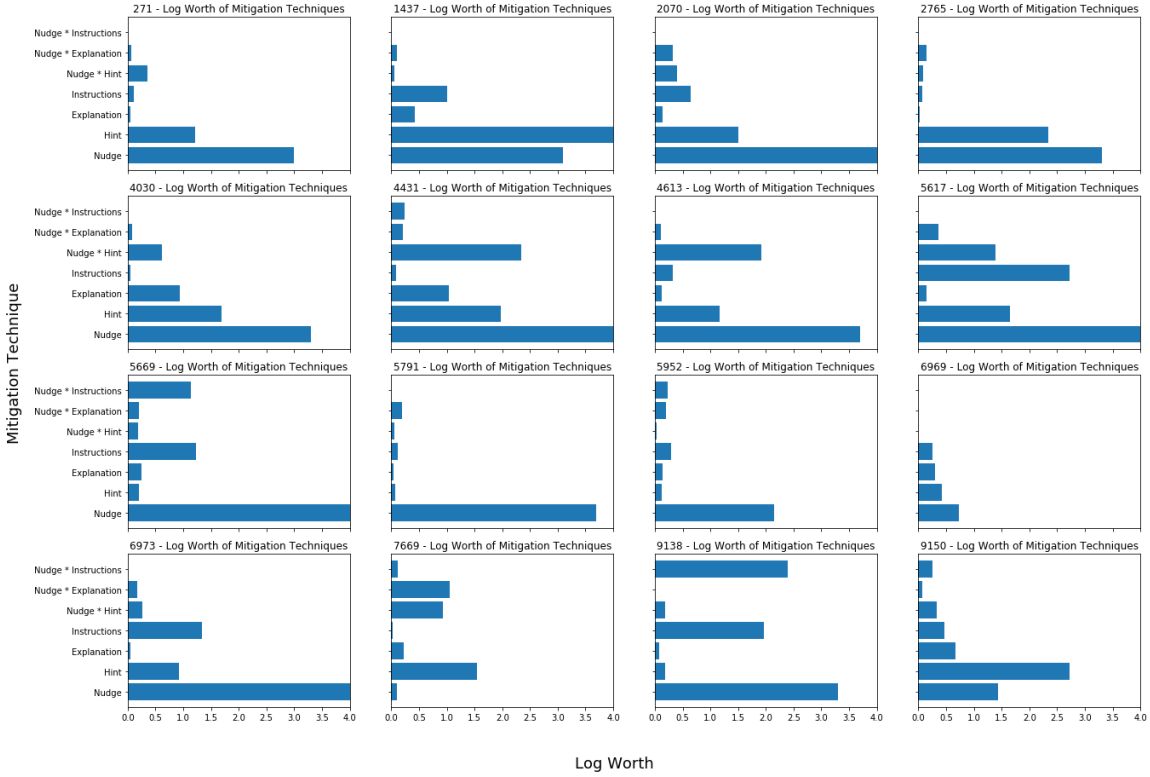

Figure 46. The log worth per effect per participant. The larger the blue bar, the more of an effect the mitigation technique had.

For all but three participants, the nudge had the highest effect on the number of efficient searches. For each of those three participants, the hint had the highest effect with the nudge coming in second.

Overall, the mitigation techniques of the nudge and the hint had the most effect on the number of efficient searches. The nudge had the most effect with a log worth of 10.664 and the hint had the second highest with a log worth of 8.493.

The final VSPs used during the last seven blocks are presented in Figure 47, and can also be seen in Table 22 and Table 23.

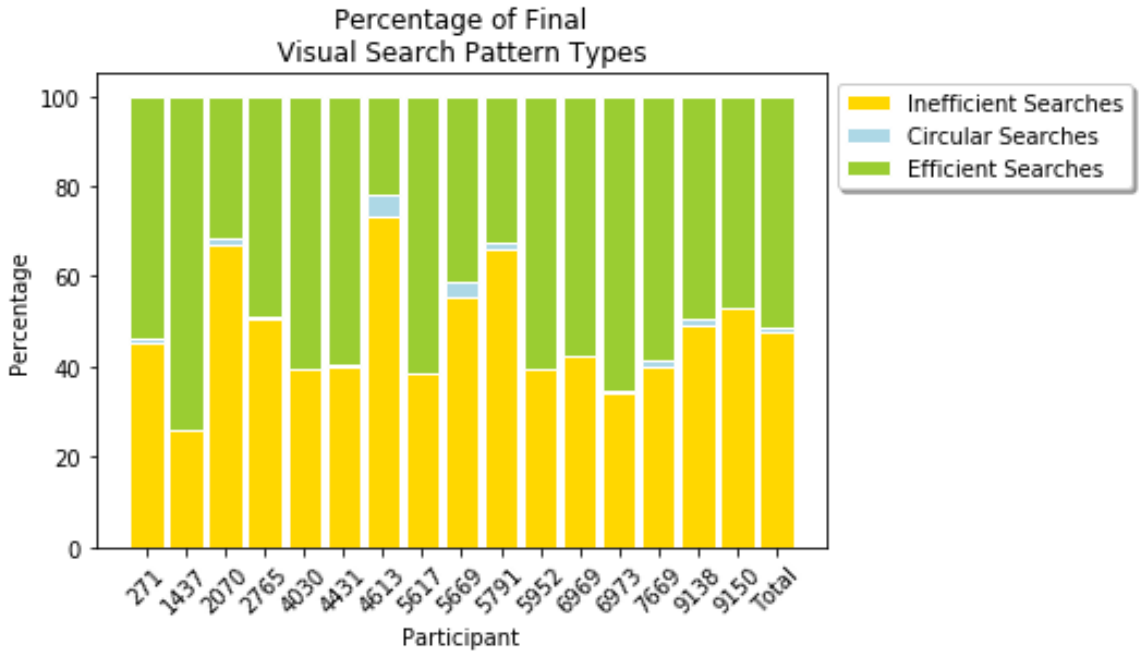

Figure 47. The VSPs per participant during the final seven blocks.

Table 22. The final VSPs for each participant

Participant	Confirmatory Searches	Non-Confirmatory Searches	Efficient Searches	Inefficient Searches	Circular Searches
271	30	110	76	64	1
1437	38	102	104	36	0
2070	30	110	45	95	2
2765	33	107	69	71	1
4030	34	106	85	55	0
4431	42	98	84	56	1
4613	29	111	32	108	7
5617	30	110	86	54	0
5669	49	91	60	80	5
5791	44	96	46	94	2
5952	32	108	85	55	0
6969	39	101	81	59	0
6973	33	107	92	48	1
7669	33	107	83	57	2
9138	56	84	70	70	2
9150	33	107	66	74	0
Total	585	1655	1164	1076	24

Table 23. The final VSPs (percentages for each participant)

Participant	Percent Inefficient	Percent Efficient	Percent Circular
271	45.39%	53.90%	0.71%
1437	25.71%	74.29%	0.00%
2070	66.90%	31.69%	1.41%
2765	50.35%	48.94%	0.71%
4030	39.29%	60.71%	0.00%
4431	39.72%	59.57%	0.71%
4613	73.47%	21.77%	4.76%
5617	38.57%	61.43%	0.00%
5669	55.17%	41.38%	3.45%
5791	66.20%	32.39%	1.41%
5952	39.29%	60.71%	0.00%
6969	42.14%	57.86%	0.00%
6973	34.04%	65.25%	0.71%
7669	40.14%	58.45%	1.41%
9138	49.30%	49.30%	1.41%
9150	52.86%	47.14%	0.00%
Total	47.53%	51.41%	1.06%

Overall, 51.4% of searches performed in the last seven blocks were efficient, 47.5% were inefficient, and just 1.1% were circular. $\frac{9}{16}$, or 56.3%, while 616, or 37.5%, of participants used a primarily efficient search. One participant, 9138, used an equal number of efficient and inefficient VSPs. These numbers constitute an increase of 32.3% of efficient searches, a decrease of 26.2% of inefficient searches, and a 6.1% decrease in circular searches.

4.3 Electroencephalography Analysis and Results

4.3.1 Machine Learning

The early stages of data exploration revealed a pattern that was not ideal for high performance using machine learning methods. This pattern was that the dataset generally had more inefficient searches than it did efficient searches. The most drastic

imbalance was 76.4% for participant 4613 and the least was 39.1% for participant 5952. Overall, the total balance was 42.0% inefficient and 58.0% efficient searches. The balance per participant can be seen graphically in Figure 48, and the total search imbalance can be seen in Figure 49. More data about the search imbalance can be seen in Table 24. Because of this imbalance, the machine learning results for this dataset are expected to be highly dependent on the individual participant and their specific data.

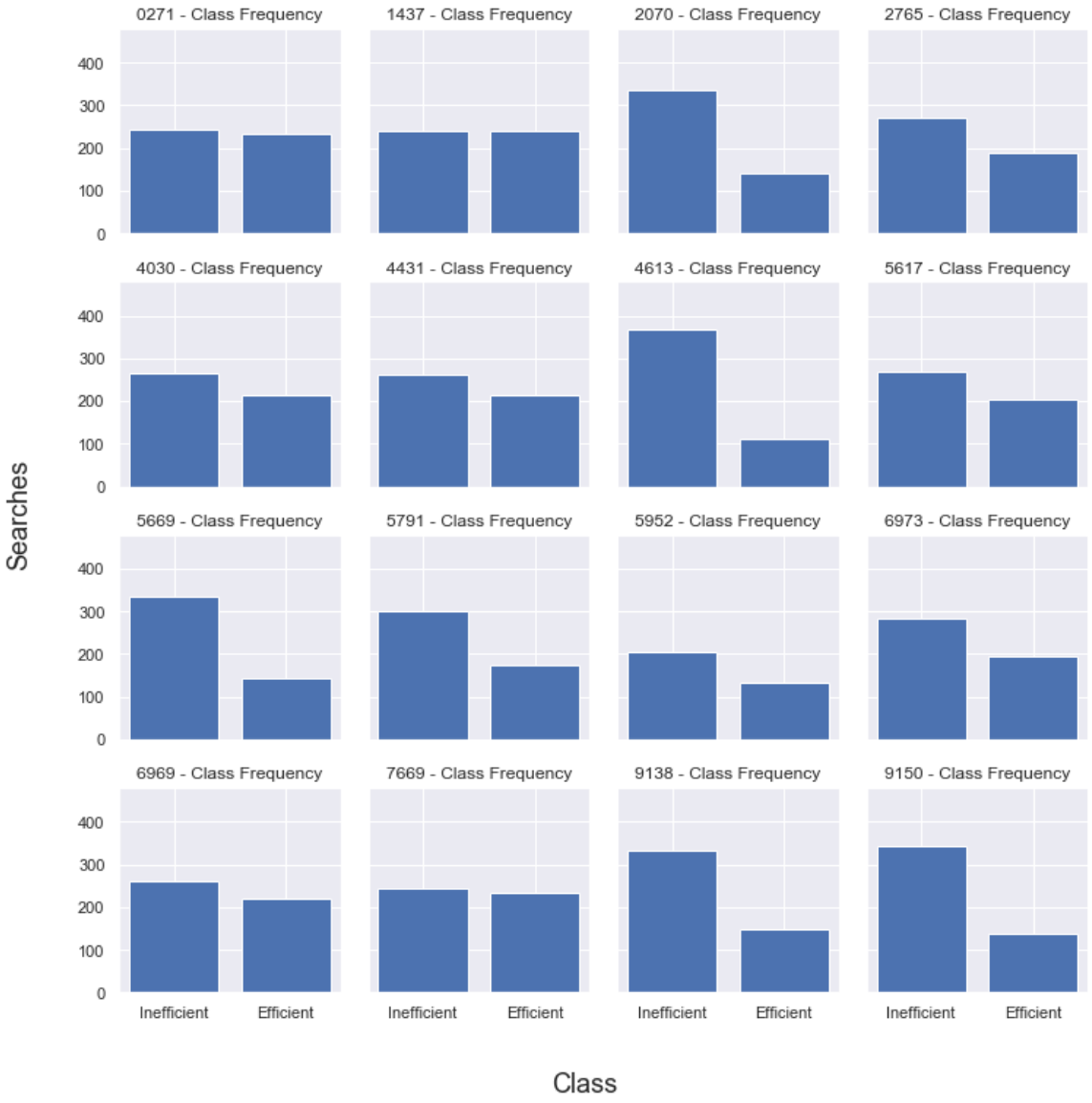

Figure 48. The class balance per participant.

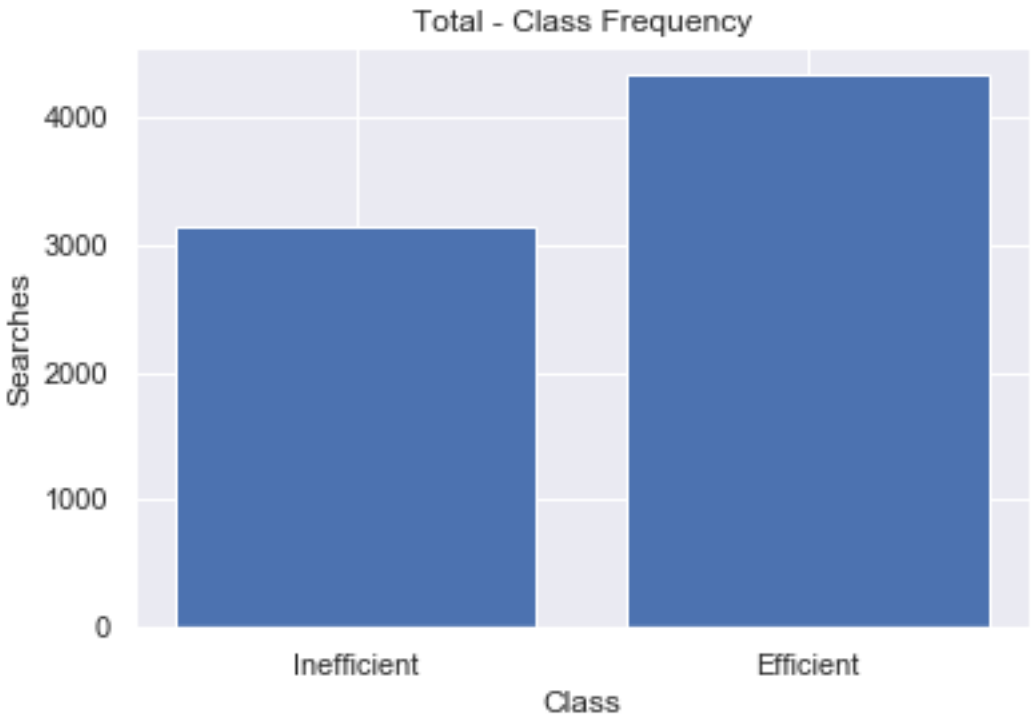

Figure 49. The class balance for the entire dataset.

