8-23-2017

Advancing our Understanding of Performance during Sleep Deprivation

Jeffrey Blake Bolkhovsky
University Of Connecticut, jeffrey.bolkhovsky@engineer.uconn.edu

Advancing our Understanding of Performance during Sleep Deprivation

Jeffrey Blake Bolkhovsky, PhD

University of Connecticut, 2017

Sleep deprivation is a concern in professional, military, and personal settings and can potentially lead to critical, life threatening errors. Predicting a person's behavior under the effects of sleep deprivation allows for the mitigation of risk due to reduced performance. The objectives of this dissertation are 1) to examine the role of sleep deprivation on cognitive and motor performance with the goal of contributing to a predictive model of complex real-world tasks, and 2) to explore non-invasive and non-disruptive physiological correlates of sleep deprivation through eye-tracking. One platform to produce predictive models is known as a cognitive architecture, which is used to simulate human performance and is informed by experimentally-determined psychophysical metrics. Currently lacking in literature, and consequently from cognitive architectures, are the generalized effects that sleep deprivation has on performance. To approach this problem, sets of motor control and cognitive tasks were designed as a framework within which to evaluate human performance under the effects of sleep deprivation. Using cognitive architectures, models of human behavior for these tasks were designed to offer insight and assess their current capacity to predict performance. Sleep deprivation studies for these tasks were run to measure the cognitive effects of sleep deprivation, and explore eye tracking as a physiological correlate. The findings in this dissertation suggest that the current state of cognitive modeling could predict human performance of simple tasks (e.g., simple reaction time), but the models failed to adequately simulate human performance during higher complexity tasks (e.g., visual search). Consequently, this work provides both experimental results and a framework from which more accurate models with higher predictive power can be designed. Finally, it was found that multiple eye-tracking measures correlated with the performance of various cognitive tasks, implying the potential use of such measures as alternate predictors of performance while under the effects of sleep deprivation.

Advancing our Understanding of Performance during Sleep Deprivation

Jeffrey Blake Bolkhovsky

B.S., Worcester Polytechnic Institute, **2011**

M.E., Worcester Polytechnic Institute, **2014**

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2017

APPROVAL PAGE

Doctor of Philosophy Dissertation

Advancing our Understanding of Performance during Sleep Deprivation

Presented by

Jeffrey Blake Bolkhovsky, B.S., M.E.

Major Advisor ____

Ki H. Chon

Associate Advisor ____

Michael K. Qin

Associate Advisor ____

Sabato Santaniello

University of Connecticut
2017

Table of Contents

Table of Figures

Table of Tables

Chapter 1: Introduction

1.1 Overview

A powerful insight, made by Card et al. (1986) when discussing measured psychological variables, provides understanding of the importance of modeling: "It is a great mistake to think that, by merely listing a miscellaneous collection of results, psychology is thereby rendered usable to support practical engineering projects (Card, Moran, & Newell, 1986)." One example of this idea coming to life is in cognitive modeling, which takes the various measured physiological and cognitive parameters and puts them into one complete system for prediction and insight into human behavior (Kingdon, 2008). Cognitive models have proved to be powerful tools; however, they are still incomplete. They lack information in areas such as performance decrement with sleep deprivation and fine motor control of the eye. With this in mind, this proposal looks at not just measuring arbitrary psychological and physiological variables, but to determine which ones are lacking or incomplete in current engineering constructs. This proposal is twofold, starting with the engineering and modeling approach that combines collected psychological and cognitive results in current literature for evaluation; then moving on to observational evidence from human experiments to collect missing information.

Cognitive modeling is performed by using cognitive architectures, which are comprehensive software systems that contain a collection of various results from psychology literature (e.g., visual acuity and memory recall time) (Kingdon, 2008). There are several cognitive architectures in the field, such as Adaptive Control of Thought—Rational (ACT-R) and Executive Process-Interactive Control (EPIC) that have collected and put many of these parameters together. Generally, cognitive architectures organize different measured performance information into distinct categories or modules, an example of this can be seen in Figure 1.1a, which shows an outline of the EPIC cognitive architecture. In this architecture, some of the modules presented include the visual processor, which holds information such as visual onset time, and the manual motor processor, which contains information about expected movement time of a person's

hand (Kieras & Meyer, 1996). A cognitive model represents a single simulated person that performs a task within an environment. The cognitive model acts as a program, with the cognitive architecture acting as the programming language. The model generates two outputs: actions taken, which is combined with an environmental input and fed back into the model, and quantified performance time and accuracy. Some example inputs and outputs of a cognitive model are shown in Figure 1.1.

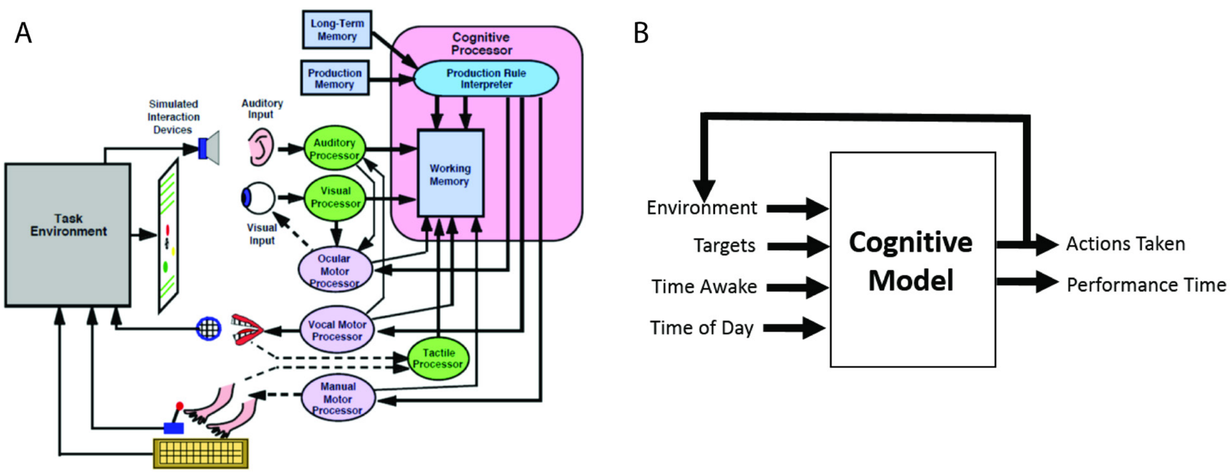

Figure 1.1. Depiction of the EPIC Cognitive Architecture.

(A) The EPIC Cognitive Architecture Diagram, from (Kieras & Meyer, 1996). (B) Illustration of inputs and outputs of a cognitive architecture, cognitive model, and task environment.

The interactions between a cognitive model and cognitive architecture allow cognitive models to predict human performance for complex tasks using the conglomerate of results from simpler ones found within the architecture. However, with incomplete knowledge of human performance (e.g., performance decrement with fatigue, visual search accuracy), commonly used cognitive models are unable to fully describe simple tasks (Gunzelmann, Gross, Gluck, & Dinges, 2009), and thus before cognitive architectures can be used for complex predictions, they must fully realize their capacity for simple ones.

As stated, one of the major areas where information on human behavior is missing is in performance decrement during fatigue, frequently induced by sleep deprivation. In general, it has been established that sleep deprivation has the most significant effect on alertness and attention, frequently

causing lapses or short periods of non-action following a stimulus (Alhola & Polo-Kantola, 2007; Lim & Dinges, 2008; H. L. Williams, Lubin, & Goodnow, 1959). In addition, it has been shown that sleep deprivation can result in an overall slowing of responses (e.g., motor tasks) (Alhola & Polo-Kantola, 2007; Lim & Dinges, 2010). With such broad findings many have argued that sleep deprivation causes a decrement in overall executive function (generalized mental function focused on accomplishing tasks), which affects our overall behavior (Binks, Waters, & Hurry, 1999; Nilsson et al., 2005).

As of this moment sleep deprivation is not implemented in any cognitive architectures, mainly due to the lack of information of how sleep deprivation affects each individual module within a cognitive architecture. Thus, this proposal looks to create cognitive models with added performance decrement due to sleep deprivation, based on current literature, and verify the expected findings or identify current flaws with experimental trials. Some of the major variables discussed in this context include reaction time, motor performance, working memory, and visual search. This effort looks to examine whether the current work in the field is widely applicable, and to provide missing performance information that will lay the foundation for the advancement of cognitive architectures for complex task modeling in the future.

1.2 Statement of Purpose

There is currently a lack of knowledge in the field of cognitive science of many of the basic effects that sleep deprivation has on human performance. Systems such as cognitive architectures put a collection of measured variables into one structure to predict human performance. Currently cognitive architectures do not include information on the effects of sleep because much of this information is still unknown. Though there are a multitude of mathematical models to describe performance during sleep deprivation, they lack the depth to describe the multipart effects of sleep deprivation over human cognition. Without more information on the effects of sleep deprivation on cognitive processes, current models are too broad to contribute significant advancements to the field and predict complex tasks.

To advance the field of cognitive science, and provide information for the advancement on the prediction of complex tasks, we will take a three-stage approach in both examining current models and then using empirically based experimental measures to acquire new information that can be used to advance such models. This will be done by modeling human performance during sleep deprivation for a subset of motor and cognitive tasks. An experiment examining motor control during sleep deprivation will then be run to compare to the previous models. Next, an experiment examining cognitive task performance during sleep deprivation will be run, once again to compare with current models. Finally, the performance of all tasks will be quantifiably assessed to determine how each examined task is affected by sleep deprivation.

1.3 Specific Aims

As mentioned above, there will be three primary steps in this dissertation; one, which is focused on modeling, and the other two, which are experimentation based. Thus, the research direction will be split into three primary aims. During the first aim, the subset of tasks being examined within the scope of this dissertation will be decided, followed by modeling based on the current literature. Thus, having Aim one focused on modeling. Aim two will examine the motor control tasks specified in Aim one using empirically based experimental procedures. Finally, Aim three will examine cognitive tasks, once again using experimental methods. The three aims can be more formally defined as follows:

> Aim 1: Modeling: Evaluate current cognitive models' ability to predict performance under the effect of sleep deprivation and identify where there is insufficient information on performance degradation during sleep deprivation.
>
> Aim 2: Motor Control: Quantify the effects of sleep deprivation on a subset of manual targeting tasks in humans.
>
> Aim 3: Cognitive Tasks: Quantify effects of sleep deprivation on a subset of cognitive tasks in humans.

The section that follows will describe what the goal of each aim is, in more detail. In addition, it will describe the purpose for each aim and how they relate to each other.

1.3.1 Aim 1: Modeling

The primary purpose of Aim one is to gain insight into how sleep deprivation will affect people over a variety of tasks. While looking at expected performance, it is possible to examine cognitive architectures, such as ACT-R and EPIC, to learn their modeling extent as well as current limitations. To this end, a specific motor control task will be chosen, as well as a set of cognitive tasks that examine different modes of human performance.

Regarding the motor control task, a tilt-based targeting task was used as the principle measurement focus. There are number of factors for choosing such a task. The first factor in the decision describes how this type of task has been shown to conform to Fitt's Law, a psychological model of human performance, and has been used to provide information regarding how the human psychomotor system processes targeting tasks (P.M. Fitts & Posner, 1967). Since Fitts' law can be used as a multifaceted tool for measuring human performance, it will allow for easy comparison to similar tasks in literature, which will be presented in more detail in Chapter 2. Tilt based tasks have also become much more prevalent with current technology, and allow for complex 3-dimensional inputs. This task will be modeled using the EPIC cognitive architecture, due to its increased ability to model motor actions over other architectures, which will once again be discussed in more detail in Chapter 2 (Kieras & Meyer, 1996).

Three cognitive-based tasks were chosen to describe multiple types of processing. The first is the psychomotor vigilance task (PVT), which measures vigilant attention or sustained concentration. The PVT was chosen because it is used as the standard for measuring performance during sleep deprivation (Lim & Dinges, 2010). The other tasks selected are working memory based tasks, and visual search tasks. These tasks were selected because they each have components separate from simply vigilant attention, of

which the information surrounding their performance during sleep deprivation is sparse. These tasks were all modeled with the ACT-R cognitive architecture

1.3.2 Aim 2: Motor Control

Aim two is centered on two primary purposes, the first being to verify cognitive models by describing how well humans perform manual targeting tasks (e.g., ability to hit a target) and to ascertain that they behave as expected. The second purpose is to advance our understanding of manual targeting tasks by determining what the observation evidence describes regarding the relationship between performance and sleep deprivation in the area.

It was stated that in Aim one a tilt-based targeting task is modeled. With the goal of Aim two to ascertain the ability for a person to perform a tilt-based motor task analogous to the one modeled in task one, a tilt-based task similar to the one designed for the model was developed. Thus, using an android based tablet, the targeting based tilt task was designed. Following the design of the motor control task, a 24-hour sleep deprivation study was executed. During this study participants remained awake for a minimum of 24 hours and performed the manual control task at specified intervals to collect human performance data.

1.3.3 Aim 3: Cognitive Tasks

Aim three, like Aim two, seeks to verify the performance predicted by cognitive models with regards to simple reaction time, working memory, and visual search. Then once again, with experimental evidence it is possible to determine the relationship between human performance and sleep deprivation to consequently allow for more complex and higher fidelity modeling in the future.

To allow for comparison between modeling and measured results from humans, three tasks are designed to serve as analogs to the cognitive tasks modeled above. First, a psychomotor vigilance task is used to allow for baseline comparison to, not only the models, but also to other results from literature. Next, an n-back task is designed to allow for the evaluation of working memory. Finally, a visual search

task based on searching for small objects on a screen-based interface is designed to allow for the analysis of the performance of visual search. These tasks are all designed for a single computer system as to keep consistency and simplicity for the experiment.

Just as before, a 24-hour study was designed to allow for the testing of all of these cognitive tasks, and to examine how performance changes with time awake. Once again, all task starts are administered at regular intervals to allow for tracking performance over the period of the experiment. In addition, some physiological parameters are measured during these tasks to gain value information on human behavior that is not currently available in literature.