Table 24. Dataset class distribution

Participant	Inefficient Searches	Efficient Searches	Percent Inefficient Searches	Percent Efficient Searches
0271	233	242	49.05%	50.95%
1437	239	241	49.79%	50.21%
2070	337	141	70.50%	29.50%
2765	269	189	58.73%	41.27%
4030	266	213	55.53%	44.47%
4431	262	215	54.93%	45.07%
4613	367	113	76.46%	23.54%
5617	268	205	56.66%	43.34%
5669	336	144	70.00%	30.00%
5791	299	175	63.08%	36.92%
5952	132	205	39.17%	60.83%
6973	196	282	41.00%	59.00%
6969	220	260	45.83%	54.17%
7669	245	233	51.26%	48.74%
9138	332	148	69.17%	30.83%
9150	342	138	71.25%	28.75%
Total	3144	4343	41.99%	58.01%

4.3.1.1 Frequency Features per Search Dataset

The mean frequency power features is the same dataset referred to in Section 3.5.2. This dataset's 320 features are the mean power of the five frequency bands at each electrode in the 64 electrode EEG cap. The extraction method for these features is described in Section 3.5.1.2 and was completed on epoched data that consisted of the participants' EEG data from 2 seconds before they indicated which color the target letter's circle was.

Due to issues with the Cognionics recording software, certain trials were dropped from various participants. Therefore, while the total number of observations for each participant should have been 480, some participants did not have the full 480 observations. The total number of recorded trials per participant is shown in Table 25.

Table 25. Number of trials recorded per participant

Participant	Total Number of Trials Recorded
0271	475
1437	480
2070	478
2765	458
4030	479
4431	477
4613	480
5617	473
5669	480
5791	474
5952	337
6973	478
6969	480
7669	478
9138	480
9150	480

In hyperparameter training, the Artificial Neural Network (ANN) was tuned with 100 epochs. The final training used both the training and validation sets and thus increased the amount of data by 20%. Thus, an increase in epochs of 20% to 120 epochs is also appropriate. The ANN trained for a maximum of 120 epochs with an early stopping callback stopping the training of the model as soon as the training accuracy hit 100%.

The overall accuracy and balanced accuracy of the Linear Discriminant Analysis (LDA), random forest classifier (RFC), and ANN models are shown in Figure 50 and Figure 51 respectively. The data is shown in Table 26. The average and standard deviations for each model's accuracy and balanced accuracy is shown in Table 27.

When examining the accuracy, LDA models performed the best, followed by ANN models. RFC models performed the worst, on average. For balanced accuracy, LDA

models performed the best, followed by RFC models, and last are the ANN models. When examining the balanced accuracy, the LDA models performed the best on $\frac{10}{16}$ models, or 62.5%, of models, while the RFC models performed the best on $\frac{5}{16}$ models, or 31.3%. An ANN model performed the best on $\frac{1}{16}$ models, or 6.3%.

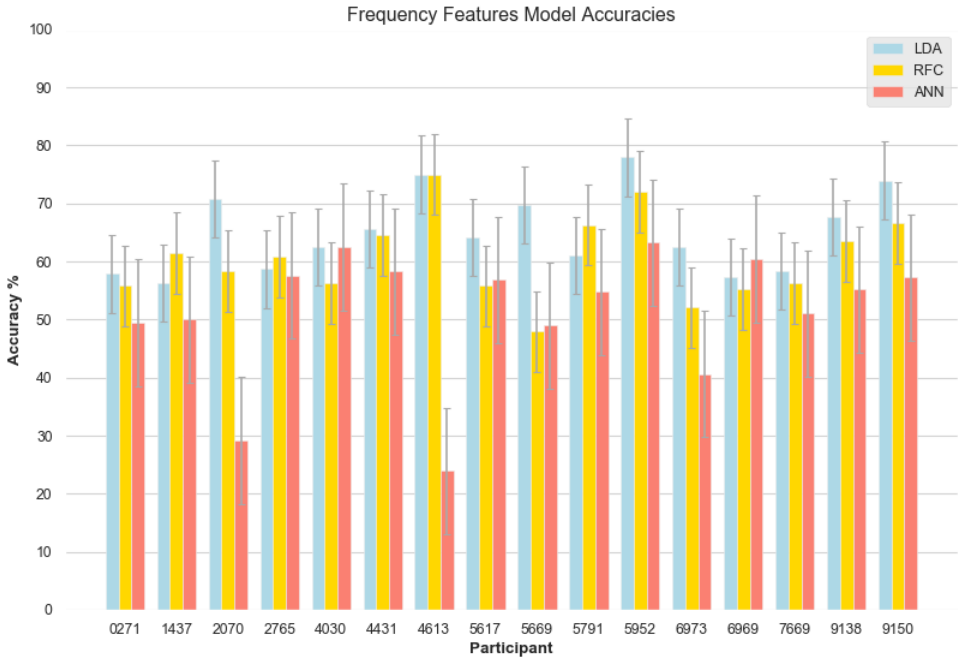

Figure 50. Overall model accuracy of the Frequency Features dataset.

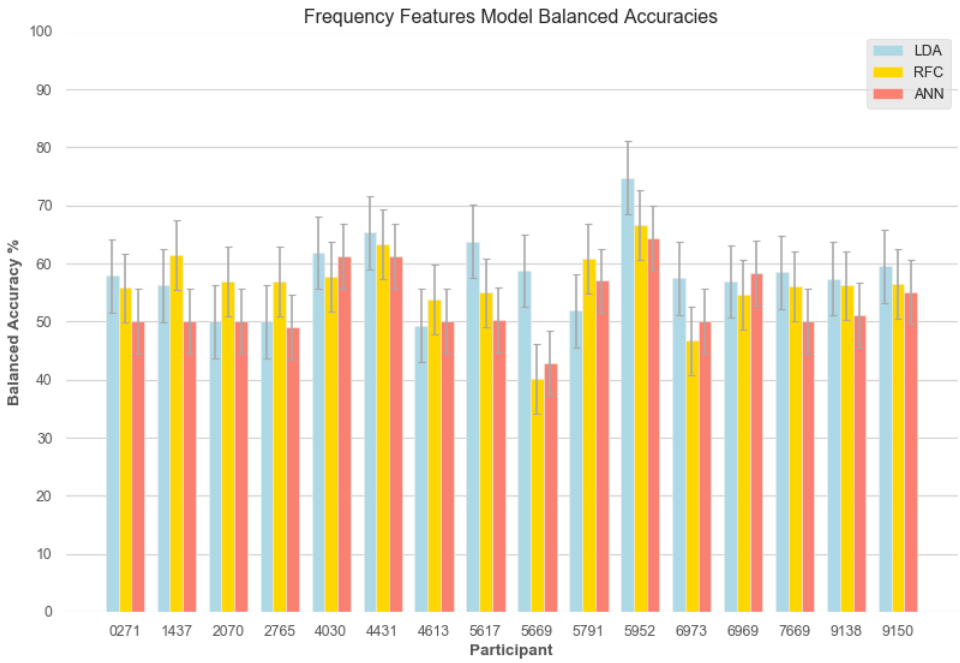

Figure 51. Overall model balanced accuracy of the Frequency Features dataset.

Table 26. Per participant model results for the frequency features dataset

Participant	LDA Accuracy	LDA Balanced Accuracy	RFC Accuracy	RFC Balanced Accuracy	ANN Accuracy	ANN Balanced Accuracy
271	57.89%	57.85%	55.79%	55.78%	49.47%	50.00%
1437	56.25%	56.25%	61.46%	61.46%	50.00%	50.00%
2070	70.83%	50.00%	58.33%	56.93%	29.17%	50.00%
2765	58.70%	50.00%	60.87%	56.92%	57.61%	49.07%
4030	62.50%	61.87%	56.25%	57.74%	62.50%	61.21%
4431	65.62%	65.36%	64.58%	63.32%	58.33%	61.17%
4613	75.00%	49.32%	75.00%	53.78%	23.96%	50.00%
5617	64.21%	63.82%	55.79%	54.95%	56.84%	50.29%
5669	69.79%	58.80%	47.92%	40.20%	48.96%	42.90%
5791	61.05%	51.90%	66.32%	60.83%	54.74%	57.02%
5952	77.94%	74.75%	72.06%	66.71%	63.24%	64.45%
6973	62.50%	57.49%	52.08%	46.69%	40.62%	50.00%
6969	57.29%	56.91%	55.21%	54.63%	60.42%	58.39%
7669	58.33%	58.49%	56.25%	56.14%	51.04%	50.00%
9138	67.71%	57.42%	63.54%	56.21%	55.21%	51.06%
9150	73.96%	59.56%	66.67%	56.51%	57.29%	55.15%

Table 27. Accuracy and Balanced accuracy by model for the frequency feature dataset

Model	Accuracy Average	Accuracy Std Dev	Balanced Accuracy Average	Balanced Accuracy Std Dev
LDA	58.1%	6.3%	58.1%	6.3%
RFC	56.2%	6.0%	56.2%	6.0%
ANN	57.3%	9.8%	53.2%	5.6%

Of the LDA models, 81.3% achieved greater than a 50% balanced accuracy. Participants 4431, 5617, and 5952 performed significantly greater than 50% by achieving a score greater than two standard deviations more than 50%. 87.5% of the RFC models achieved greater than a 50% balanced accuracy, while participants 4431 and 5952 performed significantly greater than 50%. Only 37.5% of ANN models achieved greater than a 50% balanced accuracy, while participants 4030, 4431, and 5952 performed significantly greater than 50%. The greatest overall balanced accuracy

was achieved on participant 5952, with balanced accuracy scores of 78.0%, 72.1%, and 64.5% with the LDA, RFC, and ANN models respectively. The lowest overall balanced accuracy was achieved on participant 5669, with balanced accuracy scores of 69.8%, 78.0%, and 42.9% with the LDA, RFC, and ANN models respectively.

The average Area Under the Receiver Operating Characteristic Curve (AUROC)s across all participants for LDA, RFC, and ANN models were 0.581 (std=0.063), 0.562 (std=.060), and 0.573 (std=0.098) respectively. The AUROC for each model is shown in Figure 52 and Table 28.

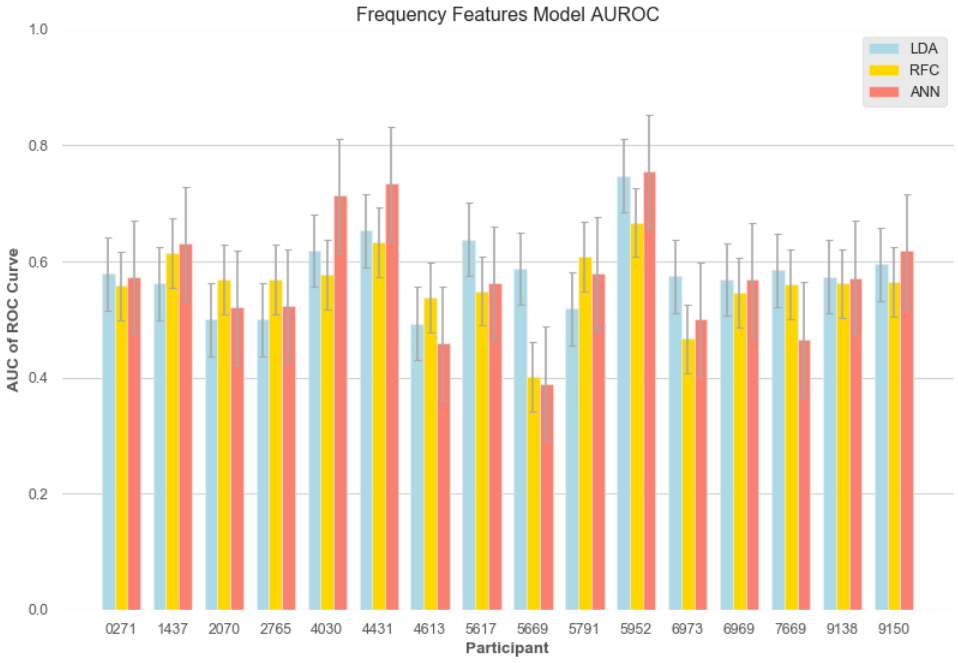

Figure 52. Overall model AUROC of the frequency feature dataset.

Table 28. Overall model AUROC of the frequency feature dataset

Participant	LDA AUC	RFC AUC	ANN AUC
271	0.578457	0.557846	0.573582
1437	0.562500	0.614583	0.630208
2070	0.500000	0.569328	0.520746
2765	0.500000	0.569201	0.523635
4030	0.618692	0.577446	0.713471
4431	0.653576	0.633172	0.734971
4613	0.493151	0.537820	0.458904
5617	0.638211	0.549458	0.562782
5669	0.588008	0.401956	0.389604
5791	0.519048	0.608333	0.579048
5952	0.747516	0.667118	0.755194
6973	0.574899	0.466937	0.500000
6969	0.569056	0.546329	0.568619
7669	0.584889	0.561442	0.466348
9138	0.574242	0.562121	0.572222
9150	0.595588	0.565126	0.618172

Observing the confusion matrices (Figure 53) for the top performing and worst performing models lends insight into a possible reasons as to why the models achieved their performance. The class imbalance for participant 5669 was 70.0% inefficient while the same imbalance for 5952 was 39.2%. Examining the confusion matrix for 5669 indicates that the model learned the opposite relation - if the participant performed an efficient search then the model tended to classify it as an inefficient search. This could perhaps be due to the prevalence of the inefficient class represented in the data. This is contrasted to participant 5952's confusion matrix - because the classes were relatively more balanced, the model had more chances to learn the true relationship. Thus the model's performance on participant 5952 was higher.

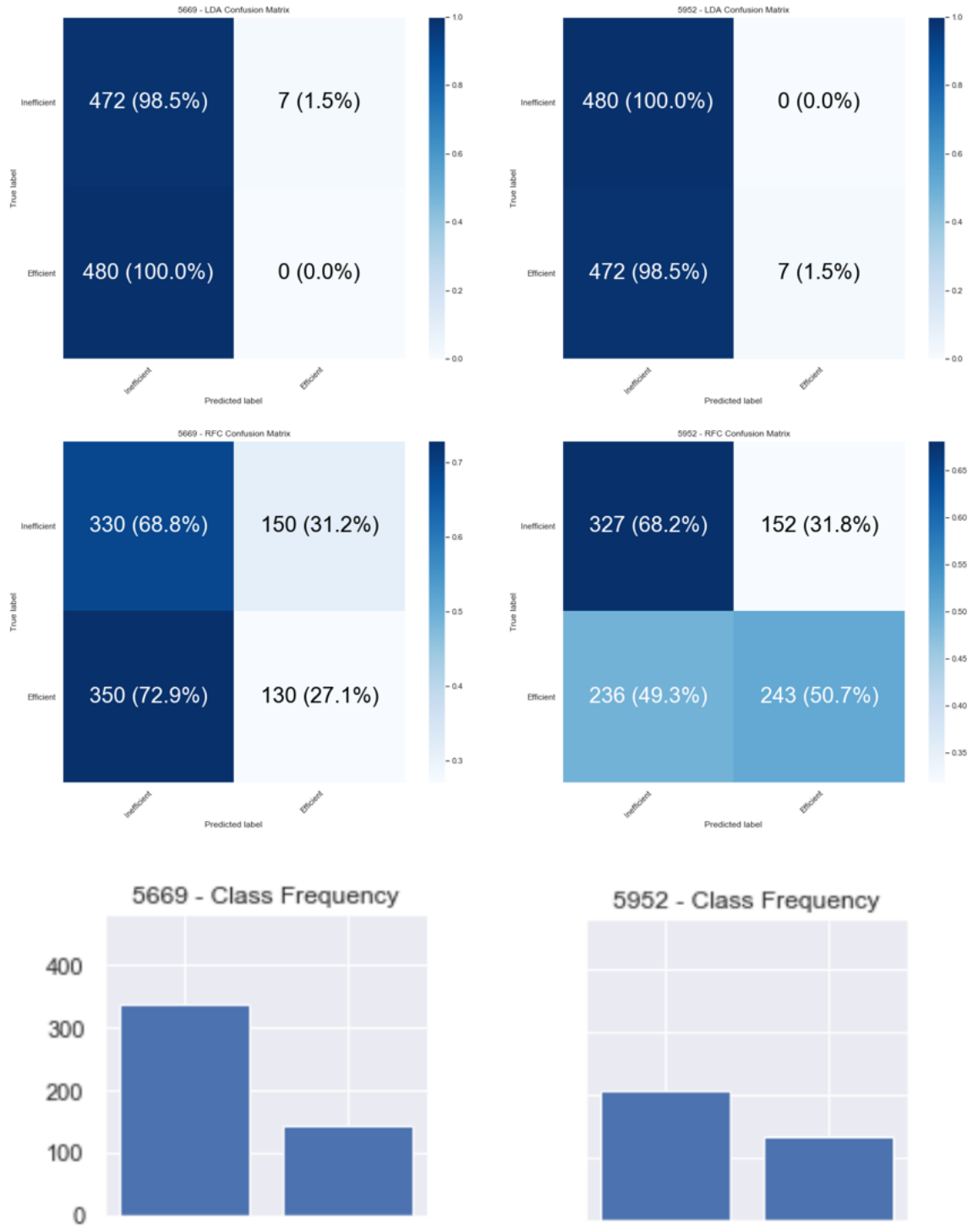

Figure 53. The top performer vs the lowest performer on the frequency feature dataset.

The observations from this dataset were obtained by observing the entire two seconds prior to the participant's decision as to what color the target letter's circle was. The critical assumption made was that there would be a consistent difference between brain activity of a person conducting an inefficient vs an efficient visual search during this two-second period. A possible explanation of inefficient searches, and thus confirmative searches, stems from the neuroscientific perspective. The neuroscientific perspective relates various cognitive biases as being characteristic of biological neural networks. Thus, cognitive biases could be a result of the neural characteristics of the brain [103]. Cognitive biases might occur in the same neural networks as motor functions and thus there would be no distinguishable brain activity that relates to an inefficient search. While high-performing results from this dataset indicate that this perspective might not reflect the full truth, accepting this perspective would account for the overall low performance of the machine learning models.

4.3.1.2 Time Series Features

The Time Series per Search dataset is the same dataset referred to in Section 3.5.2. This dataset consists of 2-second time series signals consisting of 500 frames at 250 Hz. These signals are the 2 seconds prior to a participant pressing either the "c" or "z" key indicating that they have decided which color the target letter's circle is, The 64 features present in this dataset correspond to the 64 EEG electrodes in the 10-20 International Standard electrode placement as seen in Figure 24. The labels for this dataset are either "efficient" or "inefficient" and represent the VSP that the participant used for the 2-second window. As previously noted, while there should be 480 observations for each participant, errors with the Cognionics recording software caused some trials to be dropped. For information on the amount of observations per participant, please see Table 25.

The Long Short-Term Memory (LSTM) and Temporal Convolutional Network (TCN) models were both trained with a maximum of 60 epochs each, however a callback was implemented that halted the training should the training accuracy reach 100%. On both models, for all participants except 2765, the training accuracy was able to reach 100%.

The overall accuracy and balanced accuracy of the LSTM and TCN models are shown in Figure 54 and Figure 55 respectively. The data is shown in Table 29. The average and standard deviations for each model's accuracy and balanced accuracy is shown in Table 30.

When examining both the accuracy and the balanced accuracy, LSTM models performed the best compared to TCN models. When examining the balanced accuracy, LSTM models performed the best on $\frac{10}{16}$ models, or 62.5%.

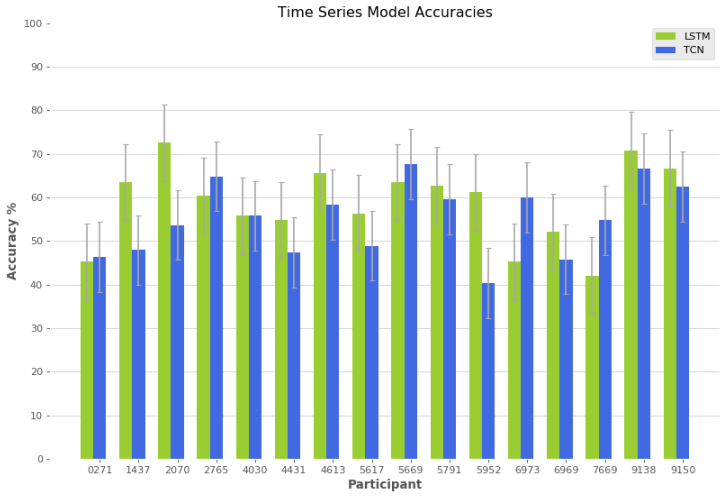

Figure 54. Overall model accuracy of the Time Series dataset.

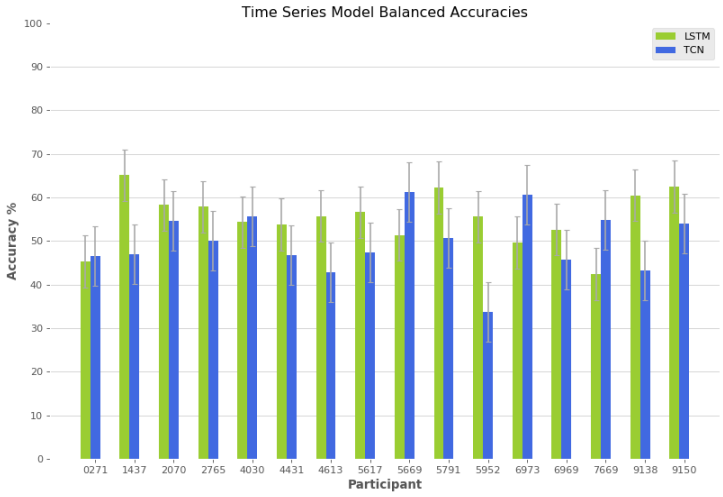

Figure 55. Overall model balanced accuracy of the Time Series dataset.

Table 29. Per participant model results for the time series dataset

Participant	LSTM Accuracy	LSTM Balanced Accuracy	TCN Accuracy	TCN Balanced Accuracy
271	45.26%	45.33%	46.32%	46.50%
1437	63.54%	65.12%	47.92%	47.06%
2070	72.63%	58.29%	53.68%	54.69%
2765	60.44%	57.86%	64.84%	50.05%
4030	55.79%	54.36%	55.79%	55.70%
4431	54.74%	53.81%	47.37%	46.70%
4613	65.62%	55.71%	58.33%	42.82%
5617	56.38%	56.62%	48.94%	47.46%
5669	63.54%	51.39%	67.71%	61.18%
5791	62.77%	32.27%	59.57%	50.77%
5952	61.19	55.59%	40.30%	33.76%
6973	42.56	49.68%	60.00%	60.68%
6969	52.08	52.61%	45.83%	45.83%
7669	42.11	42.44%	54.74%	54.79%
9138	70.83	60.52%	66.67%	43.24%
9150	66.67	62.50%	62.50%	53.92%

Table 30. Accuracy and Balanced accuracy by model for the time series dataset

Model	Accuracy Average	Average Std Dev	Balanced Accuracy Average	Balanced Accuracy Std Dev
LSTM	58.7%	8.8%	55.3%	5.9%
TCN	55.0%	8.0%	49.7%	6.8%

Of the LDA models, $\frac{13}{16}$, or 81.3%, achieved a greater than 50% balanced accuracy score. Three models, for participants 1437, 5791, and 9150, performed significantly better than 50%. Of the TCN models, $\frac{7}{16}$, or 43.8%, achieved greater than 50% balanced accuracy. No models performed significantly better than 50%. The greatest balanced accuracy for the LSTM models was participant 1437 with a 65.1%, while the lowest was 7669 with a 42.4%. The greatest balanced accuracy for the TCN models was participant 5669 with a 61.2%, while the lowest was 7669 with a 42.8%.

The average AUROC across all participants for LSTM and TCN models was 0.541 (std=0.049) and 0.497 (std=0.068) respectively. The AUROC for each model is shown in Figure 56 and Table 31.

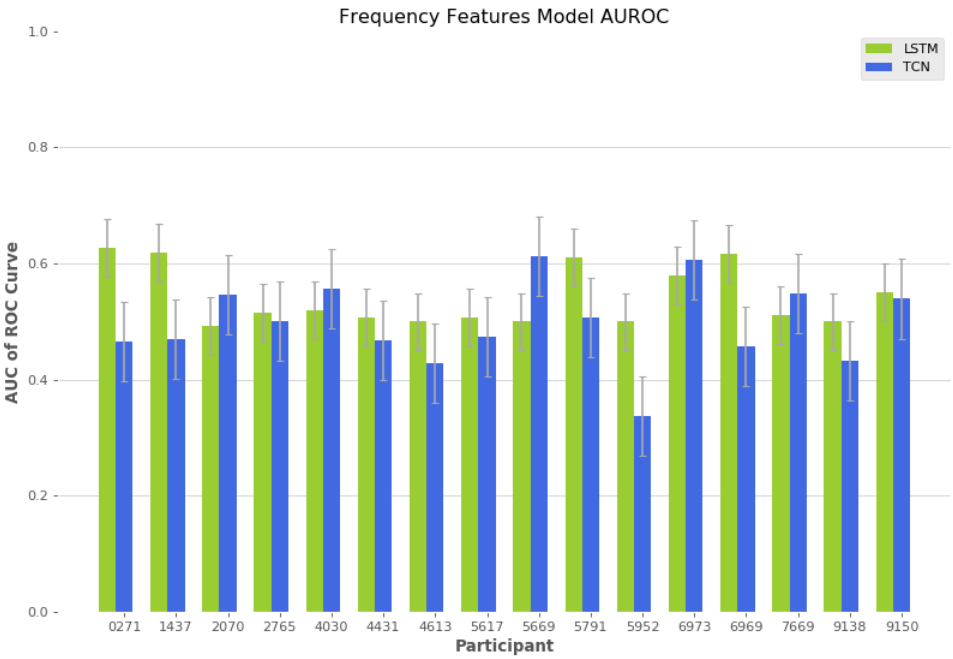

Figure 56. Overall model AUROC of the Time Series dataset.