Chapter 2: Literature Review and Background

2.1 Sleep Deprivation

There are two types of sleep deprivation: partial and acute. Partial sleep deprivation is induced when a person does not receive the required amount of sleep per day. The time required for sleep varies between individuals, but it tends to be around 7 to 8.5 hours a day (Alhola & Polo-Kantola, 2007). Reducing that amount of time results in partial sleep deprivation. Acute sleep deprivation is caused when a person is prevented from sleeping; this usually occurs over a time span of 24-72 hours for most sleep studies (Alhola & Polo-Kantola, 2007). This dissertation will focus on acute sleep deprivation, and this section will go over some of the background and literature discussing how human performance is affected by acute sleep deprivation.

Many studies have shown that sleep deprivation has a significant effect on attention and vigilance, causing lapses or temporary failures in concentration (Alhola & Polo-Kantola, 2007; Lim & Dinges, 2008; H. L. Williams et al., 1959). Slowing of general cognitive function has also been observed, even with activities that are unaffected by lapses, and across a variety of different tasks (Alhola & Polo-Kantola, 2007). However, there have been claims that sleep deprivation affects tasks differently based on their properties (e.g., required concentration, complexity), and also that sleep loss causes a non-specific effect across different categories of cognitive tasks (Balkin, Rupp, Picchioni, & Wesensten, 2008; Durmer & Dinges, 2005; Pilcher, Band, Odle-Dusseau, & Muth, 2007). Although there are conflicting views in literature on the effects of sleep deprivation, there are several hypotheses that explore the effects of sleep deprivation on performance that are well described by Lim & Dinges, 2010.

One hypothesis, based on controlled attention, finds that tasks which require more cognitive activity or concentration are less affected by sleep deprivation (Lim & Dinges, 2010). Following this idea, tasks which tend to be more monotonous and less engaging are consequently more affected by a lack of sleep (Pilcher et al., 2007). Pilcher et al. compared a variety of different tasks when examining this problem,

including the PVT and one known as the wombat, a task which is complex, stress inducing, attention demanding, and includes consistent decision making. The difference between the changes in performance between the two tasks was one of the main pushes that spurred the controlled attention hypothesis. These results were supported by other studies finding differences based on tasks that had varying levels of required attention (Magill et al., 2003; Smith & Maben, 1993).

Another prominent hypothesis, which separates tasks into different categories, is neuropsychological, and it describes how sleep loss is manifested in lower activation in the prefrontal cortex region of the brain (Chee & Choo, 2004; Lim & Dinges, 2010). This description points to tasks that are pre-frontal cortex based as being affected more by sleep loss (Durmer & Dinges, 2005). Multiple studies have been run looking at tasks that show activity in the prefrontal cortex with performance during sleep deprivation (Jones & Harrison, 2001; Kerkhof & Van Dongen, 2010). For example, in one study researchers examined brain activity using positron emission tomography (PET) to examine brain activity during 85 hours of sleep deprivation, finding reduced activity in multiple regions in the brain, including the prefrontal cortex, which corresponded to decreased performance during sleep deprivation (Drummond et al., 2005). The results of these studies have provided evidence that supports the neuropsychological hypothesis.

Finally, the third hypothesis suggests that the vigilance, or sustained concentration, required of a task is the primary factor in predicting the extent to which the performance is affected by sleep loss. (Lim & Dinges, 2010). This hypothesis has been fairly predominant in literature as the PVT, a solely vigilance-based task, has been the primary performance assay for sleep deprivation (Dorrian, Rogers, & Dinges, 2005; Van Dongen, Maislin, Mullington, & Dinges, 2003). The PVT has not only been shown to be sensitive to sleep, but several studies have also shown its sensitivity to circadian rhythm as well as the homeostatic modulation of sleep (Dinges et al., 1997; Doran, Van Dongen, & Dinges, 2001). As a consequence, claims that the measures of reaction time in attention and vigilance based tasks are the

leading methods for assessing performance during sleep deprivation have gained strength (Lim & Dinges, 2010).

Though the hypothesis describing performance decrement with sleep deprivation are so distinct they are not mutually exclusive. With so much uncertainty surrounding the topic of performance and sleep loss, it becomes difficult to predict how any particular type of task is affected by sleep loss. In fact, many argue that performance decrement due to sleep loss is so extensive that it is reasonable to assume that sleep loss exerts a non-specific effect on performance in general (Balkin et al., 2008; Lim & Dinges, 2010). Thus, it is imperative to examine various tasks and empirically determine how sleep deprivation impacts performance on each of them in addition to, how well this performance is predicted by the current standards of attention based measurements (Lim & Dinges, 2010).

2.2 Modeling

2.2.1 Two-Process Models

Sleep deprivation, in terms of performance modeling, has been described by many mathematical functions, however, the most prominent ones are two-process models. These two-process models have served as the major conceptual framework for sleep research (Achermann & Borbély, 1994; Borbély, Daan, Wirz-Justice, & Deboer, 2016, p.; Jewett & Kronauer, 1999). The two processes that drive these frameworks are known as homeostatic and circadian (Borb & Achermann, 1999; Borbély et al., 2016). While there are other factors and minimal phenomenon that affect performance during sleep deprivation, the homeostatic and circadian processes are the most prominent that interact continuously (Borbély et al., 2016).

The homeostatic process is a representation of sleep debt, with which a greater sleep debt leads to reduced performance. The value of this process increases while a person remains awake, and decreases while a person sleeps. There have been many measurements that have shown markers representing homeostasis, including eye movement, electroencephalography (EEG), and slow wave activity (SWA) in

the brain (Borbély et al., 2016). The other primary process, called the circadian pacemaker, represents value ranges that oscillate throughout a 24-hour day cycle, usually in accordance with the day and night cycle. Markers of the circadian process have been measured using core body temperature as well as measurements of melatonin (Borbély et al., 2016).

One frequently used manifestation of this model is known as the Sleep, Activity, Fatigue, and Task Effectiveness (SAFTE). The equation describing SAFTE is shown in Eq. 2.1 below (Hursh et al., 2004).

$$\mathrm{E}(t) = 100 * \frac{R(t)}{\mathrm{R_c}} + \mathrm{C}(t) + \mathrm{I} \qquad \textbf{(2.1)}$$

The SAFTE equation describes task effectiveness as *E(t)*, which is based on three parts: $100*R(t)/R_c$ representing the reservoir level or a representation of the homeostatic process, *C(t)* representing the circadian process, and *I* representing sleep inertia, a period of lower performance just after waking up (Hursh et al., 2004). In this model, the homeostatic process is treated as a reservoir in which the *R(t)* variable decreases with time awake, and increases with time asleep. The circadian portion is represented by the sum of cosines which have a period of 24 hours, and peaks that adjust for times of day based on previous studies. Finally, sleep inertia affects task effectiveness during the first two hours of wakeful periods (Hursh et al., 2004). Mathematical models like the one described above allow for an estimation of task performance but lack task specificity.

Many models that examine performance during sleep deprivation use or are based off of the SAFTE model, as it is commercially available (Dawson, Darwent, & Roach, 2017). For example, SAFTE was adapted for a model created by McCauley et al 2013., which added additional time dependencies based on circadian misalignment during sleep. This model, still working on the basis of a two-process system, added additional modulators to the circadian rhythm portion, allowing for the prediction of performance on a more specific, per-person basis (McCauley et al., 2013). Another modification is discussed by Ingre et al., 2014, which extended the two-process model into three processes. The primary aim was to add an additional modulation component to account for inconsistencies in sleep caused by events such as jetlag

or sudden schedule shifts, primarily for use with airline operations (Ingre et al., 2014). Sleep-based performance models, such as those mentioned above, have led the way to allow for greater risk assessment in the future.

2.2.2 Cognitive Architectures and Models

To increase the understanding of the relationship between sleep deprivation and human performance, some groups have chosen to simulate sleep deprivation tasks using computational cognitive modeling within a cognitive architecture (Gunzelmann et al., 2009). Unlike many standard mathematical modeling techniques, cognitive architectures act as a blueprint for cognition and focus on predicting human behavior during specific tasks (Duch, Oentaryo, & Pasquier, 2008). The architecture incorporates various basic information processing mechanisms predictably used by humans, that have been collected from literature (e.g., memory retrieval, typing speed, saccadic velocity), and allows a computer to simulate tasks based on human abilities (Kieras & Meyer, 1997).

One of the drawbacks of computation architectures is the lack of accurate ways for predicting the changes in the processing mechanisms of cognitive models due to sleep deprivation (Gunzelmann et al., 2009). Gunzelmann et al. (2009) have made an attempt at characterizing the effects of sleep deprivation within the Adaptive Control of Thought—Rational (ACT-R) cognitive architecture. The architecture adaptations consisted mainly of the manipulation of constants that influenced the information processing systems of the architecture based on the time an individual had spent continuously awake. The manipulation performed leads to a steady increase in errors of commission, median reaction time, and number of lapses as the simulated time awake increased. These changes in performance were recorded during measurements of psychomotor vigilance tests during sleep deprivation and were found to match the empirical data from other experiments.

While this work was informative, cognitive architectures are constantly evolving and some of the variables used in the previous study are no longer part of the newer systems, as cognitive models are updated and changed (Collins, Juvina, & Gluck, 2016). Thus, there is a need to gain more information on

how conditions, such as sleep deprivation, affect human performance in order to better incorporate the information into newer models. The following section will go into greater depth regarding the idea behind a cognitive architecture, and discuss some of the architectures currently in use.

2.2.2.1 Human Model Processor

Cognitive architectures began with the idea of taking various measurements for human cognition collected over a large number of studies and putting them into a system which can mimic human behavior (Anderson, 2013). Card et al. (1986) put this idea into a more concrete concept known as the model human processor. As seen in Figure 2.1, the goal of such a model is to separate out various pieces of cognition (e.g., sensory input, long-term memory) into categories and subcategories, and determine which cognitive processes, such as visual acuity and memory recall time, fit into each of those areas (Card et al., 1986). This model separates cognition into distinct modules, each with specific roles in cognition as a whole. Shown below, in the example, are four modules: perceptual processor, memory, cognitive processor, and motor processor.

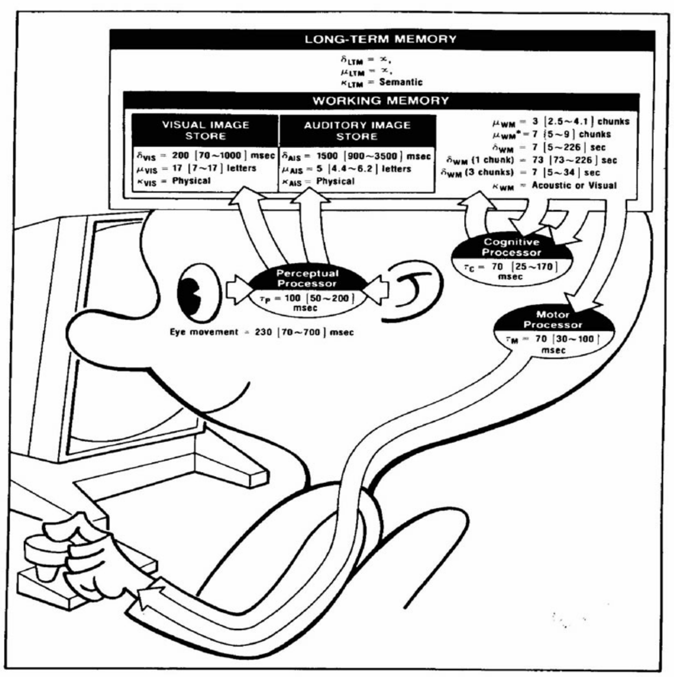

Figure 2.1. The Human Model Processor described by Card et al. (1986)

The definitions and organization of these modules leads to the capacity for simulating human behavior. Though, there are a number of ways to go about implementing these variables into a unified system, as many cognitive architectures have been developed. This dissertation will focus on examining only a few of the more prominent architectures.

2.2.2.2 Adaptive Control of Thought – Rational

The ACT-R architecture is structured around the idea that major areas of the brain process different types of information, and thus the primary modules are separated in a similar manner (Anderson et al., 2004). An overview of this system of modules can be seen below in Figure 2.2. The integration of these modules is built to mimic cognition, and is not meant to reproduce or model neurophysiology.

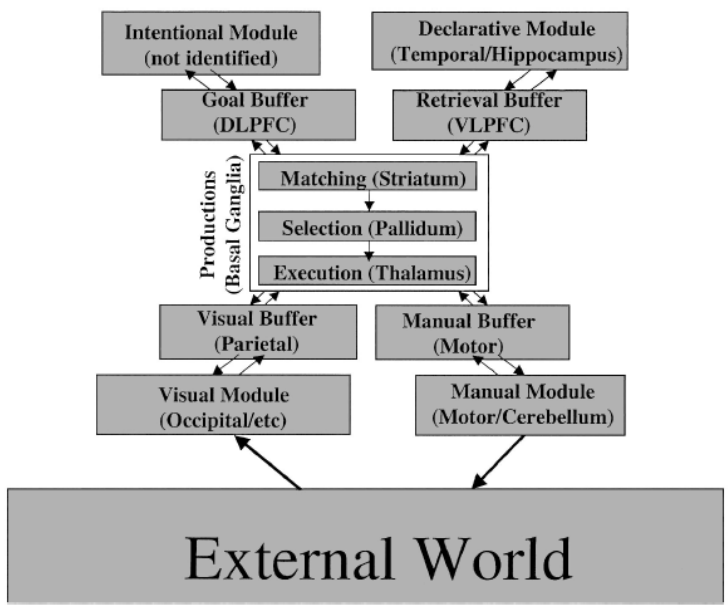

Figure 2.2. The ACT-R cognitive architecture. From Anderson et al (2004).

The four modules found within ACT-R (Fig. 2.2) each serve a specific function in order to produce a flow of information. The visual module examines identities and positions of objects in a model environment. The manual module controls and processes the movements of the hands and mouth (for the purposes of speech). The declarative module handles long term memory, and information learned by the

model. The memories of the model are stored in what are known as chunks, or blocks of multiple pieces of information used to associate two concepts (e.g., associating the concepts of leaf and green). Finally, the intentional module keeps track of the intentions and goals of the cognitive model. A central system, known as the production system, coordinates the information being processed by all of these systems mainly by passing the information that goes into and out of each module (Anderson et al., 2004).

The primary modes of control for the model are done through what is known as a production system. The production system is based on production rules, which are if-then like statements with conditions and outcomes. One simple example of such a production rule could be described by the statement "If there is a big red circle with the word start on it, then press it". These production rules act as sets of instructions for the model, which describe what choices the model has and what actions it can take (Anderson et al., 2004).

2.2.2.3 Executive-Process/Interactive Control

The EPIC (Executive-Process/Interactive Control) cognitive architecture, developed by Kieras et al. (1997,) has a lot of the same basic principles as ACT-R, though it is more of a physical architecture. The physical aspect is exemplified by the fact that EPIC has several options for physical output and control (e.g., pressing button, saying words, moving a joystick), and each of these alternatives has their own dedicated module. Thus, for the purposes of modeling tasks that use physical outputs, EPIC is more comprehensive. Another prominent feature that EPIC offers is the condensing of environmental input (e.g., visual, auditory), memory, and production rule memory into one processor. This pushes the system to interact as one whole object, rather than having to pass information to and from different modules. In addition, EPIC runs cognitive processes in parallel, which allows for modeling parallel thinking (Kieras & Meyer, 1996). An overall diagram of the EPIC architecture can be seen in Figure 1.1 (Anderson et al., 2004; Kieras & Meyer, 1997).

Though the motor processor is shown in one block, it takes input from the multiple motor processing modules available within the system (Kieras & Meyer, 1996). The central control unit combined with

these motor modules creates a high utility framework for the integration of perceptual and motor behavior. Another feature lending to a unified system in EPIC is portrayed by task environment being contained in the architecture itself (Kieras & Meyer, 1996). Most other environments must operate with an external program that simulates the task. With all these options, it is important to note that, in general, the production rules and memory of the architecture are significantly less complex than others, and this allows for less variation in higher cognitive processes with increased faculty for physical actions.

2.3 Motor Control

Ideally, it is evident by this point that this dissertation examines how sleep deprivation affects tasks, and how current models work to estimate task effectiveness. However, since the goal is to examine specific tasks, there is a need to understand how to evaluate the performance of those tasks. As stated in Aim 2, the goal is to evaluate how sleep deprivation affects motor control tasks, thus this section will discuss how one would accomplish that, as well as explore the relevant literature surrounding it.

One method for evaluating performance of a motor control task is Fitts' Law. Fitts' Law describes a psychological model of human performance, and has been used to provide information regarding how the human psychomotor system processes targeting tasks (Paul M Fitts, 1954; P.M. Fitts & Posner, 1967; Jagacinski, Repperger, Moran, Ward, & Glass, 1980; Kieras, Wood, & Meyer, 1997; MacKenzie, Kauppinen, & Silfverberg, 2001; Mottet, Guiard, Ferrand, & Bootsma, 2001; Zaal & Thelen, 2005).

One example of such study is done by Mackenzie et al. 2012., which examined a tilt-based interface using a tablet. The study had sixteen participants perform a targeting task to determine if tilting as an input method would conform to Fitts' law. The targeting task consisted of a simplified "ball maze" task, which involved rolling a ball into a hole by tilting the surface holding the ball. The study found that, like many other human-computer interactions, tilting did in fact conform to Fitts' law (MacKenzie & Teather, 2012).

Fitts' Law describes the time it takes for a person to "hit" a target by any method. The definition of hitting a target is very broad and includes, but is not limited to, touching it with a pen, hitting with a projectile object, or in a virtual setting such as a video game (MacKenzie, 1992). Fitts' law has been shown to apply in a variety of situations including those listed above. Along with simple accuracy measurements, Fitts' law is a standard for examining human performance, especially in the field of human computer interaction (Kim, 2015; MacKenzie, 1992). One of the factors in using Fitts' law is the task difficulty, referred to as index of difficulty (ID), which is shown in Eq. 2.2, and is derived based on the movement distance required to move to a target and the width of the target (Paul M Fitts, 1954; MacKenzie, 1992). The index of difficultly is defined in bits/sec.

$$\text{Index of Difficulty} = \text{Log}_2(1 + \frac{\text{Distance}}{\text{Width}}) \qquad \textbf{(2.2)}$$

Fitts' Law notes that the relationship between the ID and the average movement time required for a task is linear, thus, the formula describing Fitts' law, shown in Eq. 2.3, defines the linear relationship of movement time to the ID.

$$\text{Movement Time} = a + b * \text{ID} \qquad \textbf{(2.3)}$$

The variable *a* is the time intercept, or the minimum amount of time required for completing a task. The variable *b* is the slope, or the increase in movement time, as the difficulty of a task increases. The inverse of the slope is the index of performance (IP), which, represents a quantitative value for throughput. IP and ID work as two ways to describe movement times, as ID accounts for task difficulty, while IP accounts for the ability to perform with increasing difficulty.

In addition to the performance parameters based on Fitts' law, alternative measures of accuracy are available for such experiments (Williams et al., 1959). There are a number of ways to examine targeting tasks, such as hit rate and/or misses. However, many tasks will continue until a target is reached, especially those performed in the field of HCI. For these types of tasks there are alternative error

measurements that can be used, such as movement variability, movement error, and number of target exits and reentries.

To examine these accuracy measures it is first important to understand the task. In a movement task in which one would roll a ball into a target, similar to the task presented in the study described by Mackenzie et al., 2012, there are three parts of the path to examine. The first is the ideal path, the most direct path to hit the target. The second is the actual path that a person takes, from this point we will refer to the distance between any point on the actual path and the ideal path as y_i. Finally, there is the average path of the actual path that a person took, and we will refer to the distance between the average path and the ideal path as $\overline{y}$ (Williams et al., 1959). A diagram of these three paths and the two values mentioned are shown in Figure 2.3.

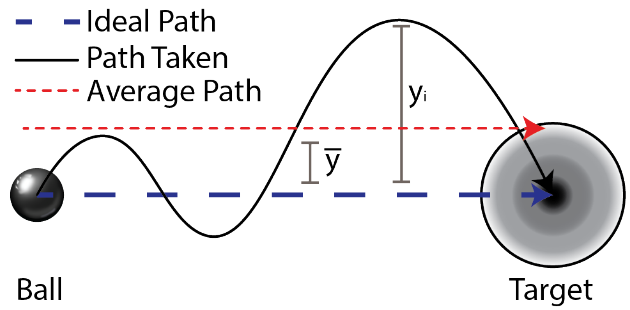

Figure 2.3. Targeting Task Paths.

Paths described in a targeting task, where the distance between the ideal path and actual path is defined as y_i and the distance between the average path and the ideal path is defined as $\boldsymbol{\overline{y}}$.

Movement variability, shown in Eq. 2.4, examines the extent to which the movement path taken by a person lies along a mean line parallel to that of the original task axis. This is examined by looking at each sample point's distance from the ideal path (y_i) and taking the standard deviation of the sample differences from the ideal path ($\overline{y}$) for every X ms or X mm of which there are n points (H. L. Williams et al., 1959).

$$MV = \sqrt{\frac{\sum (y_i - \overline{y})^2}{n-1}} \quad \textbf{(2.4)}$$

Movement error can be described as the average deviation of n sample point's distance from the ideal path (y_i) (H. L. Williams et al., 1959). Movement error, unlike movement variability, compares the movement of a person's path directly to the ideal task axis. The calculation for movement error was done using the same calculation in Mackenzie et al., 2012, and is shown in Eq. 2.5.

$$ME = \frac{\sum (y_i)}{n} \quad \textbf{(2.5)}$$

Finally, for a task in which a person cannot miss a target, it is possible to examine the number of target reentries, if the task has a dwell time. A dwell time, in this particular case, refers to a period during which a target must remain hit (i.e., a person holds their finger on a target) before it is fully considered acquired. Instead of looking at how often the target was missed, it is possible to see how many times a person exited and subsequently reentered the target after reaching it.

Earlier, there was a description of task using tilt-based interaction, which leads into the next point for this dissertation, that the motor tasks specifically will look at tilt-based interactions. While tilting has been examined as a user interface input since the 1990's, the technology was not prevalent (Harrison, Fishkin, Gujar, Mochon, & Want, 1998). The increasing availability of cheap embedded parts has allowed hardware, such as a gyro sensors, to become common in mobile devices (Lane et al., 2010). The implications of tilt as an input method for HCI are far reaching. Tilting goes well beyond the other control systems in that it allows for a complex input parameter in a 3-dimensional space. This can be compared to a mouse or a touch screen, which only allows 2-dimensional inputs (MacKenzie & Teather, 2012). Tilt devices can also be used in situations where the use of a computerized system is extremely advantageous, but there are physical limitations associated with the task that constrain the use of other devices, such as in extreme environments. Some major examples of this include underwater diving or working in cold

environments, where the worker's hands are frequently covered by gloves or other substances, and using a touch screen or pressing buttons becomes difficult. With the advancement of tilt based interaction, it is becoming more important to characterize and quantify a person's performance and accuracy using such an interface.

2.4 Cognitive Tasks

2.4.1 Attention

Vigilant attention, or attention requiring sustained concentration, is used as the standard for measuring performance during sleep deprivation (Lim & Dinges, 2010). The most common method for measuring vigilant attention is the psychomotor vigilance test (PVT), which measures a person's reaction time to the presentation of a concurrent stimuli (Lim & Dinges, 2008). PVT has been used in the past to examine the effects of sleep deprivation, showed a correlation with both lack of sleep as well as with the circadian rhythm periodicity (Lim & Dinges, 2008). It has been proven to work very well as a performance measure for many practical applications (e.g. driving) (Dinges, 1995).

The effects of sleep deprivation on vigilant attention have been shown to affect people in various aspects of performance. The first and most obvious effect is a general slowing in reaction times. Results with increased errors of omission and commission have also been observed (Lim & Dinges, 2008). Frequently, these errors were exemplified as lapses, or extended periods of inactivity following a stimulus, normally lasting 500 to 1000ms (Lim & Dinges, 2008). With so much research and concrete findings surrounding the PVT, we used it as a baseline from which to compare other performance measures, such as memory and the ability to perform visual search.

2.4.2 Working Memory

Active recall from working memory (WM) is a commonly used cognitive measure as it can predict performance in such a large array of tasks (Kane, Conway, Miura, & Colflesh, 2007). A popular procedure that is used to measure WM is known as the n-back task (Kane, Conway, Miura, & Colflesh,

2007). Auditory based n-back tasks have shown a pattern of activation in the prefrontal cortex, which allows us to examine n-back in the context of the neuropsychological hypothesis for performance decrement during sleep deprivation (Owen, McMillan, Laird, & Bullmore, 2005; Rodriguez-Jimenez et al., 2009).

The n-back task consists of presenting stimuli and having a person recall whether each stimulus in a sequence matches, in duration, one that was previously presented. This type of task allows for cognitive performance measurement due to its requirements for participants to frequently update, rehearse, and respond to information presented to them (Jaeggi, Buschkuehl, Perrig, & Meier, 2010).

2.4.3 Visual Search

The capacity to perform visual search tasks is necessary for the completion of many vigilance based tasks. Evaluation of visual performance, in the form of reaction time and accuracy, has been used as a primary measure to examine targeting tasks (Najemnik & Geisler, 2005; Perrott, Saberi, Brown, & Strybel, 1990). With so many environmental applications, visual search acts as a very powerful performance measure that has a low barrier to entry for making real world comparisons.

Visual search can be looked at as a complex measurement, since a visual search task includes many performance components that weave together, including visual thresholds, search strategies, and visual occlusion (Remington, 1980; A. Williams & Davids, 1998). Visual search is an actively studied topic and is frequently modeled to predict experimental results (Wolfe, 1994). Being able to predict human performance during visual search is a complex problem, since there are so many factors to consider. Modeling and examining visual search performance has been a tough challenge to tackle due to massively parallel processing required for various visual features, such as color, motion, depth cues, and more (Wolfe, 1994). With the limited attentional resources of humans, predicting expectations for the outputs of visual search is very dependent on context and stimuli.

The guided search model is one of many that attempt to incorporate various features from a stimulus screen such as visual parameters and relative positioning. The complexity of the problem has driven research to include modeling (Wolfe, 1994). There has been little to no research into the effects of sleep deprivation on visual search, requiring a deeper investigation into the subject.

Chapter 3: Modeling and Tasking

3.1 Introduction

In this Chapter, the goals cited in Aim one of this dissertation will be approached. These goals include design and modeling of experimental tasks during sleep deprivation. The primary focus of this Chapter will be on modeling itself, and while the basics of the types of task to approach will be discussed, they will be discussed in terms of the current mathematical models and cognitive architectures that are used as predictive tools.

3.2 Motor Control Task

3.2.1 Model

As stated before in Chapter 2, the motor control task was focused on examining specifically tilt-based interaction. Thus, for designing the model for the motor control task, the task design was inspired by the tilt-based interaction done by Mackenzie et al., 2012. As a reminder, this task involved tilting a surface to roll a ball into a hole.

Although ACT-R has been used in modeling tasks such as this in the past, its reduced capacity to implement tilt-based control makes it ill-suited for the task used here. Therefore, the EPIC cognitive architecture was used for modeling the motor control task (Gunzelmann et al., 2009; Kieras & Meyer, 1997). More precisely, the EPIC architecture was chosen due to its capacity to model gyroscopic tasks (MacKenzie & Teather, 2012). Once the task was modeled, variables that define performance-based parameters in EPIC could be edited similarly to those used the ACT-R architecture in previous studies (Gunzelmann et al., 2009). The implications should transfer to EPIC fairly easily because ACT-R's perceptual-motor capabilities are based on EPIC's.

The task was designed to have a ball start in the center of the surface, with a target that could appear randomly in one of 16 positions around the screen and in two different sizes. The task environment and production rules were prepared within the EPIC cognitive architecture. A screen shot of the task

environment is shown in Figure 3.1, where an EPIC model can actively perform the tilt task. The control of the simulation was based on Fitts' Law of movement, as it has been shown that gyroscopic control tasks on a mobile device adhere to Fitts' model (MacKenzie & Teather, 2012). Thus, task completion time was based on the parameters of time intercept and index of performance as well as the properties of the task.

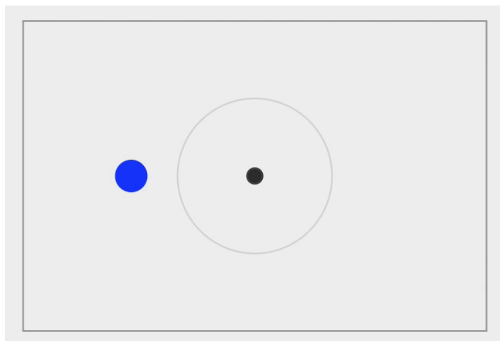

Figure 3.1. Model Tilt Task Environment.