Table 31. Overall model AUROC of the time series dataset

Participant	LSTM AUC	TCN AUC
271	0.626909254	0.464951198
1437	0.619047619	0.470588235
2070	0.492957746	0.546908316
2765	0.515625	0.500546448
4030	0.519318182	0.55695922
4431	0.506666667	0.467045455
4613	0.5	0.42823109
5617	0.506372549	0.474632527
5669	0.5	0.611842105
5791	0.611111111	0.50769995
5952	0.5	0.337619048
6973	0.579545455	0.606818182
6969	0.616851441	0.458333333
7669	0.511303191	0.547914818
9138	0.5	0.432432432
9150	0.550549451	0.53915493

The confusion matrices for the best and worst performing participants for both models can be seen in Figure 57. Contrary to the frequency models, the worst performing participant in both the LSTM and TCN models was participant 7669, which had a very balanced dataset with only 51.26% of searches being inefficient.

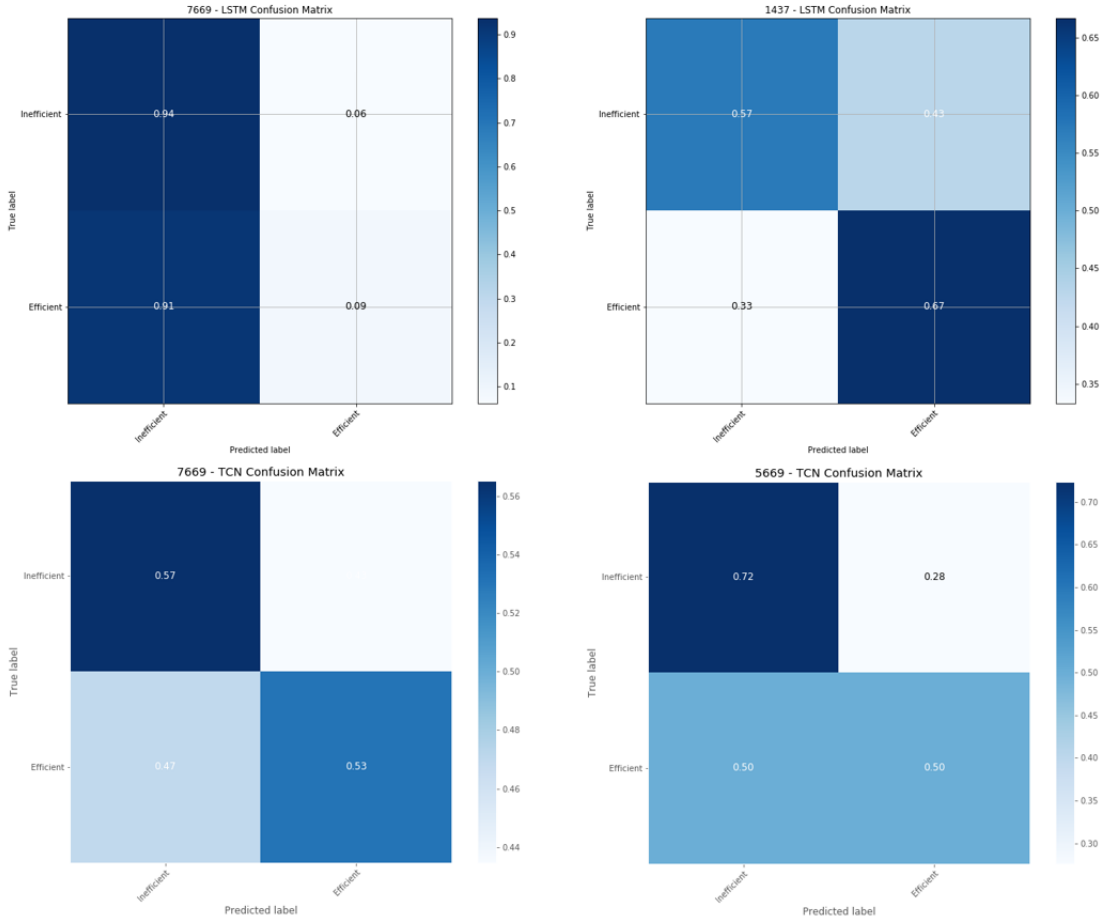

Figure 57. The top row of images are confusion matrices from the LSTM model while the bottom row are from the TCN model. The left column is the lowest performers while the right column is the highest performers.

The poor AUROC and model performance across the participants indicates that either the time series of the EEG data or the method of segmenting the data with a label of an inefficient or an efficient search is not an appropriate method for detecting efficient searches.

4.3.1.3 Feature Importance

It's necessary to track performance metrics for a machine learning model to determine how well the model is learning relationships between the features and the target variables, but determining feature importance can provide greater insight into the relationships present that are being modeled. A useful tool for identifying feature importance is a random forest (RF) classifier. RF classifiers excel at identifying important features because they compute feature importance as part of their model fitting process.

Table 32 shows the top ten features based on the number of times that the frequency feature appears in a participant's 50 most important features. These features are represented visually in Figure 58.

Table 32. Salient Features across all Participants

Feature Location/ Frequency	**Count**
FC2/Beta	9
T7/Gamma	7
O1/Beta	5
TP8/Beta	5
Cz/Delta	5
Fp1/Theta	5
CPz/Theta	4
TP8/Gamma	4
P8/Beta	4
TP7/Beta	4

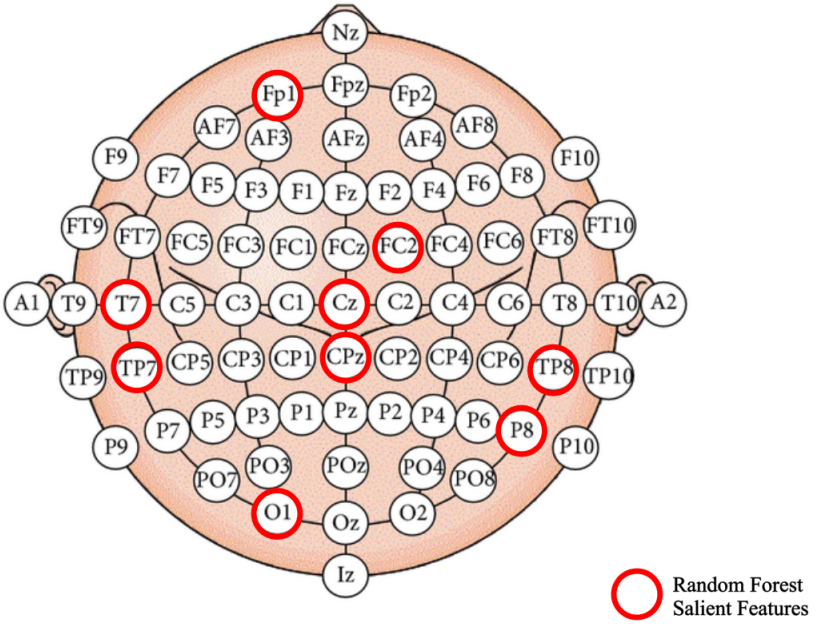

Figure 58. EEG Electrode Locations and Salient Features

To identify the most important features in an efficient VSP, features that were common among participants that had an RF area under the curve (AUC) greater than 0.50 were analyzed. These high-performing features can be seen in Table 33 and in Figure 59.

Table 33. Salient Features in Top performing Participant Models

Feature Location/ Frequency	Count
FC2/Beta	7
T7/Gamma	5
O1/Beta	4
Cz/Delta	4
CPz/Theta	4
Fp1/Theta	4
TP7/Beta	4
P2/Delta	4
TP8/Gamma	3
TP8/Beta	3

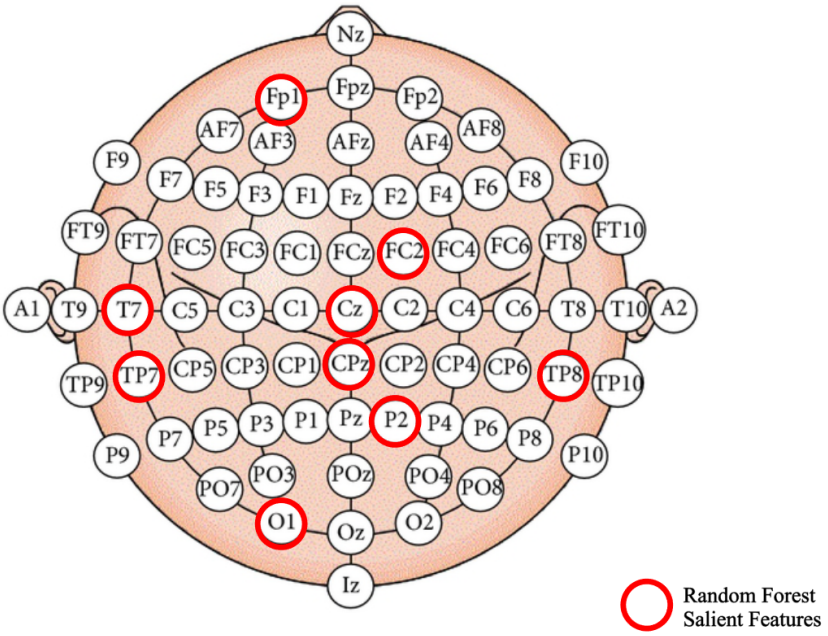

Figure 59. EEG Electrode Locations and Salient Features with High Performance

The salient features were not consistent across participants. This could be because the machine learning models were unable to associate brain activity with efficient

searching, or because there is no specific brain activity associated with efficient searching.

4.3.2 Cross-Participant

The results that have been discussed so far are within-participant results. Because the frequency features dataset, specifically the LDA and RFC, had the best performance out of all of the models, cross-participant models were analyzed as described in Section 3.5.4. The dataset had 7,487 observations which spanned across all 16 participants. Of these 7,487 observations, 3,144, or 42.0%, belong to the "inefficient" class. A train, validation, and test approach was applied in which models were trained on 13 participants, validated on 2 random participants, and then ultimately tested on a single participant. This process was repeated so that every participant's data was tested by itself. The following results are performance metrics on the test dataset.

The balanced accuracy for the cross-participant test performance is displayed in Figure 60 and in Table 34. The participant number along the bottom x-axis is the participant that was used as the test set. A balanced accuracy score above 50% was achieved on $\frac{14}{16}$ participants by at least one model. The mean test balanced accuracy for the LDA, RFC, and ANN models was 50.97%($\pm$1.52%), 54.01%($\pm$3.74%), and 50.05%($\pm$0.14%). Although above 50% accuracy was achieved on 15 participants, the mean test balanced accuracy of 51.68%($\pm$2.88%) was not significantly greater than 50%. The highest balanced accuracy was 59.46% and was obtained by the RFC model on participant 5952.

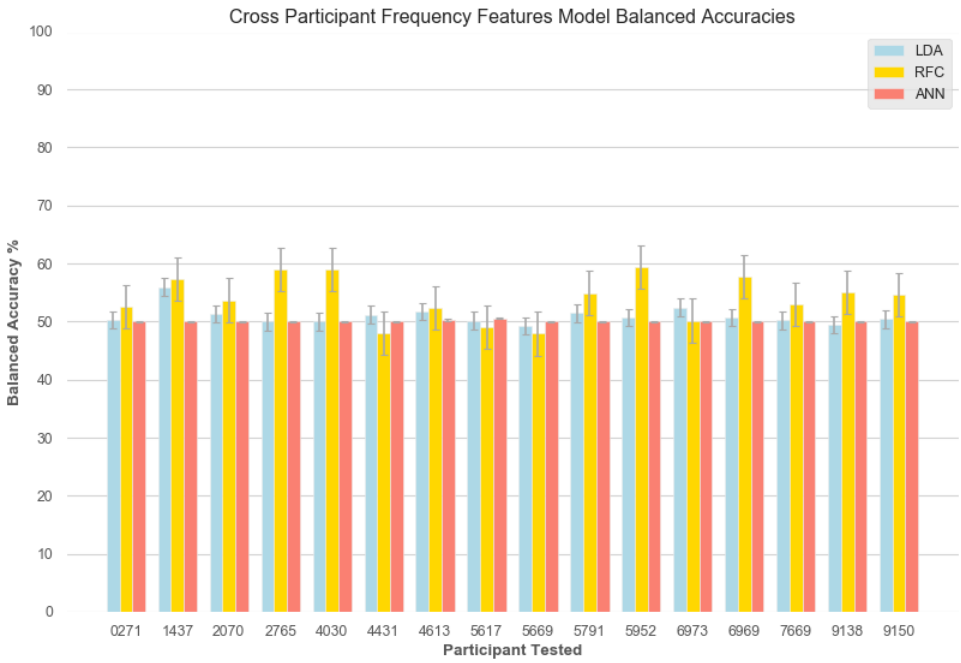

Figure 60. Cross-Participant Frequency Feature Balanced Accuracies

Table 34. Cross-Participant Frequency Features Model AUROC

Participant	LDA Balanced Accuracy	RFC Balanced Accuracy	ANN Balanced Accuracy
271	50.30%	52.51%	50.00%
1437	55.96%	57.41%	50.00%
2070	51.35%	53.70%	50.00%
2765	50.00%	58.90%	50.00%
4030	50.05%	59.05%	50.00%
4431	51.18%	48.02%	50.00%
4613	51.73%	52.40%	50.27%
5617	50.17%	49.12%	50.55%
5669	49.26%	47.92%	50.00%
5791	51.45%	54.94%	50.00%
5952	50.73%	59.46%	50.00%
6973	52.45%	50.18%	50.00%
6969	50.72%	57.71%	50.03%
7669	50.25%	53.03%	50.00%
9138	49.47%	55.10%	50.00%
9150	50.42%	54.67%	50.00%

The cross-participant's model performance was compared with the within-participant model performance by comparing their AUROCs. The mean cross-participant AUROC was 0.510, 0.540, and 0.524 for the LDA, RFC, and ANN models respectively. The cross-participant model AUROC was slightly smaller than the within-participant AUROC which was 0.518, 0.562, and 0.573 for the LDA, RFC, and ANN models respectively. The cross-participant AUROC is displayed in Figure 61 and in Table 35.