The environment is simulated in the EPIC cognitive architecture's task environment, and includes the moving circle (small black circle), the target (medium blue circle), and the focus of the model's vision (grey circle).

The production rules had the model search for the target and then move the ball to the center. The values of time intercept and index of performance were changed over time, based on the performance effective changes described in the SAFTE model (Eq. 2.1), to simulate wakefulness between 6 and 24 hours awake (Gunzelmann et al., 2009; Hursh et al., 2004). The equation was implemented within the EPIC model using the parameters described by Hursh et al., 2004, and used to modify various performance variables within the EPIC architecture's manual aimed movement module. A diagram describing how the variables were changed and derived is shown in Figure 3.2 (Hursh et al., 2004).

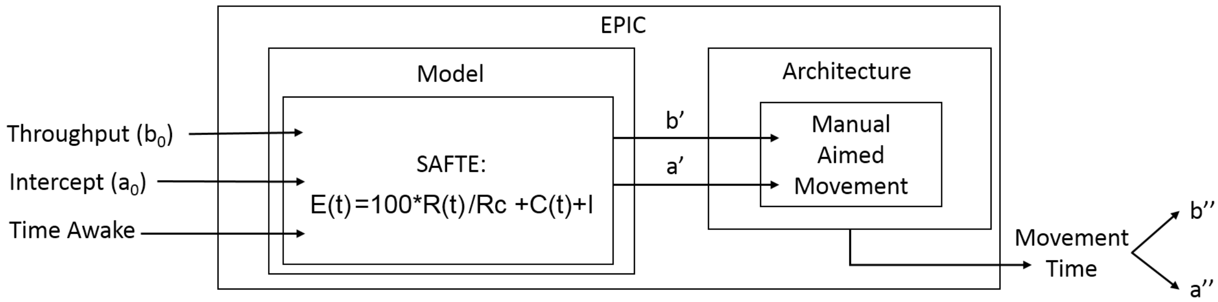

Figure 3.2. Manipulation of Performance Variables and Implementation of SAFTE within EPIC. The variables b_0, b', and b'' represent the input performance, the varied performance, and the derived parameters respectively. The variables a_0, a', and a'' represent the input intercept, the varied intercept, and the derived parameters respectively.

The index of performance parameter (*b*), used in movement calculations, was changed by decreasing the mean value based on time awake and on the effectiveness in the SAFTE model. The time intercept (*a*) parameter was also changed by increasing it based on the number of hours awake and effectiveness. These changes were designed in an effort to simulate lapses, and were meant to integrate with the EPIC library established by Kieras et al. (1996).

3.2.2 Results

The tilt task model was run to simulate 10 participants, each doing 240 actions or trials per 2-hour period, totaling to 2400 runs for every simulated period of wakefulness, or for a total of 24,000 runs. The variables examined from the model output were minimum task time, index of performance, and a noise gain value added to the Fitts' equation, as those values represented the ability of a person to perform a task. The results are shown in Figure 3.3. It was found that average predicted movement time increased over time $R^2 = 0.61$, $p < .01$. The average predicted throughput decreased over time, $R^2 = 0.57$, $p < .01$, while the average intercept increased over time, $R^2 = 0.60$, $p < .01$.

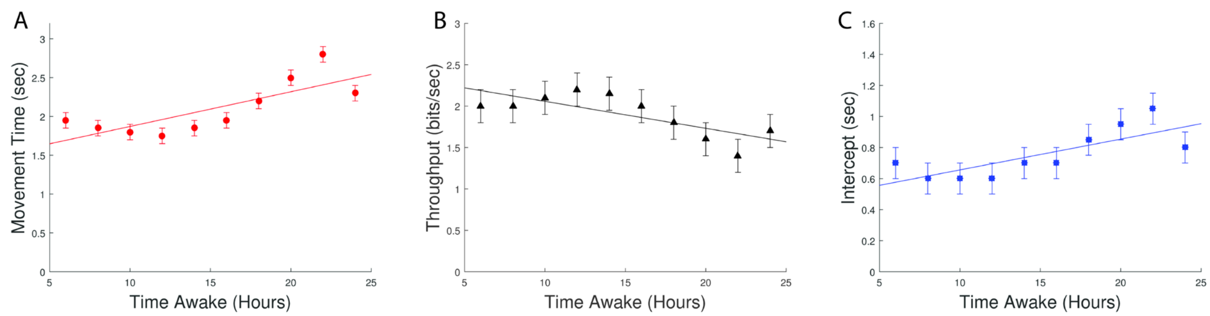

Figure 3.3. **Model Prediction of Tilt Task Performance.**

(A) The average predicted movement time (±SEM). (B) The average predicted throughput (±SEM). (C) The average predicted intercept (±SEM).

The performance data can also be visualized in the average performance lines for each time awake and difficulty curves, as shown in Figure 3.4, which displays how movement times change with the difficulty of the task. It is possible to see a general increase in movement time with hours awake along with an influence of circadian rhythm, especially after 24 hours in the model results. The performance lines and difficulty curves once again show the significant change in performance across various measures as time awake increases.

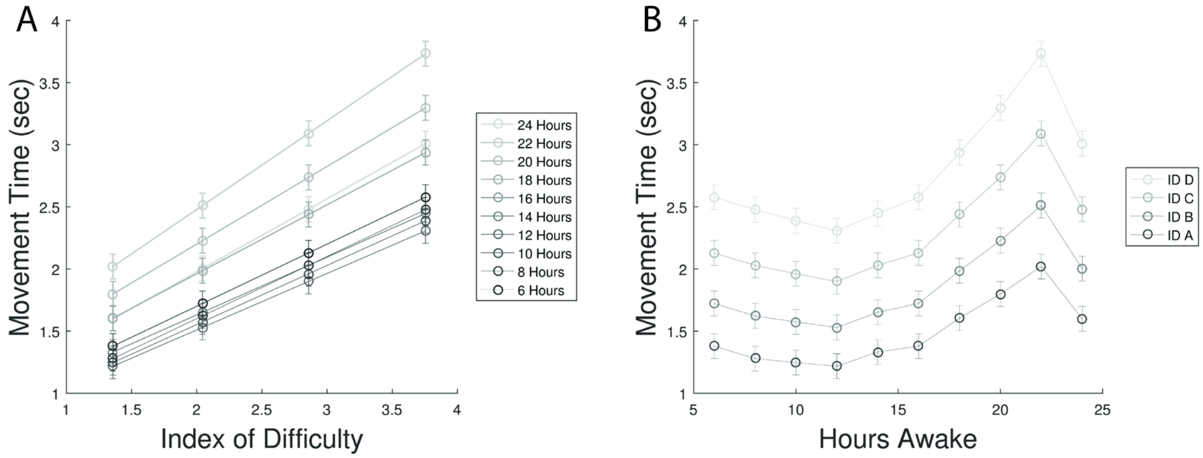

Figure 3.4. **Model Tilt Task Performance Lines.**

(A) Average predicted performance lines representing movement times for the levels of IDs and sleep deprivation (±SEM). (B) Predicted difficulty curves for each ID showing movement times for 4 IDs, with ID values A through D representing increasing task difficulty (±SEM).

The performance lines, based on the output of EPIC, show a culmination of the three metrics: movement time, throughput, and intercept. Unfortunately, the EPIC architecture does not have a way of simulating accuracy, and therefore only these predictions can be used.

3.3 Psychomotor Vigilance Task

3.3.1 Model

The PVT used in the model was based off a singular simple stimulus response task, frequently presented as a bright red stimulus on the screen (Dorrian et al., 2005). For the modeling of the task, the ACT-R cognitive architecture was chosen as similar task implementations have already been modeled with the use of ACT-R (Gunzelmann et al., 2009). As mentioned in Chapter 2, Gunzelmann et al., 2009 implemented a way to manipulate the performance of ACT-R models based on the SAFTE model. Since then, the ACT-R architecture has changed significantly, and although the previous work is no longer applicable with current implementations, it was used as the basis for how the PVT model was approached (Collins et al., 2016; Gunzelmann et al., 2009).

The stimulus chosen for the implementation of a PVT model within ACT-R was simply an "x" appearing on the screen. A visualization of the implementation of the model as the task environment, while a stimulus appears on screen, is shown in Figure 3.5.

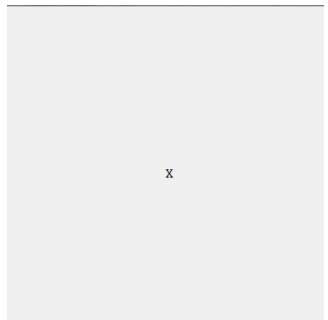

Figure 3.5. **The Task Environment of the ACT-R model during the PVT task.**

During the model run, the ACT-R model is designed to wait for a stimulus to appear, at which point the model will press the "x" key in response to the appearance of the stimulus. The approach taken to change the model behavior during sleep deprivation was more complex than simply affecting the reaction time of the model. Knowing that the primary affect that sleep deprivation causes is the appearance of lapses, a production rule, that represents lapsing, was included. In addition, since false starts are known to appear during the PVT, a production rule representing an early response was included as well. The design of the model began with a chance of having a false start before it saw the stimulus. Once the stimulus appeared, the model would either attend to the stimulus or miss it. If the model happened to miss, it would then either attend to the stimulus or lapse again after 50 ms. A basic action flowchart of this setup is shown in Figure 3.6. The 50 ms time period is derived from the cognitive cycle parameter, which is a basic parameter of ACT-R that describes how often a production rule can be run (Anderson et al., 2004; Bothell, 2004).

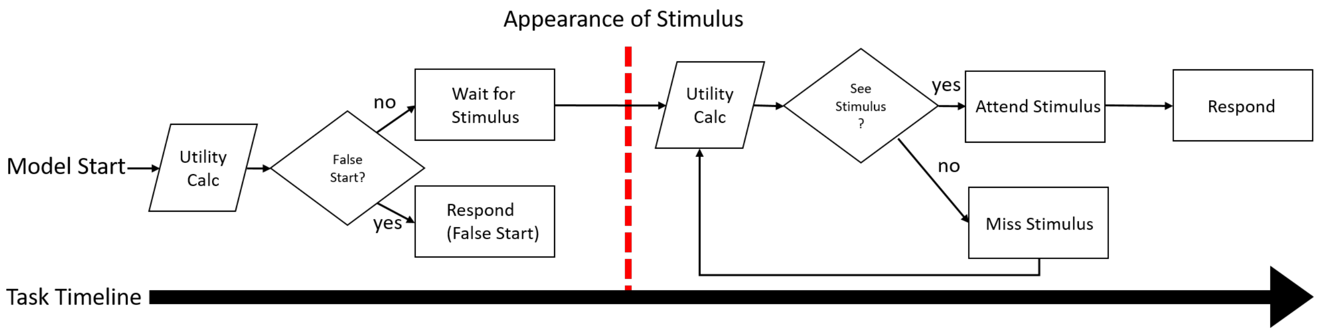

Figure 3.6. **A Task Timeline of the PVT run by the Cognitive Model implemented in ACT-R.** Rectangles represent production rules, and both utility calculations are modulated by task effectiveness.

Similar to the tilt task model, the SAFTE equation was implemented into ACT-R to assist with the prediction of performance with sleep deprivation. However, rather than implementing the SAFTE model by using it as a modulator of task effectiveness, or reaction time in this model, it was used to modulate the chances for which production rules fired, such as when the model would experience a false start, and when it would either miss or attend a stimulus. This implementation allows for the prediction of performance change during sleep deprivation. While it is a novel approach to the problem, this method was inspired by, and utilizes the principles initially set out by Gunzelmann et al., 2009.

$$U_i(n) = U_i(n-1) + \alpha[R_i(n) - U_i(n-1)] + \varepsilon \qquad \textbf{(3.1)}$$

This implementation was done by using the utility functionality of production rules in ACT-R, shown in Eq. 3.1 (Bothell, 2004). The primary purpose of the utility function is to determine which production rule will fire if multiple production rules are set to activate under the same set of conditions. The utility of a specific run, or "nth" usage of a production rule, known as $U_i(n)$, is determined by four factors: its previous utility $U_i(n-1)$, its learning rate α, its learning reward $R_i(n)$, and noise ε. The production rule with higher utility is always run, though the noise parameter introduces the probability that a production rule with lower utility can be run if the utilities are close enough. Each time a production rule is run, such as the miss stimuli rule, it can have its future utility changed so that a competing production rule, in this case attend stimuli, may be called instead. The implementation of false starts was

included by taking the task effectiveness parameter *E(t)* from the SAFTE equation (Eq. 2.1) and using it to modify the base utility of the false start production rule, increasing its likelihood of being called with less task effectiveness. The inclusion of lapses resulted by taking *E(t)*, and using that to modulate the learning reward, $R_i(n)$.

$$R'_i(n) = R_i(n) * [1 - E(t)] \tag{3.2}$$

The choice behind the utilization of the variables for utility and learning is twofold. First, this method is most akin to the prior implementation of the previous version of this model. Second, these two are the most direct parameters to examine to allow for the modeling of false starts and lapses (Gunzelmann et al., 2009). The new learning reward, $R'_i(n)$, implemented in the model is described in Eq. 3.2. This interaction causes the utility of not attending to the target to fall quickly with high task effectiveness, which will increase the probability that the model will find the stimulus quickly based on the noise for determining which activation rule fires. On the contrary, the utility for missing a target will fall slowly with low task effectiveness, causing a decreased chance for finding the target initially, but increasing chances of doing so over time.

3.3.2 Results

The PVT model was run to simulate 16 participants, as this number of participants was planned for the experiment discussed in Chapter 5. For each subject, the model simulated 100 stimuli per 2-hour period, totaling to 1000 actions for every simulated period of wakefulness or 10,000 total runs.

During the psychomotor vigilance task, four parameters were examined: minor lapses (>500ms), major lapses (>1000ms), average reaction time, and false starts (Dorrian et al., 2005). For each participant, the values were normalized to have unit variance. The model did not predict any major lapses over the expected time period. The averages of the normalized measures of these variables were taken and are shown in Figure 3.7.