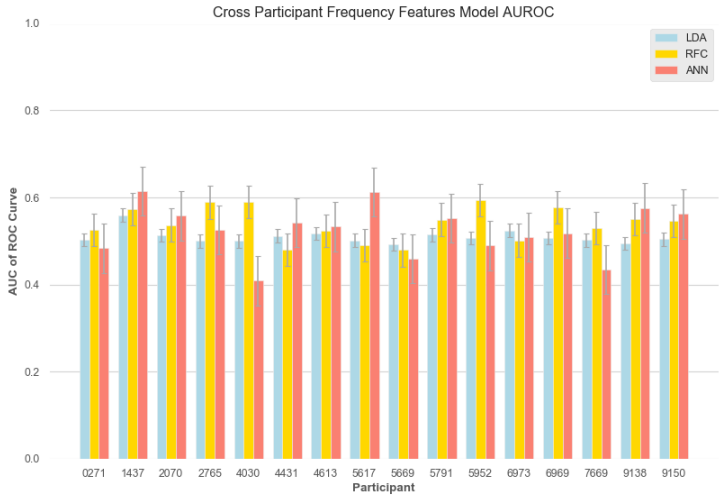

Figure 61. Cross-Participant Frequency Feature AUROCs

Table 35. Cross-Participant Frequency Features Model AUROC

Participant	LDA AUC	RFC AUC	ANN AUC
271	0.503015	0.525130	0.483551
1437	0.559576	0.574090	0.614655
2070	0.513458	0.536966	0.558263
2765	0.500000	0.588954	0.525255
4030	0.500468	0.590473	0.409139
4431	0.511805	0.480224	0.542358
4613	0.517301	0.523981	0.532927
5617	0.501720	0.491172	0.612805
5669	0.492560	0.479167	0.459408
5791	0.514496	0.549441	0.552690
5952	0.507317	0.594568	0.489763
6973	0.524479	0.501809	0.508711
6969	0.507168	0.577098	0.518164
7669	0.502461	0.530349	0.434860
9138	0.494709	0.550961	0.575810
9150	0.504195	0.546720	0.562177

The highest AUROC achieved was 0.613 and was achieved by the ANN model on participant 5617. This AUROC indicates that the models can perform well on the dataset as a whole, but overall there is also a lack of consistent results. This lack of consistent results could be due to many various factors. There was small amount of data per participant, however, an additional factor is that some participants might have very noisy data. As seen in Figure 52, participant 5669 has an average AUROC of 0.477. When all of the participants' data is combined to generate cross-participant

models, noisy data such as this can reduce overall performance.

4.4 Error Analysis

One hypothesized source of error stems from how efficient and inefficient VSPs are conducted within the brain. As seen from the class balance (Figure 48) and the post-experiment questionnaires, some participants naturally used an efficient VSP while some naturally resorted to an inefficient VSP. The fact that some participants naturally used an efficient VSP and some used an inefficient VSP could lend credence to the theory of neuroscientific perspective. This theory states that cognitive biases might occur in the same neural networks as motor functions and thus there would be no distinguishable brain activity that relates to an inefficient search.

Another source of error could stem from the epoching process - the decision to conduct an inefficient versus an efficient search could be a split-second decision that only occurs once, not an ongoing decision that lasts the duration of the two-seconds before the participant selects an answer. Additionally, once a participant decides to search inefficiently or efficiently, they could not consider the question of how to search again. This theory could also explain why the time-series models performed worse when compared to the frequency-feature models.

4.5 Summary

Behavioral results from the Efficient Search Experiment (ESE) determined participants' initial Visual Search Pattern (VSP)s as well as the effects of the mitigation techniques on these VSPs. It also answered questions as to whether efficient searches were more accurate and faster when compared to inefficient searching.

First, participants had an overall accuracy of 95.03% when selecting which color the target letter's circle was. A two-sample paired t-test was performed to see if there

was a difference in the accuracies of inefficient and efficient searches. When searching efficiently, there was a statistically significant increase in accuracy ($t(15)=5.59$, $p=0.00005$) of 2.41%.

Next, the average search time of participants was 2.17 seconds. A two-sample paired t-test was performed to see if there was a difference in the average search times of inefficient and efficient searches. When searching efficiently, there was a statistically significant decrease in search times ($t(15)=5.53$, $p=0.00005$) of 0.30 seconds.

Initially, participants overwhelmingly performed inefficient searches. In the first eight blocks, 73.68% of searches were inefficient, 19.14% were efficient, and 7.18% were circular. At the end of the experiment, participants performed more efficient searches than inefficient searches. In the last seven blocks, 47.53% of searches were inefficient, 51.41% were efficient, and 1.06% were circular.

With a log worth of 10.664, the mitigation technique of the nudge was the most effective in increasing the number of efficient searches. In effectiveness, the nudge was closely followed by the hint with a log worth of 8.493.

The results from the behavioral analysis served to answer Research Question 1 and Research Question 3:

Research Question 1

What visual search patterns do participants naturally use during a visual search task?

Results: 73.68% *of participants initially used an inefficient VSP, while* 19.14% *used an efficient search and* 7.18% *used a circular search.*

Research Question 3

For a participant who is performing an inefficient search, can mitigation techniques change the participant's search patterns to an efficient search pattern that will persist for the remainder of the search tasks?

Results: In the last seven blocks, efficient searches were increased by 32.27% *to* 51.41%*, inefficient searches were decreased by* 26.15% *to* 47.53%*, and circular searches were decreased by* 6.12% *to* 1.06%.

Results from the Electroencephalography (EEG) analysis determined whether a difference could be identified between inefficient and efficient searching. Two datasets were considered: the first used features extracted from the average frequency values of the alpha, beta, delta, gamma, and theta frequency bands, while the second used the raw time series values of the electrodes. The results of the machine learning models can be seen in Table 36.

Table 36. Accuracy and Balanced accuracy by model for both the frequency-feature and time-series datasets

Model	Accuracy Average	Accuracy Std Dev	Balanced Accuracy Average	Balanced Accuracy Std Dev
LDA	58.1%	6.3%	58.1%	6.3%
RFC	56.2%	6.0%	56.2%	6.0%
ANN	57.3%	9.8%	53.2%	5.6%
LSTM	58.7%	8.8%	55.3%	5.9%
TCN	55.0%	8.0%	49.7%	6.8%

The results from the EEG analysis served to answer Research Question 2:

Research Question 2

Can physiological signals such as EEG, Electrooculography (EOG), and Electrocardiography (ECG) be associated with an efficient visual search?

Result: Four models were created that achieved an average balanced accuracy of greater than 50%*. A Linear Discriminant Analysis (LDA) (*58.1%*), a random forest classifier (RFC) (*56.2%*, an Artificial Neural Network (ANN) (*53.2%*), and a Long Short-Term Memory (LSTM) (*55.3%*) model.*

V. Conclusions and Recommendations

5.1 Conclusions of Research

This research was successful in its objective of detecting and mitigating an inefficient search during a visual search task. The initial Visual Search Pattern (VSP)s of participants were identified and the effect of various mitigation techniques on these VSPs was determined. A relationship between physiological signals and an inefficient VSP was also found. In order to achieve these goals, a human-participant experiment was designed and executed during which electrophysiological and behavioral data was recorded and analyzed.

The first research question (Section 3.3.1) investigated what VSP participants initially used. It was hypothesized that the majority ($> 50\%$) of participants would initially use an inefficient VSP. This hypothesis proved correct: initially, participants overwhelmingly performed inefficient searches. In the first eight blocks, 73.68% of searches were inefficient, 19.14% were efficient, and 7.18% were circular.

Research question two (Section 3.3.2) investigated whether physiological signals can be associated with an efficient visual search. To investigate this question, multiple machine learning models were investigated. However, only Electroencephalography (EEG) physiological signals were investigated to answer this research question. Machine learning models were able to obtain an average within-participant cross-validation balanced accuracy of 58.1%, 56.2%, 53.2%, 55.3%, and 49.7% with the Linear Discriminant Analysis (LDA), random forest classifier (RFC), Artificial Neural Network (ANN), Long Short-Term Memory (LSTM), and Temporal Convolutional Network (TCN) models respectively. Cross-participant machine learning was also explored on the frequency features dataset. The mean test balanced accuracy for the LDA, RFC, and ANN models was 50.97%($\pm$1.52%), 54.01%($\pm$3.74%), and 50.05%($\pm$0.14%). The

LDA models achieved greater than 50% balanced accuracy on $\frac{13}{16}$ participants, the RFC models achieved greater than 50% balanced accuracy on $\frac{14}{16}$ participants, and the ANN models achieved a greater than 50% balanced accuracy on $\frac{6}{16}$ participants. The highest balanced accuracies were 74.75%, 66.71%, and 64.45% for the LDA, RFC, and ANN models respectively. The LSTM models achieved greater than 50% balanced accuracy on $\frac{13}{16}$ participants and the TCN models achieved a greater than 50% balanced accuracy on $\frac{7}{16}$ participants. The highest balanced accuracies were 65.12% and 61.18% for the LSTM and TCN models respectively.

The third research question (Section 3.3.3) sought to determine whether mitigation techniques applied during a visual search could change a participant's inefficient VSP to an efficient VSP for the remainder of the experiment. With a log worth of 10.664, the mitigation technique of the nudge was the most effective in increasing the number of efficient searches. In effectiveness, the nudge was closely followed by the hint with a log worth of 8.493. In the last seven blocks, efficient searches were increased by 32.27% to 51.41%, inefficient searches were decreased by 26.15% to 47.53%, and circular searches were decreased by 6.12% to 1.06%.

Participants had an overall accuracy of 95.03% when selecting which color the target letter's circle was. A two-sample paired t-test was performed to see if there was a difference in the accuracies of inefficient and efficient searches. When searching efficiently, there was a statistically significant increase in accuracy ($t(15)=5.59$, $p=0.00005$) of 2.41%.

Lastly, the average search time of participants was 2.17 seconds. A two-sample paired t-test was performed to see if there was a difference in the average search times of inefficient and efficient searches. When searching efficiently, there was a statistically significant decrease in search times ($t(15)=5.53$, $p=0.00005$) of 0.30 seconds.

5.2 Significance of Research

Current research into inefficient searches during a visual search task analyze search patterns once the task is complete. These methods do not allow for a classification of an inefficient search in real-time. Furthermore, the majority of visual research has hypothesized that most humans will naturally use an inefficient search method. This research's findings reinforce that hypothesis. Military operators use visual searches every day in their job. This includes pilots scanning instrument gauges, intel analysts scanning satellite imagery, and doctors scanning patient x-rays. All of these military members can fall prey to an inefficient, and thus a biased, visual search. Therefore, whether through proper training and instruction, or through a mitigation system, a way to both detect and mitigate these inefficient searches would significantly help these operators to properly do their jobs. The results of this work add on to the knowledge repository for how humans perform a visual search and as such can be used to develop efficient search methods. These results of how various mitigation techniques affect a visual search are useful in determining the best method of mitigating an inefficient or biased search. Additionally, the results presented in this work advance the use of physiological signals to detect cognitive biases. While certain participants' models performed well, overall the models for each dataset only marginally perform better than chance. However, these results do suggest that it is possible to classify an efficient or inefficient search from EEG signals.

5.3 Recommendations for Future Research

5.3.1 Efficient Search Experiment (ESE) Changes

This work modified an existing visual search experiment to include a dynamic mitigation system based on each participant's own VSPs. As this was the first

iteration of this experiment, there is room for improvement if a future experiment were to be conducted. A piece of feedback given consistently by participants was that they were not aware that there would only ever be one instance of the target letter appearing in the stimuli per trial. The participants that informed the experiment administrators said that once they realized there was only ever one instance of the target letter that it then changed their search patterns. A recommendation for a modification to the ESE is to inform the participants during the training day that there is only ever one instance of the target letter present in any given trial.

Another feedback piece given by participants was that they felt as if they were "cheating" when using an efficient search pattern. A recommendation for a modification to the ESE is to inform the participants during the training day that they are welcome to search the stimuli in whichever manner they feel is the most accurate and efficient.

Instead of using gaze tracking, a future addition to the ESE could be the detection of inefficient searching through the use of EEG signals. During training, initial data could be gathered on a participant and fed into a machine learning model. During the execution of the actual experiment, the live signals from the EEG electrodes could be given to the model in real-time. The model would then output the likelihood that the participant is conducting an inefficient search and then apply the appropriate measures.

5.3.2 Participant Selection for Future Trials

As noted in the limitations (Section 1.6) the participants that completed the ESE were not diverse. For the results of this work to apply to the larger population, a more diverse group of participants is needed. A wider range of ages, backgrounds, education levels, diversity, and gender should be included. However, the ultimate goal of this work is to limit the effects of an inefficient search in military operators.

Thus, the demographics of the participants in this work more accurately reflect the target population than the population as a whole.

5.3.3 Machine Learning

Research Question Two investigated if phsyiological signals could be used to determine whether an inefficient search was occuring. However, this work investigated only EEG. Future work could investigate Electrooculography (EOG), Electrocardiography (ECG), and galvanic skin response (GSR). In addition to determining a link between confirmation bias and EEG, Minas et al. also determined a link between GSR and confirmation bias. Thus, investigating GSR could be beneficial.