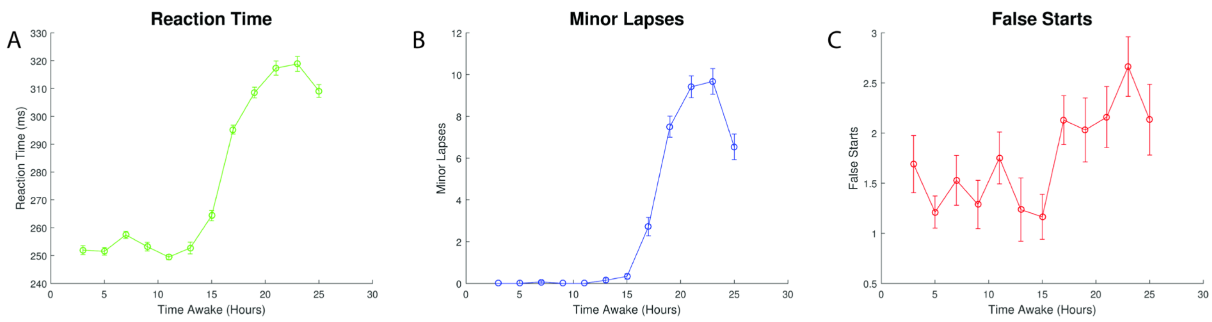

Figure 3.7. **Model Predictions of PVT Performance.**

(A) Reaction Time (±SEM) against time awake. (B) Minor Lapses (±SEM) against time awake. (C) False starts (±SEM) against time awake.

When examining the reaction time, there was a statistically significant increase starting at 17 hours awake and continued increasing until it reached a peak at 23 hours awake, with a peak difference of 69ms (28%), compared to reaction times at prior hours, where $F(11,180) = 245$, $p < .01$. Reaction time also had a drop after 25 hours awake. The number of minor lapses also had a significant increase starting at 17 hours awake and continued increasing until it reached a peak at 23 hours awake and also followed by a drop in the number of lapses at 25 hours awake, $F(11,180) = 129$, $p < .01$. The maximum difference in minor lapses for the model was 9.7 lapses. For false starts, there was a significant increase after 23 hours awake, of as much as 1.5 (128%) lapses, compared to the earlier hours $F(11,180) = 3$, $p < .01$. False starts had a drop after 25 hours awake, which was no longer significantly different from trials prior to 23 hours. The drop in metrics after 25 hours for all three results are primarily due to the circadian recovery that resulted from the SAFTE implementation.

3.4 Working Memory Task

3.4.1 Model

The task design for working memory, when developing with the model, was that of an n-back task. As stated previously, the n-back task is commonly used for the measurement of working memory (Jaeggi et al., 2010). An auditory n-back task was designed, in which the model would be presented tones

of different durations from which to compare. The n-back task would also have versions with and without distractors. This version would provide distraction by modulating the frequencies of the presented tones. The model was designed with the use of the ACT-R cognitive architecture because this architecture is more suitable for cognitive-based tasks.

The task presented within the architecture consisted of presenting a tone every two seconds with a randomized duration or either 100 or 200 milliseconds. The designed model was presented a 1-back task and a 2-back task. For the 1-back task, the model compared each tone with one presented right before it and responded by pressing "s" if the two tones are the same duration and "d" if they are different. The same procedure was followed for the 2-back task, except that the model compared each tone with the tone two back.

With the intention of examining how working memory is affected by sleep deprivation, the focus for implementing the SAFTE model is on the recall of memories. For ACT-R, memories are stored in units known as "chunks", which act as units that hold information that can be recalled by the model (Anderson et al., 2004). These chunks have a property known as activation, which, affects how well they can be recalled. The activation parameter determines which chunk is chosen if multiple chunks satisfy the description of the information that the model is trying to recall. If the activation parameter is too low, the chunk cannot be brought up all. This particular instance is the ACT-R representation of forgetting.

$$A = B + S + P + \varepsilon \tag{3.3}$$

Whenever an attempt is made to retrieve a chunk, the activation parameter A is determined using the calculation described in Eq. 3.3. The calculation of A is based on 4 parameters: B, the base level activation; S, the spreading activation value, which reflects the effect that the contents of the buffers have on the retrieval process; P, the partial matching value, which describes the degree to which a chunk needs to match a retrieval request; and ε, the noise value for the activation. The modeling of sleep deprivation within the model was primarily carried out by modulating the activation parameter. The activation

parameter was changed by modulating the noise by *G*(1-E(t))* where *G* is a constant and *E(t)* is the task effectiveness from SAFTE (Eq. 2.1). This is done in order to modulate the ability of the model to recall a tone. The noise utility, or the noise added to the utility function shown in Eq. 3.1, is also modulated by *G*(1-E(t))* for this model. Modulating the utility noise allows the model to "accidentally" compare tone frequency rather than duration. An outline of the model, with its production rules, is shown in Figure 3.8.

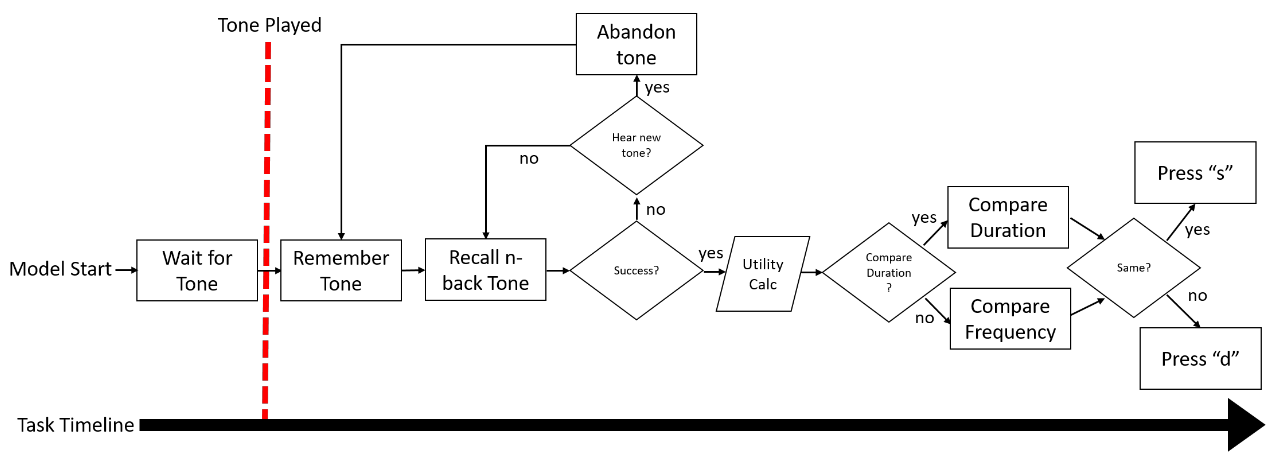

Figure 3.8. A Task Timeline of the N-Back Task in ACT-R.

Rectangles represent production rules, and both utility calculations as well as the ability to recall a tone are modulated by task effectiveness.

The tasks presented to the model are: 1-back 1-pitch, 2-back 1-pitch, and 1-back 4-pitch. The implementation of the model starts with the model waiting for a tone. The model then commits the tone to memory and attempts to recall the appropriate n-back tone for comparison. The modulation of the activation noise controls whether the correct tone, or any tone at all can be recalled, and the model continues until it recalls a tone or until the next one is played. The model then attempts to compare the current tone with the tone it recalled. Because the model is supposed to compare tone duration, the utility of the production rule for duration comparison is higher, however, the utility noise, modulated by *E(t)*, can cause the compare frequency rule to be fired instead. Finally, the model responds and presses a key based on whether or not the tones were considered the same.

3.4.2 Results

The n-back model was run to simulate 16 participants, just as the PVT model. Once again this was done to emulate the experiment that will be discussed in more detail in Chapter 5. For each participant, the model was run for 12 sessions. Each session was modeled at a different level of wakefulness, starting at 3 hours awake and incrementing until it reached 25 hours awake. During each session, the model performed 200 trials for each of the 1-back 1-pitch, 2-back 1-pitch, and 1-back 4-pitch tasks. This resulted in 9600 trials across all participants for every simulated period of wakefulness, or 115200 total trials of the model.

Regarding the n-back task model, performance was examined by measuring accuracy and misses for all tasks. Accuracy was determined by measuring how many tones were correctly responded to, and misses were determined by the number of tones that were not responded to before the next tone played. Misses were counted as incorrect in terms of accuracy. For all measures, average normalized values were examined. The results are shown in Figure 3.9.

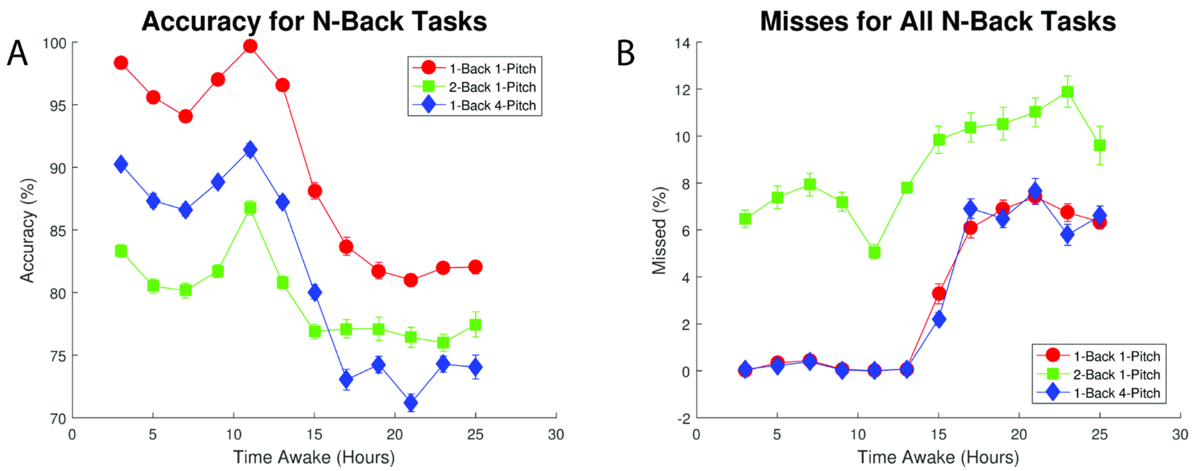

Figure 3.9. Model Predictions of N-Back Performance.

(A) N-back accuracy (±SEM) against time awake. (B) N-back misses (±SEM) against time awake.

Significant differences in model performance using a one-way ANOVA with Dunn-Sidak post-hoc analysis were examined. The 1-back 1-pitch task average accuracy had a significant decrease starting at 15 hours awake and continued to 21 hours awake, where $F(11,180) = 244.12$, $p < .01$, with no increase in accuracy after that time. Misses for 1-back 1-pitch had an increased value after 17 hours awake, and this continued increasing until 21 hours awake, $F(11,180) = 147.65$, $p < .01$. For the 2-back 1-pitch task, the average accuracy had a significant decrease after 17 hours awake, $F(11,180) = 23.22$, $p < .01$, with no increase in accuracy following that point. Misses for 2-back 1-pitch showed a significant increase after 17 hours, $F(11,180) = 14.39$, $p < .01$. Regarding the 1-back 4-pitch task, the average accuracy had a significant decrease after 17 hours awake $F(11,180) = 156.19$, $p < .01$, with no more significant changes after 19 hours. Misses for 1-back 4-pitch had a significant increase at approximately 17 hours, and an even greater difference at 19 hours $F(7,180) = 117.55$, $p = < .01$.

In addition, change in accuracies and miss rates across the different n-back tasks during hours awake was determined. When examining accuracy, 50% was chosen as the baseline, as that is the expected outcome from random choice. Based on this, accuracy decreased by 38% during 1-back 1-pitch, 29% during 2-back 1-pitch, and 49% during 1-back 4-pitch. Though there was less of an accuracy

decrease during the 2-back 1-pitch than during the 1-back 1-pitch, the overall accuracies of the 2-back 1-pitch were lower. Finally, there is an increase of the miss rate by 7.4% during 1-back 1-pitch, 6.8% during 2-back 1-pitch, and 7.6% during 1-back 4-pitch.

3.5 Visual Search Task

3.5.1 Model

The design of the visual search model was based off a visual search for finding ships appearing on the horizon. This design was based off of the game Dangerous Waters, initially developed by Sonalyst Systems, to serve as a platform for visual search tasks for Navy operations. A screenshot of the game is shown in Figure 3.10.

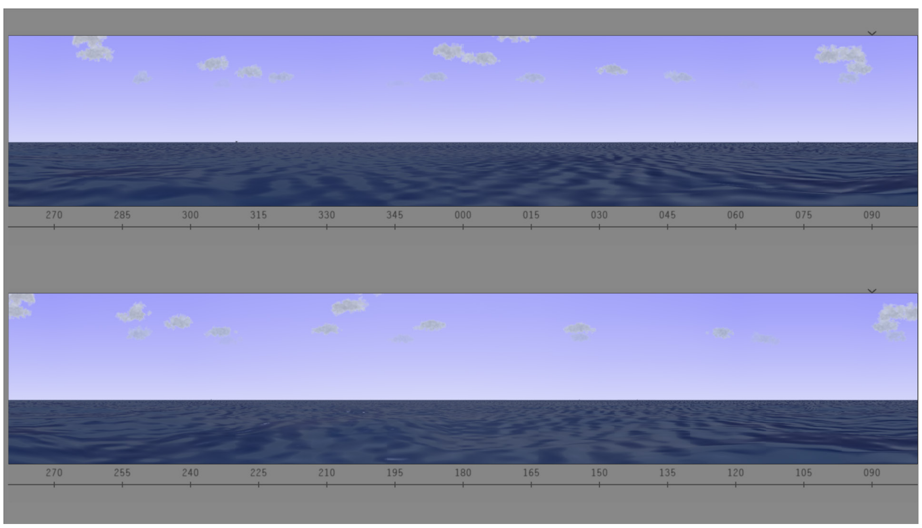

Figure 3.10. Screen Presented to the Model during the Visual Search Task.

The model for the visual search task was developed in collaboration with Dr. Michael Schoelles from Rensselaer Polytechnic Institute. The model was developed to take input from an external environment, which provides stimuli to the model, unlike other models which had constructed their environments within ACT-R. The production rules for this model are detailed in Figure 3.11. The rules guide the model to scan different areas of the map, looking for targets to appear. Once a target is found, the model will press space and speak the bearing at which the target appeared.