Possible future work could include increasing the size of the datasets. Increasing the size of the datasets used for machine learning allows for more data to be used in the train, validation, and test datasets and allows the models more opportunities to learn the relationship.

The data used in this experiment depended wholly on the epoching of the dataset. In this experiment, the epochs were determined by the two seconds that occurred before the participants pushed the key to indicate their answer. Future work should include variations on epoching to determine the best epoch central point and time window.

This research used the mean power spectral density and the raw time series signals of the EEG signals as features. In a recent study on cognitive workload estimation using EEG, results indicated that the variance of power spectral density was an important feature [104]. Using the variance of power spectral density in addition to the mean power spectral density could improve machine learning results.

5.4 Summary

This work explored the detection and mitigation of inefficient searches, or a confirmation bias, during a visual search task. Using behavioral and electrophysiological signals, an inefficient search during a visual search was detected. This work determined that the majority of participants naturally employed an inefficient search pattern. Once an inefficient search was detected, mitigation techniques were employed to encourage the participants to use an efficient search. Behavioral analysis indicates that the most effective mitigation techniques were the use of a nudge to raise the cost of the search and a hint to inform the participants how to search efficiently. These two mitigations techniques increased the number of efficient searches such that the majority of the searches performed in the final block of the experiment were efficient. Additionally, by using machine learning, various models were created that were able to classify whether an inefficient search was occurring. Balanced accuracy scores of greater than 50% were achieved on $\frac{13}{16}$, $\frac{14}{16}$, $\frac{6}{16}$, $\frac{13}{16}$, and $\frac{7}{16}$ participants by the LDA, RFC, ANN, LSTM, and TCN models. These results indicate that detection of an inefficient search is possible but more work is necessary to improve the performance of the models. Improvements in the experiment and for future machine learning tasks are suggested that could improve performance when classifying an inefficient search. Overall, this work has shown that an inefficient search in a visual search can be successfully detected and mitigated.

Appendix A. Pre- and Post-Experiment Questionnaires

Pre-Experiment Questionnaire (ONLY Experiment Day)

How many hours of sleep do you get on average? _____

How many hours of sleep did you have last night? _____

How would you characterize your sleep last night?
Circle one choice: Very Poor, Poor, Fair, Good, Very Good

Did you consume any products with caffeine today?
Circle one choice: yes or no
If yes:
What product(s) did you consume?
_____
When did last consume this product?
_____
Approximately how much (mg / ounces / cups) of this product have you consumed today? _____

Have you had seizures before? yes or no
Have you ever had brain surgery? yes or no
Do you have a history of brain tumors? yes or no
Do you have a history of head trauma? yes or no
Please list any other brain-related health issues that you may have (if any)?
_____
_____

Do you have any reason(s) to believe that your ability to accomplish tasks during this study today would be abnormal (for example: distracted, overly tired, hungry, stressed, injured)?
_____
_____

If yes:
Do you still want to participate in the cognitive study today? Circle one choice: Yes / No
If no:
Would you like to reschedule participation for another day?
_____

Figure 62. Pre-Experiment Questionnaire

Post-Experiment Questionnaire (ONLY Experiment Day)

Computer experience:

What sort of electronic devices do you use?
Circle all choices that apply:
Personal computer/Desktop/Laptop
TV/Game Console
Smartphone/Tablet
Enterprise Server
Other, _____

How often do you use electronic devices?
Circle one choice: Daily, A few times a week, Once a week, Never, Prefer not to answer
How often do you play video games?
Circle one choice: Daily, A few times a week, Once a week, Never, Prefer not to answer

Do you use electronic devices in your job?
Circle one choice: Yes, No, Prefer not to answer

Age: _____

Are you male or female? Male___ Female____ Prefer not to answer ____

What's your highest education level?

A. Lower than high school
B. Graduated from high school
C. Some college, no degree
D. Associate's Degree
E. Bachelor's Degree
F. Master's degree
G. Ph.D. degree

Have you had pilot training or been trained in the scanning of instruments? yes or no

Psychological Knowledge:

On a scale of 1-5 (5: being you studied it extensively on your own, 4: you took a class which covered it, 3: read about it/looked it up, 2: heard the term used in discussion, 1: not familiar with the term), please rate the following:

How familiar are you with cognitive biases? _____

How familiar are you with confirmation bias and/or confirmatory search? _____

Figure 63. Post-Experiment Questionnaire part one

Performance:

Please explain how you searched for the target letter. If your search pattern changed or evolved throughout the experiment, please be sure to detail this change, when it occurred, and why you think that your search pattern changed.

_____

_____

_____

_____

_____

_____

On a scale of 1-5 (5: being you used it in 95% or more of trials, 4: 75% or more of trials, 3: 50% of more, 2: 25% or more, 1: less than 25%), please rate the following:

A confirmatory search pattern is one where you first search circles which match the color shown in the instructions.

- How often did you use confirmatory search overall in today's visual search experiment? _____
- How often did you use confirmatory search in the first 7 blocks of visual search trials (before the first break)? _____
- How often did you use confirmatory search in the last 14 blocks of visual search trials (after the first break)? _____

In the trials there were always two colors and one color had more circles than the other. An *efficient* search is one in which you only look at the lesser-represented-color circles to find the letter.

- How efficient do you believe you were overall in the first 7 blocks of visual search trials? _____
- How efficient do you believe you were overall in the last 14 blocks of visual search trials? _____

The covering of the letters so that they would not appear until gazed upon was a technique called a "nudge". This was used in order to encourage you to adopt an efficient search pattern, by adding a cost (i.e. time) to your search.

For each of the following questions – circle only one answer. Did the nudge…

Help reduce your tendency towards a biased visual search? Yes, No, Unsure

Annoy or irritate you when it was included in searching? Yes, No, Unsure

After the nudge was first introduced, if it was ever removed for a block of trials, do you believe your search for that block was efficient? Yes, No, Unsure

Figure 64. Post-Experiment Questionnaire part two

Appendix B. Cognionics EEG Trigger Values

Table 37. Cognionics EEG Trigger Values Part One

Event	Experiment Marker	Byte Value	Trigger Value
Spacebar Pressed After Instruction	Start of 1st baseline	90	23040
	End of 1st baseline	91	23296
	Start of last baseline	92	23552
	End of last baseline	93	23808
Before every trial	Block 1	1	256
	Block 2	2	512
	Block 3	3	768
	Block 4	4	1024
	Block 5	5	1280
	Block 6	6	1536
	Block 7	7	1792
	Block 8	8	2048
	Block 9	9	2304
	Block 10	10	2560
	Block 11	11	2816
	Block 12	12	3072
	Block 13	13	3328
	Block 14	14	3584
	Block 15	15	3840
	Block 16	16	4096
	Block 17	17	4352
	Block 18	18	4608
	Block 19	19	4864
	Block 20	20	5120
	Block 21	21	5376
	Block 22	22	5632
	Block 23	23	5888
	Block 24	24	6144
Trial Started After Fixation Delay (spacebar pressed)	Trial 1	31	7936
	Trial 2	32	8192
	Trial 3	33	8448
	Trial 4	34	8704
	Trial 5	35	8960
	Trial 6	36	9216
	Trail 7	37	9472
	Trial 8	38	9728
	Trail 9	39	9984
	Trial 10	40	10240
	Trial 11	41	10496
	Trial 12	42	10752
	Trial 13	43	11008
	Trial 14	44	11264
	Trial 15	45	11520
	Trial 16	46	11776
	Trial 17	47	12032
	Trial 18	48	12288
	Trial 19	49	12544
	Trial 20	50	12800

Table 38. Cognionics EEG Trigger Values Part Two

Event	Experiment Marker	Byte Value	Trigger Value
End of Trial (response submitted, z or c key press)	Mark Confirm	60	15360
	Mark NOT Confirm	61	15616
	Mark Efficient	62	15872
	Mark NOT Efficient	63	16128
	Mark Correct Response (user key press)	64	16384
	Mark Incorrect Response (user key press)	65	16640
	Mark Circular	66	16896
	Mark NOT circular	67	17152
	Mark Weird	68	17408
	Mark NOT Weird	69	17664
	Mark Miss	82	20992
	Mark NOT Miss	83	21248
	Mark Minority Only	84	21504
	Mark NOT Minority Only	85	21760
	Mark Single Majority Then Minority	86	22016
	Mark NOT Single Majority Then Minority	87	22272
Before every trial	Mark Nudge Present	70	17920
	Mark No Nudge Present	71	18176
	Mark Hint Seen	72	18432
	Mark No Hint Seen	73	18688
	Mark Hint First Appeared	74	18944
	Mark Explanation Seen	75	19200
	Mark No Explanation Seen	76	19456
	Mark Explanation First Appeared	77	19712
After Block Concludes (i.e. trial 20 finishes)	Mark Block Confirmatory	80	20480
	Mark Block Non-Confirmatory	81	20736

Appendix C. Abbreviated Informed Consent Document

Abbreviated Informed Consent Document
Cognitive Bias Estimation in Decision Making
FWR20180174H v1.1

You are being asked to participate in a research study.

Key study information you should know:

- The purpose of the study is to determine the relationships between behavior, self-reported information, and physiological measures when cognitive bias is, and is not present. Also, the efficacy of bias mitigation of bias will be determined.
- Risks or discomforts from this research are minimal, but could possibly include discomfort from computer use, temporary hair loss and there is a theoretical risk of transmitting skin-borne pathogens during a skin-cleaning process when applying sensors to the skin.
- This study will not benefit you directly.
- Taking part in this research project is voluntary. You can discontinue participation at any time without penalty or loss.

If you participate in this research, you will be performing visual search tasks on a computer. An Electroencephalograph (EEG) head cap will be applied to measure brain activity. Sensors placed near your eyes for Electrooculography (EOG) will measure eye movement and blink signals while sensors on your chest will record heart information using Electrocardiography (ECG). Galvanic Skin Response (GSR) will measure electrodermal activity (EDA) and will be placed on the palm of your non-dominant hand. Gaze tracking data will be collected non-invasively using a Smart Eye Pro eye tracker. Training will take up to 1 hour and the separate experiment will take up to 2 hours on a different day within a two week period (total of up to 3 hours).

This study will require you to use a computer. Beyond the potential discomfort experienced by everyday computer use, there are other possible sources of risk & discomfort. We will attach sensors on your head, face and arms. Some participants may experience discomfort (due to limited movement during trials). Minor skin irritation and/or discomfort may result when the electrodes are placed the head, face, chest and abdomen when you or the testers clean those locations to reduce electrical impedance in order to improve signal quality. Minimal, temporary hair loss is unlikely but may occur locally at the electrode sites. Because applying the electrical sensors to participants requires making contact with and, in some cases, scrubbing the skin with liquids and exfoliating materials, there is a theoretical risk of transmitting skin-borne pathogens during this process. All necessary equipment will be cleaned and disinfected before the procedure begins to minimize the risk of transmitting diseases.

Steps will be taken to protect your confidentiality: The data collected from your questionnaire, computer activities and the sensors on your body will be associated with a randomly-assigned participant number, but no personal information will be collected with this data. The un-identifiable data will be protected with a password on the collection computers and via CAC access on the AFIT network. The researchers will/will not collect any identifiers linked to you. No participant identifiable information will be included in any publications. Any paper data collected will be kept in a locked cabinet.

The data may be accessed by the Department of Defense for auditing purposes.

If you have questions regarding the study, contact the Principal Investigator: Dr. Brett Borghetti can be reached at (937) 255-3636x4612. If you have questions regarding your rights as a research subject, contact the AFRL IRB: 937-904-8100 or afrl.ir.protocolmanagment@us.af.mil.

Cognitive Bias Estimation in Decision Making
FWR20180174H v1.01
AFRL IRB APPROVAL VALID FROM 1 AUGUST 2019 THROUGH 31 JULY 2024

Figure 65. Abbreviated Informed Consent Document

Appendix D. International Review Board Approval Letter

DEPARTMENT OF THE AIR FORCE
AIR FORCE RESEARCH LABORATORY
WRIGHT-PATTERSON AIR FORCE BASE OHIO 45433

MEMORANDUM FOR AFIT (BRETT BORGHETTI)

FROM: 711 HPW/IR

SUBJECT: IRB Approval for the Use of Human Volunteers in Research

1. Protocol Title: Cognitive Bias Estimation in Cyber Security

2.

Protocol Number	Protocol Version	Risk Level
FWR20181047H	V1.01	Minimal Risk

3.

Approval Date	Expiration Date	Re-approval Request Due by:
2 August 2019	1 August 2024	1 July 2024

4. This study received Expedited Review using regulatory category: 32CFR219.110 (b)(4) & (b)(7)

5. Based on 21 January 2019 changes to 32 CFR 219 (typically referred to as the "Common Rule"), re-approval requirements were revised and the AFRL IRB now allows a five year approval period. A continuing review or final report is due prior to the expiration date listed above. IR administrative staff may conduct post approval monitoring over the course of this study. Changes to the study methodology or study staff must still be submitted to the AFRL IRB.

6. Summary: The study objective is identify statistical relationships, and computationally model the associations between three components: Self-reported measures of confirmation bias in decision-making; Behavior patterns during investigative decision-making; Physiological signals collected during decision-making. **Amendments:** Personnel modifications were made to the protocol. The smart eye tracking system and gaze tracks were added to the methodology to determine where the subjects are looking on the screen. The stimulation was changed from scenario based stimulus to a visual search stimulus to enable more trails and data collection, and cognitive bias mitigation was added to the protocol.

7. All inquiries and correspondence concerning this protocol must include the protocol number and name of the primary investigator. Please contact the 711 HPW/IR office using the organizational mailbox at AFRL.IR.ProtocolManagement@us.af.mil or by calling 937-904-8100.

ALLEN.RHONDA.COLLEEN.1395901547 Digitally signed by ALLEN.RHONDA.COLLEEN.1395901547 Date: 2019.08.02 08:00:37 -04'00'

RHONDA C. ALLEN, MAS, CIP
Vice-Chair, AFRL IRB

DISTRIBUTION C: Distribution authorized to U.S. Government agencies (administrative). Other requests for, or further releases of this document shall be referred to 711 HPW/IR.

Figure 66. IRB Approval Letter

Bibliography

1. A. Tversky and D. Kahneman, "Judgment under uncertainty: Heuristics and biases," *Science*, vol. 185, pp. 1124–1131, 1974.

2. National Research Council, "Measuring Human Capabilities: An Agenda for Basic Research on the Assessment of Individual and Group Performance Potential for Military Accession," Washington, D.C., Tech. Rep., apr 2015. [Online]. Available: http://www.nationalhttp://www.nap.edu/catalog/19017https://doi.org/10.17226/19017.

3. A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive Psychology*, vol. 12, no. 1, pp. 97–136, 1980.

4. M. Ashcraft and G. Radvansky, *Cognition (6th Edition)*, 2013.

5. R. S. Nickerson, "Confirmation bias: A ubiquitous phenomenon in many guises." *Review of General Psychology*, vol. 2, no. 2, pp. 175–220, 1998. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/1089-2680.2.2.175https://pdfs.semanticscholar.org/70c9/3e5e38a8176590f69c0491fd63ab2a9e67c4.pdf

6. J. D. Rajsic, "Confirmation Bias in Visual Attention," Ph.D. dissertation, University of Toronto, 2017. [Online]. Available: https://tspace.library.utoronto.ca/handle/1807/80783

7. L. G. Barron and M. R. Rose, "Multitasking as a predictor of pilot performance: Validity beyond serial single-task assessments," *Military Psychology*, vol. 29, no. 4, pp. 316–326, 2017.

8. C. Fine, *A mind of its own : how your brain distorts and deceives*.

9. J. H. Korteling, A.-M. Brouwer, and A. Toet, "A neuroscientific perspective on cognitive biases," 2017. [Online]. Available: https://osf.io/5wj6y/

10. J. Klayman and Y.-w. Ha, "Confirmation, disconfirmation, and information in hypothesis testing." *Psychological Review*, vol. 94, no. 2, pp. 211–228, 1987. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.94.2.211

11. M. Oaksford and N. Chater, "A Rational Analysis of the Selection Task as Optimal Data Selection," Tech. Rep. 4, 1994. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.319.1554{&}rep=rep1{&}type=pdf

12. J. Friedrich, "Primary error detection and minimization (PEDMIN) strategies in social cognition: A reinterpretation of confirmation bias phenomena." *Psychological Review*, vol. 100, no. 2, pp. 298–319, 1993. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.100.2.298

13. E. Shafir and R. A. Leboeuf, "Rationality," *Annual Review of Psychology*, no. 53, pp. 491–517, 2002. [Online]. Available: https://pingpong.ki.se/public/pp/public{_}courses/course13241/published/1544688977468/resourceId/19013462/content/Shafir{&}LeBoeuf(2002)Rationality.pdf

14. K. E. Stanovich and R. F. West, "Individual differences in reasoning: implications for the rationality debate?" *The Behavioral and brain sciences*, vol. 23, no. 5, pp. 645–65; discussion 665–726, oct 2000. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/11301544

15. J. A. Bargh and M. J. Ferguson, "Beyond behaviorism: On the automaticity of higher mental processes." *Psychological Bulletin*, vol. 126, no. 6, pp. 925–945, 2000. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-2909.126.6.925

16. A. Dijksterhuis and L. F. Nordgren, "A Theory of Unconscious Thought," *Perspectives on Psychological Science*, vol. 1, no. 2, pp. 95–109, jun 2006. [Online]. Available: http://journals.sagepub.com/doi/10.1111/j.1745-6916.2006.00007.x

17. R. Sugden, J. Zheng, and D. J. Zizzo, "Not all anchors are created equal," *JOURNAL OF ECONOMIC PSYCHOLOGY*, vol. 39, pp. 21–31, 2013. [Online]. Available: http://dx.doi.org/10.1016/j.joep.2013.06.008

18. F. Bacon, *Novum Organum*, J. Devey, Ed. New York: Collier P.F., 1902. [Online]. Available: http://scholar.google.com/scholar?hl=en{&}btnG=Search{&}q=intitle:NOVUM+ORGANUM+OR+TRUE+SUGGESTIONS+FOR+THE+INTERPRETATION+OF+NATURE{#}0{%}5Cnhttp://scholar.google.com/scholar?hl=en{&}btnG=Search{&}q=intitle:Novum+organum:+or,+True+suggestions+for+the+interpretation+

19. P. C. Wason, "On the failure to eliminate hypotheses in a conceptual task," *Quarterly Journal of Experimental Psychology*, vol. 12, no. 3, pp. 129–140, 1960. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/17470216008416717http://web.mit.edu/curhan/www/docs/Articles/biases/12{_}Quarterly{_}J{_}Experimental{_}Psychology{_}129{_}{%}28Wason{%}29.pdf

20. ——, "Reasoning about a rule," *Quarterly Journal of Experimental Psychology*, vol. 20, no. 3, pp. 273–281, aug 1968. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/14640746808400161

21. M. Snyder and W. W. B. Swann, "Hypothesis-testing processes in social interaction." *Journal of Personality and Social Psychology*, vol. 36, no. 11, pp. 1202–1212, 1978. [Online]. Available: http://content.apa.org/journals/psp/36/11/1202http://psycnet.apa.org/psycinfo/1980-09663-001

22. G. Gigerenzer and K. Hug, *Domain-specific reasoning: Social contracts, cheating, and perspective change*, 1992, vol. 43, no. 2. [Online]. Available: http://library.mpib-berlin.mpg.de/ft/gg/GG{_}Domain{_}1992.pdf

23. H. B. Enderton, "A Mathematical Introduction to Logic, 2nd Edition," Tech. Rep. [Online]. Available: http://agnigarh.tezu.ernet.in/{~}zubin/isc/extra/MathematicalIntroductionToLogic-Enderton.pdf

24. J. L. C. Tooby, "Foundations of culture .Pdf," J. Barkow, L. Cosmides, and J. Tooby, Eds. New York: Oxford University Press, 1992. [Online]. Available: https://www.cep.ucsb.edu/papers/pfc92.pdf

25. D. J. Koehler, "Explanation, imagination, and confidence in judgment." *Psychological bulletin*, vol. 110, no. 3, pp. 499–519, nov 1991. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/1758920

26. C. R. Mynatt, M. E. Doherty, and W. Dragan, "Information Relevance, Working Memory, and the Consideration of Alternatives," *The Quarterly Journal of Experimental Psychology Section A*, vol. 46, no. 4, pp. 759–778, nov 1993. [Online]. Available: http://journals.sagepub.com/doi/10.1080/14640749308401038

27. H.-M. Süß, K. Oberauer, W. W. Wittmann, O. Wilhelm, and R. Schulze, "Working-memory capacity explains reasoning ability—and a little bit more," *Intelligence*, vol. 30, no. 3, pp. 261–288, may 2002. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0160289601001003

28. N. Cowan, "The magical number 4 in short-term memory: a reconsideration of mental storage capacity." *The Behavioral and brain sciences*, vol. 24, no. 1, pp. 87–114; discussion 114–85, feb 2001. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/11515286

29. S. J. Luck and E. K. Vogel, "The capacity of visual working memory for features and conjunctions," *Nature*, vol. 390, no. 6657, pp. 279–281, nov

1997. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/9384378http://www.nature.com/articles/36846

30. G. A. Miller, "The Magical Number Seven, Plus or Minus Two Some Limits on Our Capacity for Processing Information," *Psychological Review*, vol. 101, no. 2, pp. 343–352. [Online]. Available: http://spider.apa.org/ftdocs/rev/1994/april/rev1012343.html

31. K. Oberauer, "Access to information in working memory: Exploring the focus of attention." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 28, no. 3, pp. 411–421, 2002. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/0278-7393.28.3.411http://spider.apa.org/ftdocs/rev/1994/april/rev1012343.html

32. M. E. Doherty, C. R. Mynatt, R. D. Tweney, and M. D. Schiavo, "Pseudodiagnosticity," *Acta Psychologica*, vol. 43, no. 2, pp. 111–121, 1979. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/0001691879900179

33. K. Fiedler, "Beware of Samples! A Cognitive-Ecological Sampling Approach to Judgment Biases," *Psychological Review*, vol. 107, no. 4, pp. 659–676, 2000. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.476.5455{&}rep=rep1{&}type=pdf

34. J. B. Soll, K. L. Milkman, and J. W. Payne, "A User's Guide to Debiasing," in *The Wiley Blackwell Handbook of Judgment and Decision Making*, 2015, no. May 2014, pp. 924–951.

35. R. P. Larrick, "Debiasing," in *Blackwell Handbook of Judgment and Decision Making*, D. J. Koehler and N. Harvey, Eds. Blackwell Publishing Ltd, 2004, pp. 316–337.

36. B. A. Clegg, R. M. Martey, J. Stromer-Galley, K. Kenski, T. Saulnier, J. E. Folkestad, E. Mclaren, A. Shaw, J. E. Lewis, J. D. Patterson, and T. Strzalkowski, "Game-based Training to Mitigate Three Forms of Cognitive Bias," *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC 2014)*, no. 14180, pp. 1–12, 2014. [Online]. Available: https://www.researchgate.net/publication/264158816{_}Game-based{_}Training{_}to{_}Mitigate{_}Three{_}Forms{_}of{_}Cognitive{_}Bias

37. B. Fischhoff, D. Kahneman, P. Slovic, and A. Tversky, "Debiasing," in *Judgment Under Uncertainty: Heuristics and Biases.*, 1982. [Online]. Available: https://philpapers.org/rec/FISDDS

38. H. R. Arkes, "Costs and benefits of judgment errors: Implications for debiasing." *Psychological Bulletin*, vol. 110, no. 3, pp. 486–498, 1991. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-2909.110.3.486

39. B. M. Hales and P. J. Pronovost, "The checklist—a tool for error management and performance improvement," *Journal of Critical Care*, vol. 21, no. 3, pp. 231–235, sep 2006. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16990087https://linkinghub.elsevier.com/retrieve/pii/S0883944106000815

40. J. S. Lerner and P. E. Tetlock, "Accounting for the effects of accountability." *Psychological Bulletin*, vol. 125, no. 2, pp. 255–275, 1999. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-2909.125.2.255

41. T. Mussweiler, F. Strack, and T. Pfeiffer, "Overcoming the Inevitable Anchoring Effect: Considering the Opposite Compensates for Selective Accessibility," *Personality and Social Psychology Bulletin*, vol. 26, no. 9, pp. 1142–1150, nov 2000. [Online]. Available: http://journals.sagepub.com/doi/10.1177/01461672002611010

42. V. M. Beck, A. Hollingworth, and S. J. Luck, "Simultaneous Control of Attention by Multiple Working Memory Representations," *Psychological Science*, vol. 23, no. 8, pp. 887–898, aug 2012. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/22760886http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3419335http://journals.sagepub.com/doi/10.1177/0956797612439068

43. G. Keren, "Cognitive Aids and Debiasing Methods: Can Cognitive Pills Cure Cognitive Ills?" *Advances in Psychology*, vol. 68, no. C, pp. 523–552, 1990.

44. K. L. Milkman, D. Chugh, and M. H. Bazerman, "How Can Decision Making Be Improved?" *Perspectives on Psychological Science*, vol. 4, no. 4, pp. 379–383, 2009. [Online]. Available: http://dx.doi.org/10.1111/j.1745-6924.2009.01142.x

45. B. Kleinmuntz, "Why we still use our heads instead of formulas: Toward an integrative approach." *Psychological Bulletin*, vol. 107, no. 3, pp. 296–310, 1990. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-2909.107.3.296

46. R. H. Thaler and C. R. Sunstein, *Nudge: Improving decisions about health, wealth, and happiness.* New Haven, CT, US: Yale University Press, 2008. [Online]. Available: https://psycnet.apa.org/record/2008-03730-000

47. C. R. McKenzie, M. J. Liersch, and S. R. Finkelstein, "Recommendations Implicit in Policy Defaults," *Psychological Science*, vol. 17, no. 5, pp. 414–420, may 2006. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16683929http://journals.sagepub.com/doi/10.1111/j.1467-9280.2006.01721.x

48. J. G. Johnson, P. Cohen, E. M. Smailes, S. Kasen, and J. S. Brook, "Television Viewing and Aggressive Behavior During Adolescence and Adulthood," *Science*, vol. 295, no. 5564, pp. 2468–2471, mar 2002. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/11923542http://www.sciencemag.org/cgi/doi/10.1126/science.1062929

49. K. Ly, N. Mazar, M. Zhao, and D. Soman, "A Practitioner's Guide to Nudging," *SSRN Electronic Journal*, mar 2013. [Online]. Available: http://www.ssrn.com/abstract=2609347

50. K. Milkman, J. Beshears, J. Choi, D. Laibson, and B. Madrian, "Following Through on Good Intentions: The Power of Planning Prompts," National Bureau of Economic Research, Cambridge, MA, Tech. Rep., apr 2012. [Online]. Available: http://www.nber.org/papers/w17995.pdf

51. P. W. Schultz, J. M. Nolan, R. B. Cialdini, N. J. Goldstein, and V. Griskevicius, "The Constructive, Destructive, and Reconstructive Power of Social Norms," *Psychological Science*, vol. 18, no. 5, pp. 429–434, may 2007. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/17576283http://journals.sagepub.com/doi/10.1111/j.1467-9280.2007.01917.x