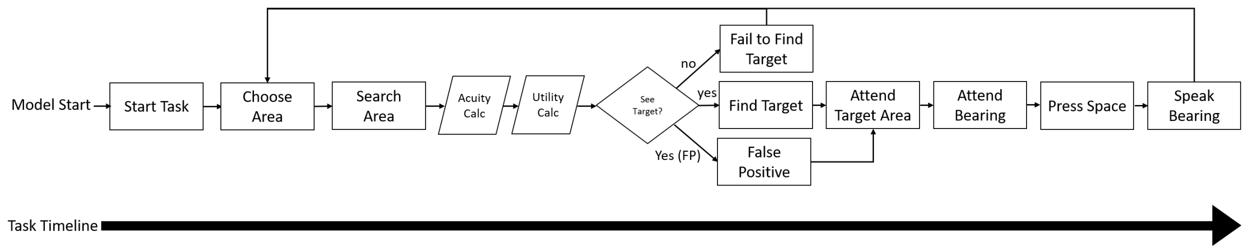

Figure 3.11. A Task Timeline of the Visual Search Task in ACT-R.

Rectangles represent production rules, and both acuity and utility calculations are modulated by task effectiveness.

For this particular model, an acuity module was added to simulate the acuity, or capacity of the model to see a target during the task. This addition was needed because the model took input from an external environment, and the ACT-R architecture does not have all the necessary parameters for that environment. The implemented acuity model is based on the model representation of acuity within the environment of cognitive architectures (Nyamsuren & Taatgen, 2013). As shown in the task timeline, an acuity calculation is performed prior to finding a target to determine if the model is able to see a target on screen. The acuity calculation, which determines a threshold for detection of a target, is shown in Eq. 3.4. A target is considered visible if its size on the screen is above the threshold.

$$\text{Threshold} = a * e^2 - b * e + \varepsilon \qquad \textbf{(3.4)}$$

The *e* variable represents the eccentricity of an object, which is a function of the size of an object and the angle from which it is viewed. The parameters *a* and b represent weighting functions that balance features such as color and shape. Finally, *ε* represents the noise of the threshold.

The implementation for sleep deprivation was done in two parts. In a similar manner to the prior two models, this two-step implementation was used to represent the two different types of errors experienced by the person or models. The first type of error, an error of omission, describes how the

model will be able to see or miss a target. This type of error leads to two primary effects on performance. The first effect is increased time to spot a target due to its being missed during the initial search. The other effect is missing the target entirely. This type of error was modeled by taking the acuity from Eq. 3.4 and modulating the a and ε parameters. The a parameter was scaled by the inverse of task effectiveness from the SAFTE equation, while the noise parameter, ε, was scaled by 1-$E(t)$ to represent increased noise and inconsistency in visual search. This implementation attempts to represent lapses during visual search as a result of decreased attention. In other words, the subject may not perceive a target even when looking at the target area (Kendall, Kautz, Russo, & Killgore, 2006).

The second type of error, or error of commission, is the detection of a false positive during the visual search. This type of error was included in this model in the same manner as it was included in the PVT model for false starts. An alternate production rule for seeing a target on the screen was included with lower utility. The utility noise, described in Eq. 3.1, was varied with task effectiveness, $E(t)$, allowing the production rule to be called "in error" and having the model respond to a false alarm.

3.5.2 Results

The visual search model was run to simulate 16 participants, just as in the previous two models. Again, model runs were performed the same way. For each participant, the model was run for 12 sessions, each representing a different time awake over a 25-hour period, starting at 3 hours. During each session, which ran for approximately 20 minutes, 5 targets appeared on the screen and increased in size over time. This resulted in totaling 80 targets across all participants for every simulated period of wakefulness, or 960 ships during all runs of the model.

During visual search, three primary measures were examined: time to detect targets, number of targets missed, and number of extra targets, or false positives, reported. The average normalized values of these measures are shown in Figure 3.12.

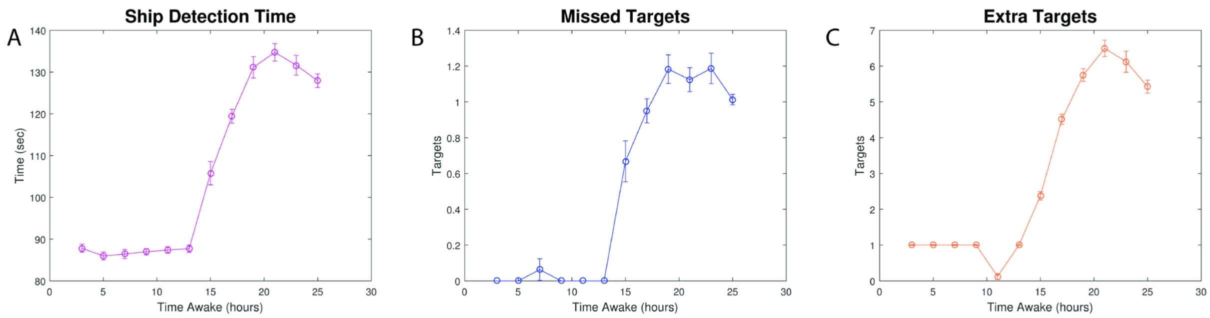

Figure 3.12. Model Predictions for Visual Search Performance.

(A) Ship detection time (±SEM) against time awake. (B) Missed targets (±SEM) against time awake. (C) Extra targets reported (±SEM) against time awake.

The target detection time showed a difference of up to 49 sec (57%), with an increased search time predicted starting at 15 hours awake, with $F(11,180) = 150.75$, $p < .01$. The number of missed targets had a difference of up to 1.2 targets on average, with a significant increase starting at 15 hours awake, $F(11,180) = 71.3$, $p < .01$. The number of extra targets, with an average increase of 6.37, also showed after 15 hours awake, $F(11,180) = 247.8$, $p < .01$. For the results of the model run, there were similar significant increases in all measures starting at approximately 15 hours awake.

Chapter 4: Effects of 24 Hour Wakefulness on Motor Control

4.1 Introduction

In Chapter 3, a motor control task that consisted of tilt-based targeting activity was developed, and results on how the EPIC cognitive architecture performed that task with simulated sleep deprivation were obtained. To verify how well a model, which is based on current literature, predicts human performance under the effects of sleep deprivation, it was necessary to design a human analog to the simulated task. This leads into Aim 2, which is to quantify the effects of sleep deprivation on a subset of manual targeting tasks in humans. To accomplish the goals for Aim 2, a 24-hour-based sleep deprivation study, in which participants consistently performed the previously mentioned tilt-based task was designed and run. In this section, the specifics surrounding the design and implementation of the experiment will be described. This section will also include the results from the experiment as well as a discussion on how these results are interpreted and consequently used.

4.2 Methods

4.2.1 Experimental Population

The experiment was conducted with the assistance of 10 student volunteers from Worcester Polytechnic Institute. This research complied with the American Psychological Association Code of Ethics and was approved by the Institutional Review Board at Worcester Polytechnic Institute. Informed consent was obtained from each participant, and participants were compensated for their time. The population consisted of 5 male and 5 female participants between the ages of 22 and 32, none of whom had any extensive experience working with gyroscopic-based devices that used tilt-type control. Participants were instructed to refrain from intake of any stimulants and depressants for 48 hours prior to the start of the experimental procedure, and reported an average of 7.7 hours (SD = 0.72) of sleep each night for 7 nights prior to the start of the experiment.

4.2.2 Apparatus and Materials

The procedure was performed using a Samsung Galaxy Tab 10.1 running Google's 4.1 (Jellybean) operating system, shown in Figure 4.1. The task was done on a screen that was 14.23 cm by 21.35 cm (800 px by 1200 px) using a ball that was 0.36 cm (25 px) in diameter. The software was developed in Java using the Android SDK, while tilt control was implemented using the device's built in gyro sensor. Pitch and roll values were converted to tilt magnitude and direction. The task implemented in the software allowed a user to tilt the device to control a circle that was shown on screen. The goal of each individual task trial was to move a ball (represented by a small white circle) to a target (represented by a large cyan circle) location on the screen. The software provided the capacity to change task features including, but not limited to, the starting position of the ball, the effect of pitch and roll on the velocity and direction of the ball, and the size, position, and visual aspects of the targets. The application was set up to run multiple trials, in which a single trial consisted of a participant moving a circle on the screen of the device to the target by tilting it.

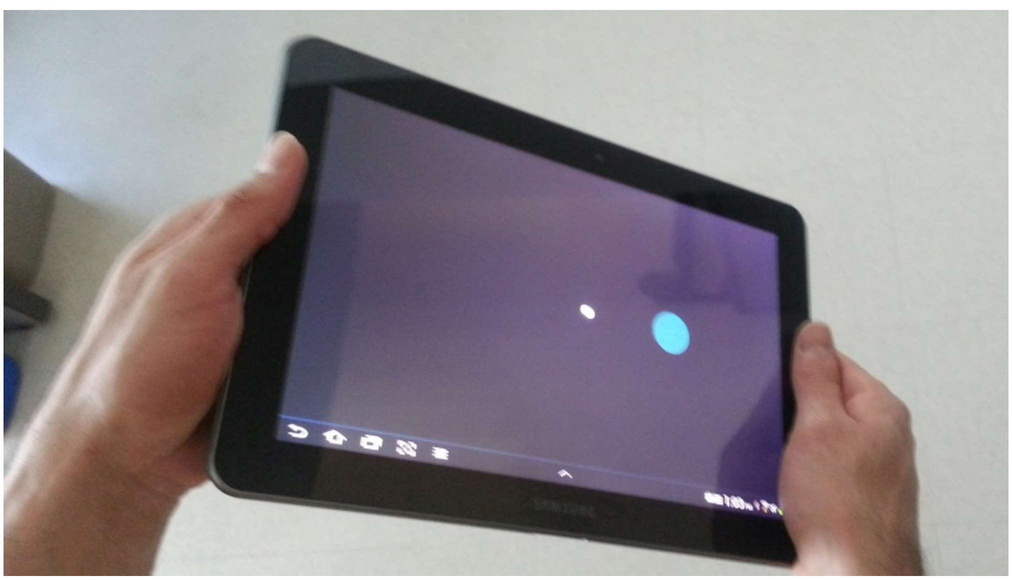

Figure 4.1. Tilt Task Platform, Samsung Galaxy Tab 10.1.

The ball (small white circle) in the center is moving to the target (large cyan circle) on the right.

The movement of the target was controlled by the pitch and roll values produced by the user when tilting the device. In addition, the movement of the ball was influenced by a gain parameter, which was determined empirically prior to the experiment to allow for adequate object manipulation and to control the speed of trials. The velocity of the ball along with the angle of movement was calculated using Eqs. 4.1 and 4.2, with pitch and roll in degrees from horizontal.

$$\text{Ball Velocity} = \text{gain} * \sqrt{\text{roll}^2 + \text{pitch}^2} \tag{4.1}$$

$$\text{Movement Angle (from horizonal)} = \arctan(\frac{\text{roll}}{\text{pitch}}) \tag{4.2}$$

The total number of trials in a set included 8 target positions, 1 movement gain, 1 ball size, 3 target sizes, 2 target distances, and 5 repetitions, for a total of 240 trials per set.

4.2.3 Design and Procedure

The dependent variables being examined were movement time, throughput, minimum movement time, movement variability, movement error, and target reentries. Table 4.1 shows the independent variables used. A set of 5 trial repetitions was included for each combination of independent variables. These combinations were all presented in a random order for all trials to eliminate sequence effects.

Table 4.1

Independent Variables and Task Values for the Tilt-Control Task

Independent Variables	Values
Target Position (° from horizontal)	0, 45, 90, 135, 180, 225, 270, 315
Movement Amplitude (cm (px))	2.22 (125), 4.45 (250)
Target Size (cm (px))	0.71 (40), 1.07(60), 1.79 (100)
Time Awake (hours)	6, 8, 10, 12, 14, 16, 18, 20, 22, 24

Participants were asked to refrain from consuming any stimulants and depressants for 48 hours prior to the start of the experiment. They were also advised to keep a normal and consistent sleep schedule for 1 week prior to the study. All participants submitted journals containing their sleeping and eating habits for the week before the experiment, which indicated compliance.

Each participant completed a set of 240 movements over a 20-to-30-minute session. A session occurred every two hours over a period of 24 hours. The time at which participants awoke during the day of the trial was monitored, and participants began the experiment four hours after waking up. The first two sessions were considered learning sessions and were not included in final data analysis. All subsequent trials were run every two hours, until the point at which the participant had remained awake for 24 hours.

Once the experiment began, a start screen was provided between each task. A trial began once a “begin” button appeared on the screen and was pressed by the participant. At this point, a target appeared on the screen, and the ball in the center moved as the device was tilted, shown in Figure 4.2. A target was considered acquired or successfully hit once the circle remained within the target area for a total of 500 ms.

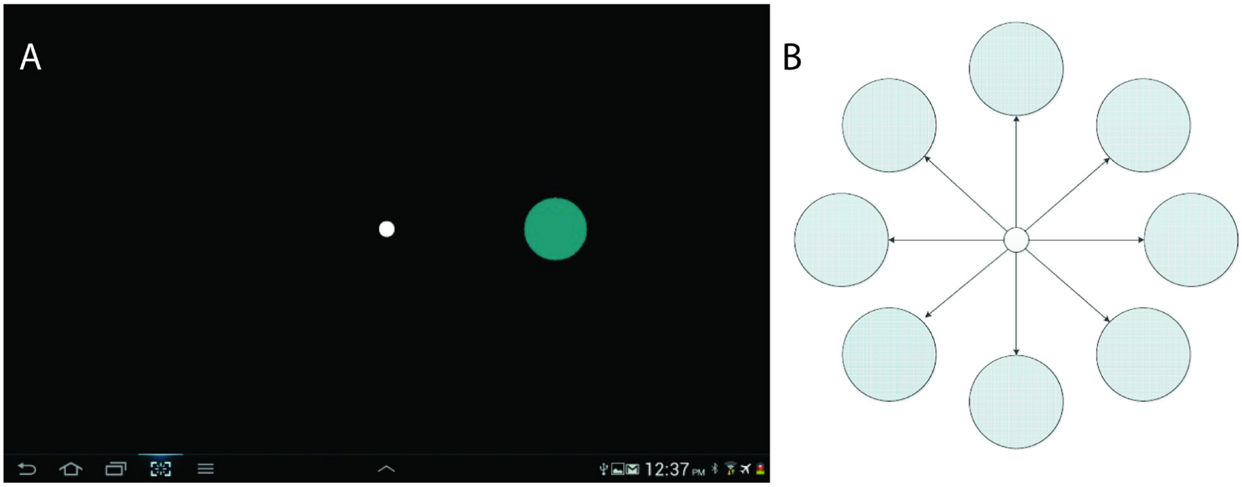

Figure 4.2. Tilt Task Interface.