52. R. B. Cialdini, R. R. Reno, and C. A. Kallgren, "A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places." *Journal of Personality and Social Psychology*, vol. 58, no. 6, pp. 1015–1026, 1990. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.58.6.1015

53. D. E. Broadbent, "Perception and Communication," Tech. Rep., 1958. [Online]. Available: https://pure.mpg.de/rest/items/item{_}2300885{_}4/component/file{_}2300884/content

54. J. A. Deutsch, D. Deutsch, D. E. Broadbent, K. L. Chow, D. A. Hamburg, and F. Morrell, "ATTENTION: SOME THEORETICAL CONSIDERATIONS 1," Tech. Rep., 1963. [Online]. Available: http://deutsch.ucsd.edu/pdf/Psych{_}Rev-1963{_}70{_}80-90.pdf

55. E. Awh, A. V. Belopolsky, and J. Theeuwes, "Top-down versus bottom-up attentional control: a failed theoretical dichotomy," *Trends Cogn*

Sci, vol. 16, no. 8, pp. 437–443, 2012. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3426354/pdf/nihms-390710.pdf

56. S. Yantis and S. Jonides, "Abrupt Visual Onsets and Selective Attention: Evidence from Visual Search," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 10, no. 5, pp. 601–621, 1984. [Online]. Available: http://wexler.free.fr/library/files/yantis(1984)abruptvisualonsetsandselectiveattention.evidencefromvisualsearch.pdf

57. R. A. Abrams and S. E. Christ, "Motion Onset Captures Attention," *Psychological Science*, vol. 14, no. 5, pp. 427–432, sep 2003. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/12930472http://journals.sagepub.com/doi/10.1111/1467-9280.01458

58. J. Theeuwes, "Perceptual selectivity for color and form." *Perception & psychophysics*, vol. 51, no. 6, pp. 599–606, jun 1992. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/1620571

59. W. F. Bacon and H. E. Egeth, "Overriding stimulus-driven attentional capture," *Perception & Psychophysics*, vol. 55, no. 5, may 1994. [Online]. Available: https://link.springer.com/content/pdf/10.3758{%}2FBF03205306.pdfhttp://www.ncbi.nlm.nih.gov/pubmed/8008550

60. C. L. Folk, R. W. Remington, and J. C. Johnston, "Involuntary covert orienting is contingent on attentional control settings." *Journal of experimental psychology. Human perception and performance*, vol. 18, no. 4, pp. 1030–44, nov 1992. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/1431742

61. M. F. Land and M. Hayhoe, "In what ways do eye movements contribute to everyday activities?" *Vision research*, vol. 41, no. 25-26, pp. 3559–65, 2001. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/11718795

62. J. Duncan and G. W. Humphreys, "Visual search and stimulus similarity." *Psychological Review*, vol. 96, no. 3, pp. 433–458, 1989. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.96.3.433

63. U. Neisser, "Visual Search," *Scientific American*, vol. 210, no. 6, pp. 94–102, jun 1964. [Online]. Available: http://www.nature.com/doifinder/10.1038/scientificamerican0664-94

64. A. Treisman and S. Sato, "No Title," aug 1990. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/2144564http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.120.4986{&}rep=rep1{&}type=pdf

65. J. M. Wolfe, K. R. Cave, and S. L. Franzel, "Guided search: an alternative to the feature integration model for visual search." *Journal of experimental psychology. Human perception and performance*, vol. 15, no. 3, pp. 419–33, aug 1989. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/2527952

66. J. M. Wolfe and T. S. Horowitz, "What attributes guide the deployment of visual attention and how do they do it?" *Nature Reviews Neuroscience*, vol. 5, no. 6, pp. 495–501, jun 2004. [Online]. Available: http://www.nature.com/articles/nrn1411

67. C. Bundesen, "A theory of visual attention," *Psychological Review*, vol. 97, no. 4, pp. 523–547, oct 1990. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/2247540

68. L. Huang and H. Pashler, "A Boolean Map Theory of Visual Attention," *Psychological Review*, vol. 114, no. 3, pp. 599–631, 2007. [Online]. Available: http://www.pashler.com/Articles/Huang{_}Pashler{_}PR2007.pdf

69. R. Desimone and J. Duncan, "Neural Mechanisms of Selective Visual Attention," pp. 193–222, mar 1995. [Online]. Available: http://wexler.free.fr/library/files/desimone{%}281995{%}29neuralmechanismsofselectivevisualattention.pdfhttp://www.ncbi.nlm.nih.gov/pubmed/7605061http://www.annualreviews.org/doi/10.1146/annurev.ne.18.030195.001205

70. C. N. Olivers, J. Peters, R. Houtkamp, and P. R. Roelfsema, "Different states in visual working memory: when it guides attention and when it does not," *Trends in Cognitive Sciences*, vol. 15, no. 7, pp. 327–34, jun 2011. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/21665518https://linkinghub.elsevier.com/retrieve/pii/S1364661311000854

71. A. Mack and I. Rock, *MIT Press/Bradford Books series in cognitive psychology. Inattentional blindness.* Cambridge, MA, US: The MIT Press, 1998. [Online]. Available: https://psycnet.apa.org/record/1998-07464-000

72. D. J. Simons and C. F. Chabris, "Gorillas in Our Midst: Sustained Inattentional Blindness for Dynamic Events," *Perception*, vol. 28, no. 9, pp. 1059–1074, sep 1999. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10694957http://journals.sagepub.com/doi/10.1068/p281059

73. J. Rajsic, D. E. Wilson, and J. Pratt, "Confirmation bias in visual search," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 41, no. 5, 2015.

74. ——, "The Price of Information: Increased Inspection Costs Reduce the Confirmation Bias in Visual Search," *Quarterly Journal of Experimental Psychology*, vol. 71, no. December, pp. 1–20, apr 2017. [Online]. Available: http://journals.sagepub.com/doi/10.1080/17470218.2016.1278249

75. C. L. Prosser, *Comparative Animal Physiology, Part A, Environmental and Metabolic Animal Physiology, 4th Edition* , 1991. [Online]. Available: https://www.wiley.com/en-us/Comparative+Animal+Physiology{%}2C+Part+A{%}2C+Environmental+and+Metabolic+Animal+Physiology{%}2C+4th+Edition-p-9780471857679

76. R. K. Minas, R. F. Potter, A. R. Dennis, V. Bartelt, and S. Bae, "Putting on the Thinking Cap: Using NeuroIS to Understand Information Processing Biases in Virtual Teams," *Journal of Management Information Systems*, vol. 30, no. 4, pp. 49–82, apr 2014. [Online]. Available: http://www.tandfonline.com/doi/full/10.2753/MIS0742-1222300403

77. M. X. Cohen, *Analyzing Neural Time Series Data: Theory and Practice*, 2014.

78. A. Gérome, *Hands-On Machine Learning with Scikit-Learn and TensorFlow - O'Reilly Media*, 2017. [Online]. Available: http://shop.oreilly.com/product/0636920052289.do

79. T. Hastie, R. Tibshirani, G. James, and D. Witten, *An Introduction to Statistical Learning, Springer Texts*, 2006, vol. 102.

80. B. Binias, D. Myszor, and K. A. Cyran, "A machine learning approach to the detection of pilot's reaction to unexpected events based on EEG signals," *Computational Intelligence and Neuroscience*, vol. 2018, 2018.

81. "Understanding Random Forest - Towards Data Science." [Online]. Available: https://towardsdatascience.com/understanding-random-forest-58381e0602d2 [Accessed: 2020-01-29]

82. P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning Representations from EEG with Deep Recurrent-Convolutional Neural Networks," nov 2015. [Online]. Available: http://arxiv.org/abs/1511.06448

83. "Understanding Random Forest - Towards Data Science." [Online]. Available: https://towardsdatascience.com/understanding-random-forest-58381e0602d2 [Accessed: 2020-01-29]

84. F. Chollet, *Deep Learning with Python & Keras*. Manning Publications Co., 2018, vol. 80, no. 1. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20608803

85. G. F. Wilson and C. A. Russell, "Real-Time Assessment of Mental Workload Using Psychophysiological Measures and Artificial Neural Networks," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 45, no. 4, pp. 635–643, 2003.

86. J. C. Christensen, J. R. Estepp, G. F. Wilson, and C. A. Russell, "The effects of day-to-day variability of physiological data on operator functional state classification," *Neuroimage2*, vol. 59, no. 1, pp. 57–63, 2012.

87. G. Carlsson, "Using Topological Data Analysis to Understand the Behavior of Convolutional Neural Networks," 2018.

88. S. Tripathi, S. Acharya, R. Dev Sharma, S. Mittal, and S. Bhattacharya, "Using Deep and Convolutional Neural Networks for Accurate Emotion Classification on DEAP Dataset," Tech. Rep. [Online]. Available: www.aaai.org

89. J. Donahue, L. A. Hendricks, M. Rohrbach, S. Venugopalan, S. Guadarrama, K. Saenko, and T. Darrell, "Long-Term Recurrent Convolutional Networks for Visual Recognition and Description," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 677–691, apr 2017.

90. R. Hefron, B. Borghetti, C. Schubert Kabban, J. Christensen, and J. Estepp, "Cross-Participant EEG-Based Assessment of Cognitive Workload Using Multi-Path Convolutional Recurrent Neural Networks." *Sensors (Basel, Switzerland)*, vol. 18, no. 5, apr 2018. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/29701668http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5982227

91. S. Bai, J. Z. Kolter, and V. Koltun, "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling," mar 2018. [Online]. Available: http://arxiv.org/abs/1803.01271

92. J. Peirce, J. R. Gray, S. Simpson, M. MacAskill, R. Höchenberger, H. Sogo, E. Kastman, and J. K. Lindeløv, "PsychoPy2: Experiments in behavior made easy," *Behavior Research Methods*, vol. 51, no. 1, pp. 195–203, feb 2019.

93. "SE PRO - Smart Eye." [Online]. Available: https://smarteye.se/research-instruments/se-pro/ [Accessed: 2019-12-31]

94. "MAPPS." [Online]. Available: https://www.eyesdx.com/products/mapps/ [Accessed: 2019-12-31]

95. "Dry EEG — Headsets — Electrodes." [Online]. Available: https://www.cognionics.net/ [Accessed: 2019-12-31]

96. N. Bigdely-Shamlo, T. Mullen, C. Kothe, K. M. Su, and K. A. Robbins, "The PREP pipeline: Standardized preprocessing for large-scale EEG analysis," *Frontiers in Neuroinformatics*, vol. 9, no. JUNE, pp. 1–19, jun 2015.

97. M. B. Pontifex, V. Miskovic, and S. Laszlo, "Evaluating the efficacy of fully automated approaches for the selection of eyeblink ICA components," *Psychophysiology*, vol. 54, no. 5, pp. 780–791, 2017.

98. F. Lotte, M. Congedo, and L. Anatole, "A review of classification algorithms for EEG-based brain – computer interfaces To cite this version : A Review of Classification Algorithms for EEG-based Brain-Computer Interfaces," 2007.

99. O. Ledoit and M. Wolf, "Honey, I Shrunk the Sample Covariance Matrix," Tech. Rep., 2003.

100. S. Kumar, A. Sharma, and T. Tsunoda, "Brain wave classification using long short-term memory network based OPTICAL predictor," *Scientific Reports*, vol. 9, no. 1, dec 2019.

101. R. Wan, S. Mei, J. Wang, M. Liu, and F. Yang, "Multivariate temporal convolutional network: A deep neural networks approach for multivariate time series forecasting," *Electronics (Switzerland)*, vol. 8, no. 8, 2019.

102. F. Yu and V. Koltun, "Multi-Scale Context Aggregation by Dilated Convolutions," nov 2015. [Online]. Available: http://arxiv.org/abs/1511.07122

103. J. E. Korteling, A. M. Brouwer, and A. Toet, "A neural network framework for cognitive bias," *Frontiers in Psychology*, vol. 9, no. SEP, sep 2018.

104. R. G. Hefron, B. J. Borghetti, J. C. Christensen, and C. M. Shubert-Kabban, "Deep long short-term memory structures model temporal dependencies improving cognitive workload estimation," *Pattern Recognition Letters*.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
21–03–2020	Master's Thesis	Sep 2018 — Mar 2020

4. TITLE AND SUBTITLE

Automated Detection and Mitigation of Inefficient Visual Searching Using Electroencephalography and Machine Learning

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Gallaher, Joshua P, 2d Lt, USAF

5d. PROJECT NUMBER

19G192

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering an Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENG-MS-20-M-022

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Office of Scientific Research
Dr. James Lawton
AFOSR Program manager, Information and Networks
(703) 696-5999
james.lawton.1@us.af.mil

10. SPONSOR/MONITOR'S ACRONYM(S)

AFOSR/RTA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Decisions made during the high-stress and fast-paced operations of the military are extremely prone to cognitive biases. A commonly known cognitive bias is a confirmation bias, or the inappropriate bolstering of an unknown hypothesis. One such critical military operation that can fall prey to a confirmation bias is a visual search. During a visual search, a military operator must perform a visual scan of an environment for a specific target. However, the visual search process can fall prey to the same confirmation bias which can cause inefficient searches. This study elicits inefficient visual search patterns and applies various mitigation techniques in an effort to improve the efficiency of the searches. The effects of the various mitigations are studied and the most effective mitigations are determined. Machine learning models are trained to find the relationship between Electroencephalography (EEG) signals and inefficient visual searching.

15. SUBJECT TERMS

confirmation bias, decision making, visual search, electroencephalography (EEG), machine learning

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Brett Borghetti, AFIT/ENG
U	U	U	UU	195	19b. TELEPHONE NUMBER *(include area code)* (937) 255-6565 x4581; brett.borghetti@afit.edu