(A) Screen with an active movement task where the current target (large cyan circle) is to the right of the controllable object or ball (small white circle). (B) The 8 relative positions of the targets that can appear for the tasks.

To use Fitts' law to measure performance, the index of difficulty for each task, the distance between the initial position of the ball and the target as well as the width of the target were calculated, based on Eq. 2.2. The width used is the effective width of the task (i.e., the difference in size between the target and the ball) rather than the width of the target itself. This was used because, the effective width estimates the target width focused on by the participant, and this been determined to be more accurate for the purposes of determining difficulty of a task (MacKenzie, 1992; MacKenzie & Teather, 2012). Once the index of difficulty was determined, each task was analyzed using Fitts' Law (Eq. 2.3).

4.3 Results

4.3.1 Experimental

The ten participants performed all of the movement tasks without data loss or having to restart. The R^2 between the index of difficulty and average movement time was computed for every set of 240 trials. For over 100 sets of trials (10 participants over 10 sessions), the average R^2 value was .99 with a standard deviation of 0.015. An example of this plot is shown in Figure 4.3, along with a graph depicting

the R^2 values for all 100 sets of trials showing a minimum of .91. This provides strong evidence that movements in the tasks examined in this study were in fact Fitts' movements.

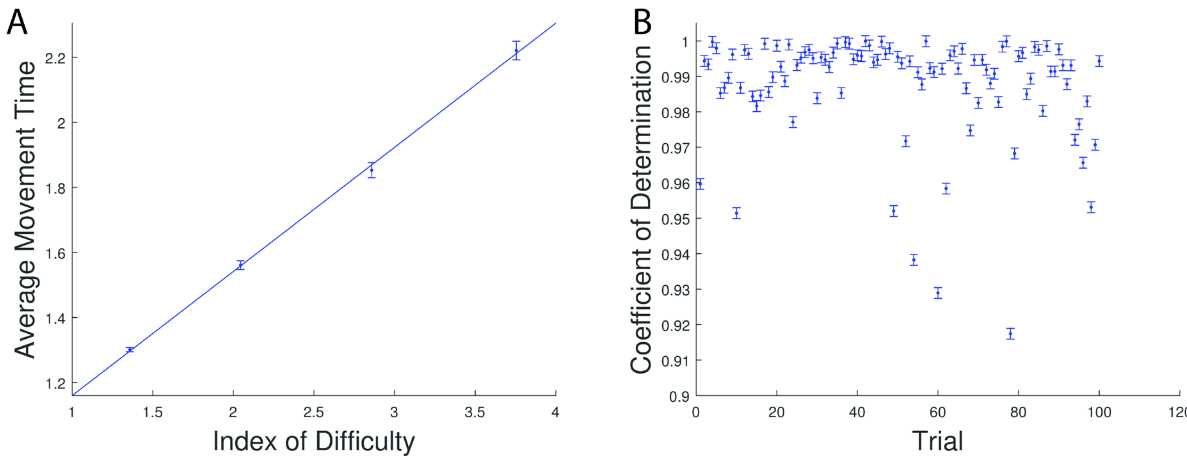

Figure 4.3. Fitts' Law Verification of Tilt Task.

(A) Example of average movement time over index of difficulty one participant. (±SEM). (B) R^2 values (±SEM) for all 100 trials sets (10 sets for 10 people), showing a minimum R^2 value of 0.91 for all participant data.

Data from the learning trials showed that performance followed the power law of learning and did not appear to be significantly affected by fatigue (Murre & Chessa, 2011). Figure 4.4 shows the average movement time, throughput, and intercept values across participants, taken for each set of trials, with error bars representing SEM. These values were then averaged and compared to hours awake. After examining overall trend using regression and performing a one-way ANOVA with Dunn-Sidak post-hoc to look for differences in performance between different levels of wakefulness, it was found that over the period of 24 hours, there was no reliable change in movement time $F(9,90) = 0.26$, $R^2 = .11$, $p = .35$. In addition, no reliable change in throughput was found $F(9,90) = .45$, $R^2 = .15$, $p = .26$. Finally no change was found with intercept $F(9,90) = .69$, $R^2 = 0.27$, $p = .12$. All results show no significant change in performance as time awake increases.

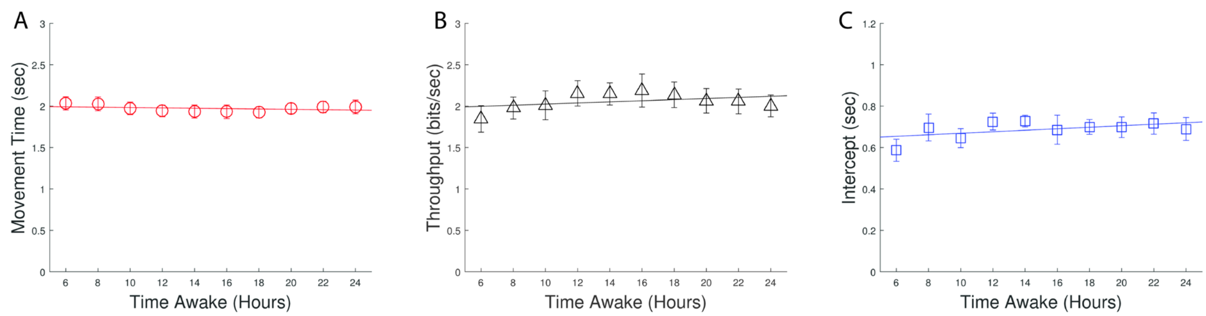

Figure 4.4. Human Performance for the Tilt Task.

(A) The average movement time across all participant (±SEM). (B) The average throughput across all participant data (±SEM). (C) The average intercept across all participant data (±SEM).

Average performance lines and difficulty curves, based on Eq. 2.2, are shown in Figure 4.5. The performance lines and difficulty curves once again show that the movement time, throughput, and intercept have no reliable changes as participants became more affected by sleep deprivation.

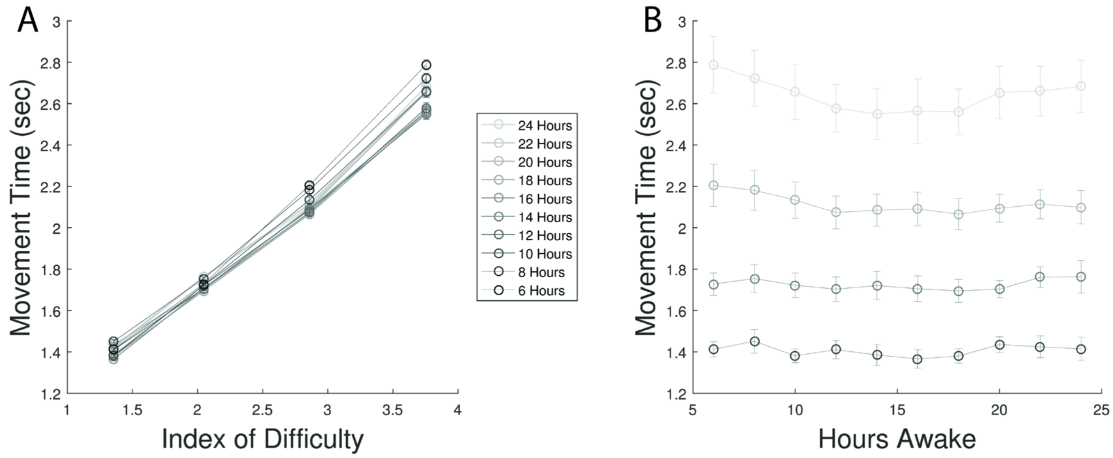

Figure 4.5. Human Tilt Task Performance Lines.

(A) Average performance lines representing movement times for various levels of sleep deprivation from participant data (±SEM). (B) Difficulty curves for each ID showing movement times for 4 IDs, with IDs A through D representing increasing task difficulty from participant data (±SEM).

Accuracy-based measurements were also taken in the form of movement variability, movement error, and movement reentries. Similar to the performance results, after regression and performing a one-way ANOVA with Dunn-Sidak post-hoc analysis, it was found that there was no noticeable change in movement variability with time awake: $F(9,90) = 0.15$, $R^2 = 0.15$, $p = .03$. There was also no noticeable change in movement error: $F(9,90) = 0.35$, $R^2 = 0.35$, $p = .01$, and no noticeable change in movement reentries $F(9,90) = 0.35$, $R^2 = 0.32$, $p = .01$. Figure 4.6 shows the data for all accuracy based measurements for all participants and difficulties. Note that, there are no corresponding predictions from the model.

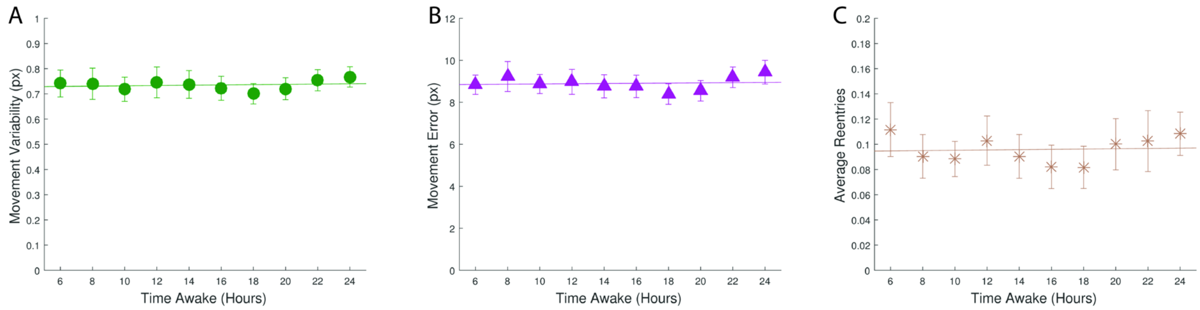

Figure 4.6. Human Accuracy Measures During Tilt Task.

(A) The average movement variability for all participants' data (±SEM). (B) The average movement error variance for all participants' data (±SEM). (C) The average number of reentries variance for all participants' data (±SEM).

4.3.2 Model Comparison

As described in Chapter 3, one of the goals of this research was to compare the results of the model with that of the human data to determine how well they matched. Here, we compare the results of performance decrement to see how well the model predicted human performance. As stated prior, the EPIC architecture did not have a system capable of handling accuracy during modeling, so these results will only examine the performance measures of movement time, throughput, and intercept. To visualize how the performance of the model and humans changed over time the results of both have been overlaid and are shown in Figure 4.7.

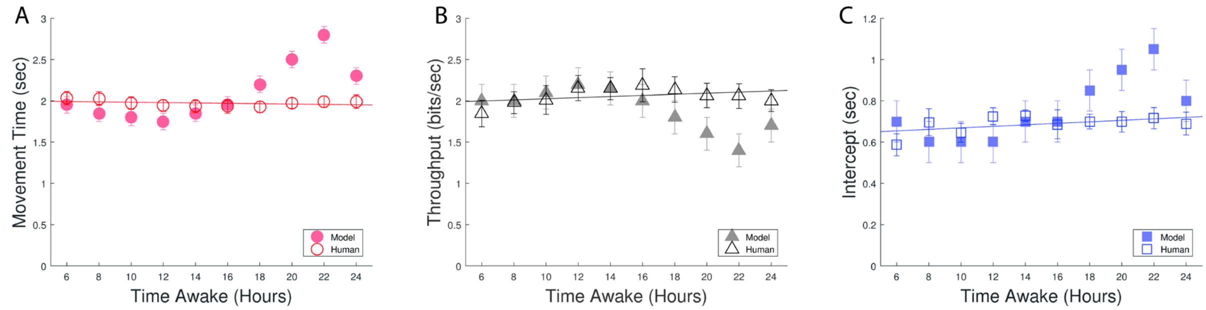

Figure 4.7. Tilt Performance Comparison between Human and Model.

(A) The average movement time across all participant data (open markers) and the model predication (solid markers) (±SEM) (B) The average throughput across all participant data (open markers) and the model prediction (solid markers) (±SEM). (C) The average intercept across all participant data (open markers) and the model prediction (solid markers) (±SEM).

Although the model showed significant changes in performance with sleep deprivation, and the results from the human did not, correlation analysis was done to search for any final predictive factors. The correlation plots for all performance measures are shown in Figure 4.8.

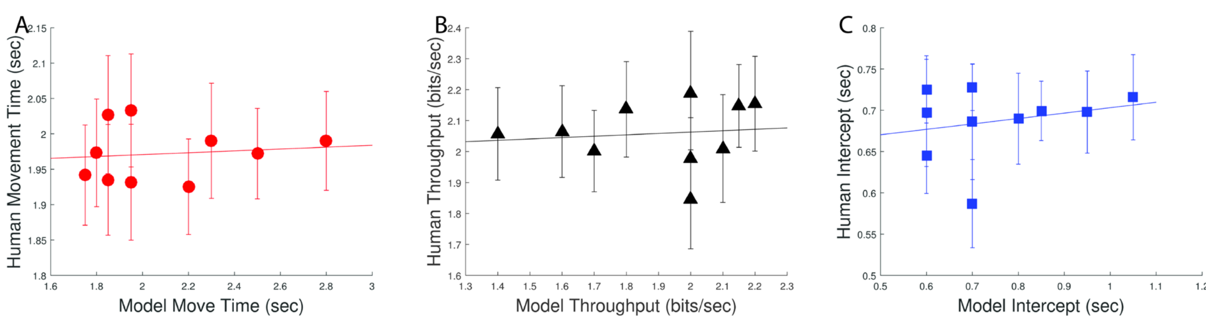

Figure 4.8. Tilt Task Performance Correlation between Human and Model.

(A) Movement time from model compared to movement times from human results (±SEM) (A) Throughput from model compared to throughput from human results (±SEM). (C) Intercept from model compared to Intercept from human results (±SEM).

The correlation between the model and human results for movement time was .12 (p = .75), which was not significant. The comparison for throughput showed an r value of .11 (p = .76), again

showing no significant correlation. Finally, the correlation coefficient for intercept was .24 (p = .5), once again showing no significant relation. Thus, for all comparison measures, there was no significant correlation between the model and the human.

4.4 Discussion

4.4.1 Experimental Analysis

The primary purpose of this study was toward the fulfillment of Aim 2, that is to determine how human performance on a tilt-based interface task changed over the course of 24 hours awake. The final results provided evidence that the average movement time did not change over the period of sleep deprivation, which is not consistent with previous sleep deprivation studies that employed the use of alternative tasks to measure performance, such as psychomotor vigilance (Lim & Dinges, 2010).

When examining the difference between this and other tasks, it is reasonable to analyze aspects of this experiment and task that may have caused these differences. It is unlikely that learning effects had a significant influence on the results, as any effects would have been detected in early trials, and would need to identically keep pace with fatigue throughout the entire experiment. The sample size could also be increased; however, the low standard error empirically suggests that increasing the sample size to the point of finding significance will lead to low ecological relevance.

The examination of the index of performance in this study represents the throughput of information being processed by the participants. This measurement represents the increase in time necessary to complete a task as the difficulty increases. It was found that over a 24-hour period of sleep deprivation, the throughput of participants did not significantly change between trials, meaning that there was no increase in time to complete tasks of different difficulties as the participants became more sleep deprived. While this particular experiment did not show any change in throughput, it is possible that a difference in performance could be found if certain parameters of the experiment were altered, such as the hours awake, velocity gain, or the range of index of difficulties.

The intercept, representing the minimum time required to complete a task, also did not significantly change with increasing sleep deprivation. This value is affected by many of the same parameters that affect throughput. However, unlike throughput, the intercept could also have been affected by the start screen interface, which removed some of the psychomotor vigilance elements of this task. A button between each trial was meant to create consistency between trials as well as to provide definitive start and stop times for each task that were controlled by the participant. However, because the start time was defined by the participants, one opportunity for lapses, a primary cause of decreased reaction time in other studies, was removed. A different type of interface or pause between trials could show an increase in intercept with increasing hours awake. However, this would change the type of task (i.e., make it a vigilance-based task).

The accuracy of the tasks was also found not to have a significant change over a wakeful period of 24 hours. This indicated that the motor skills of participants did not appear to decrease over time, and participants were able to retain their accuracy while simultaneously retaining their performance over the period of wakefulness. There do appear to be influences on task completion time from circadian rhythms, but they were muted in this task. Thus, there is not a speed-accuracy trade off masking the effects of sleep loss on response time.

Considering that all performance and accuracy measurements remained consistent over the 24-hour period, the previous assumption that performance and accuracy would decrease with increased hours awake did not hold true. This was unexpected based on previous research regarding psychomotor vigilance and wakefulness and the inclusion of the SAFTE equation in our model (Hursh et al., 2004; Pilcher et al., 2007). It was found that for this type of task, the previously general decrement with sleep loss does not accurately predict an individual’s activity for this task.

The sustained performance and accuracy over this time period for this type of control does not follow previously found parameters from psychomotor vigilance tasks. The results suggest that this task was not affected by sleep deprivation within the time period tested. The findings presented here

undermine the notion that fatigue affects all performance tasks equally as is currently predicted. In the future, the next steps would be to examine varied psychomotor tasks, such as the one examined in this dissertation to determine which aspects are more or less affected by sleep deprivation. This information can lead to valuable data that can be used to improve theories of the effects of sleep loss.

A few changes could be made to this type of experiment that would allow a deeper examination of tilt-based control. The first would be to use a wider range of task difficulty (e.g., smaller targets) to explore if differences can be found at higher levels of difficulty. Next, as mentioned earlier, the interface between trials could be changed to mimic a participant receiving a specific stimulus so that more possibilities for lapses to occur could be introduced.

4.4.2 Model Comparison

The second purpose of this study was to determine how human performance compares to the cognitive model. When modeled within the EPIC architecture, the model predicted a decrement in performance with increased time awake in the three performance measures: movement time, throughput, and intercept. However, the predicted changes were not supported in the experiment. Similar to other models, the parameters were modified on the number of hours awake alone. The relationship between hours awake and how the individual parameters vary is more complex than what was modeled and seems to be also dependent on the task at hand. This implies that the predictions for the task demonstrated in this study were not as expected based on current models of sleep loss. This once again supports the idea that the task presented in this experiment is more robust than expected and does not follow generalized performance predictions.

4.4.3 Study Limitations

In this particular experiment, the type of task presented to the participants was a more active task, and it is possible that it did not require the same level of vigilance to complete as other similar tasks. This could explain why there were no effects of sleep loss. This was unexpected because this tilt-based task

seemed to require low levels of engagement and was fairly monotonous, and thus, based on the controlled attention hypothesis, it should be susceptible to sleep deprivation (Pilcher et al., 2007). A possible alternative explanation for a lack of change in performance could be that the task in this experiment used different or additional cognitive or perceptual-motor systems than those used in previous tasks. It could also be that the task was simply more engaging than those used in the past, or that participants had not reached the amount of sleep deprivation needed to affect their performance.

In addition to the lack of consistency with psychomotor tasks, when working with psychological modeling, there are significantly more processes that need to be taken into account than just how long an individual has been awake. In the prediction model developed by Gunzelmann et al. (2009), variables that controlled throughput, accuracy, and error were varied based entirely on an individual's wakefulness. Other mathematical models that estimate performance based on sleep deprivation use a blanket variable of percent performance to all tasks in a non-specific manner (Hursh et al., 2004; McCauley et al., 2013). However, it is very likely that the relationship between task specific cognitive processes and fatigue-dependent cognitive processes needs to be clearly defined before a more accurate model for prediction can be built.

Gain was a constant in this particular experiment, but varying the gain would allow for a better examination for the effects of that parameter. Finally, while decrements are seen in other tasks within 24 hours, that sleep loss time span may not have been a long enough time to see performance degradation from sleep deprivation on this task. Thus, future studies may look to increase this time, as studies have shown that time awake is one of the most significant factors when examining between-studies variability (Lim & Dinges, 2010). However, these results suggest that perceptual-motor skills may be more robust against sleep fatigue than other components of thought.

Chapter 5: Effects of 24 Hour Wakefulness on Cognitive Tasks

5.1 Introduction

In Chapter 3, three distinct cognitive tasks were modeled within the ACT-R cognitive architecture. The three tasks looked at were psychomotor vigilance, n-back working memory, and visual search. Predicted changes in human performance were collected from these models. This leads into Aim 3, which is to quantify the effects of sleep deprivation on these cognitive tasks. In a similar fashion to accomplishing Aim 2, a 24-hour-based sleep deprivation study, in which participants consistently performed the three cognitive tasks, was put together and run. The tasks included a PVT, an n-back working memory task with and without distractors, and finally, a visual search task. In this Chapter, the implementation, results, and conclusion of this experiment will be discussed.

5.2 Methods

The experiment was conducted with the assistance of 16 student and personnel volunteers from the University of Connecticut. This research complied with the American Psychological Association Code of Ethics and was approved by the Institutional Review Board at the University of Connecticut. Informed consent was obtained from each participant, and participants were compensated for their time. The population consisted of 8 male and 8 female participants between the ages of 18 and 45. This sample size was chosen to accomplish two goals; detecting differences in performance and determining correlation between time awake and performance measures. A sample size of 16 was more than sufficient for detecting differences during the PVT, n-back, and visual search tasks, with a power of .95 based on the variance of previous studies that have reported on similar metrics (Dorrian et al., 2005; Rowland et al., 2005; H. L. Williams et al., 1959). In addition, with this sample size, it is possible to detect a correlation of at least .75 with a power of .9 (Zar, 1999). All procedures were conducted in the setup shown in Figure 5.1.

Figure 5.1. Experimental Setup for Cognitive Tasks.

Participants were asked to keep a regular sleep schedule and a sleep log in order to keep track of their sleeping habits. Participants were instructed to refrain from intake of any stimulants and depressants for 48 hours prior to the start of the experimental procedure.

Participants began the experimental protocol approximately one hour after waking, from which point they remained in the experimental area for 24 hours and were not permitted any sleep. All participants participated in 13 experimental sessions, one session conducted every 2 hours, where each included a psychomotor vigilance task, an n-back task, and a visual search task. In between sessions, participants were asked to limit excess physical activity.

5.2.1 PVT

The PVT experiment was performed using a standard mouse as input. The procedure was completed using the software developed by Khitrov et al. (2013). Participants sat within the setup shown in Figure 5.1, during which a black screen was presented where the stimuli appeared. Participants were asked to press the left mouse button as quickly as possible each time a stimulus appeared.

For each session, the psychomotor vigilance task lasted 10 minutes. The inter-stimulus interval was randomly determined in the range of 2 to 10 seconds. During this time, the reaction time between

stimulus appearance and left button click was measured, as well as false starts (e.g., response with no stimulus).

5.2.2 N-Back Task

Working memory was evaluated using the n-back paradigm. Custom software was written in order to run an auditory n-back test. Three n-back tasks were designed to examine multiple aspects of working memory in this study. The first task was a 1-back 1-pitch task with two tones of 100 ms or 200 ms duration and frequency of 500 Hz. Tones were presented at random, with random inter-stimulus intervals between 1 and 2 seconds. Participants were told to press the “s” key on the keyboard for each tone if it was the same duration as the tone 1-back, and the “d” key on the keyboard if the tone was different in duration to the tone 1-back. This was done for a period of three minutes.

The procedure was then repeated, except the participant compared each tone with the tone 2-back, referred to as the 2-back 1-pitch task. This type of test is more mentally taxing and was included to avoid potential ceiling effects that may be present in the 1-back task. Finally, another 1-back test was conducted, identical to the original, but each tone could be played at a frequency of 700, 750, 900, or 1250 Hz. During this third task, otherwise known as 1-back 4-pitch, the participant still compared the duration of the tones, however this task allowed us to see if adding distractors (as tone frequency components) induces a stronger effect from sleep deprivation during n-back tasks.

5.2.3 Visual Search

Visual search was done using customized software that simulated searching for ships on a horizon. An example screen can be seen in Figure 5.2. The ASL D6 Desktop EYE-TRAC™ system was used to record gaze position, blinking, and pupil dilation.

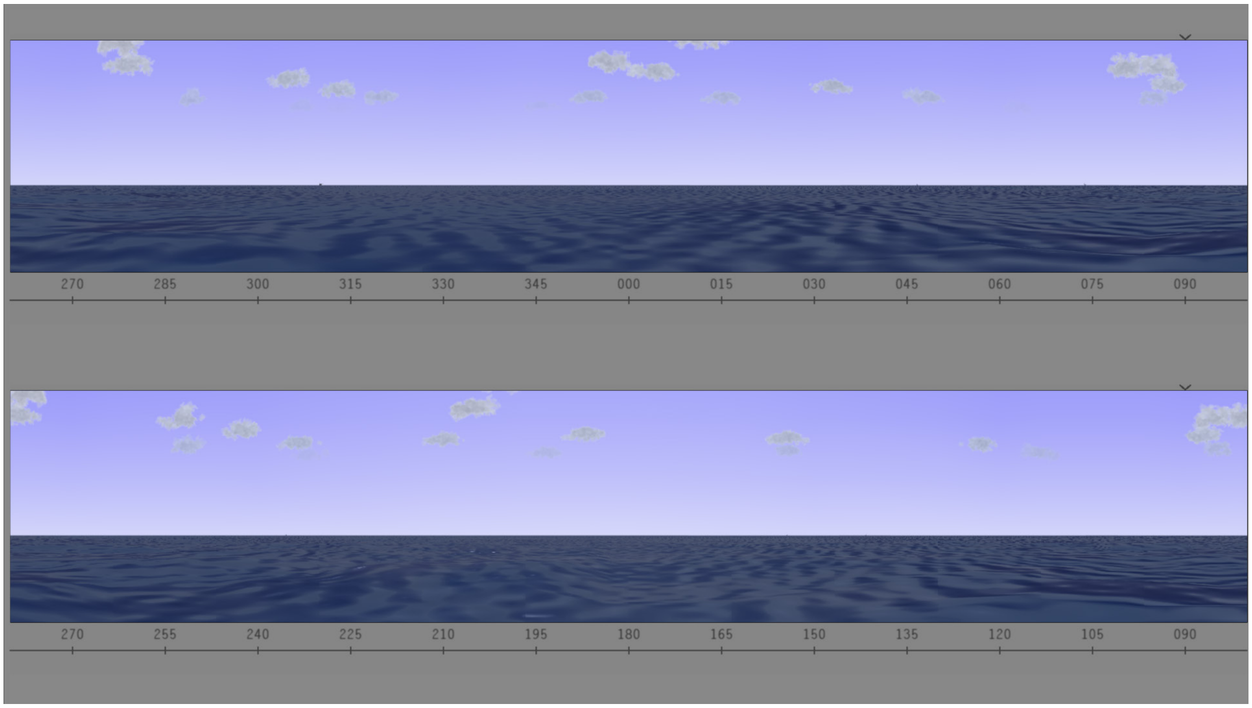

Figure 5.2. The Visual Search Task.

The initial task view for each session had no visible ships. Over time, ships would appear along the horizon, starting very small and getting larger as the ships approached the observer in the simulation. Each ship that appeared on the screen was identical. For each session, a random number of between 3 and 7 ships would appear at pseudo-random locations. Participants were asked to report the bearing of a ship as soon as they found it, alongside pressing the space key on the keyboard to confirm detection. This task lasted 20 minutes for each session, and the performance parameters being examined were: 1) target detection time, which was the time to respond to a ship from initial appearance; 2) misses, which were cases of ships not being detected before the end of the test session; and 3) false positives, which included reports of ships that did not exist or were not present on the screen at the time.

5.3 Results

5.3.1 Experimental

The results for the procedures will be discussed separately to allow for evaluation of each task type, followed by a comparison between tasks. For the evaluations of each task, we performed the examination using one-way ANOVA with Dunn-Sidak post-hoc analysis to account for multiple comparisons. The results reported are based on the sleep habits of participants, which were reported as an average of 7.5 hours (SD = 0.6) of sleep each night for 7 nights prior to the start of the experiment.

5.3.1.1 PVT

During the PVT, four parameters were examined: minor lapses (>500ms), major lapses (>1000ms), average reaction time, and false starts (Dorrian et al., 2005). For each participant, values were normalized to have unit variance. The averages of the normalized measures of these variables were taken, and are shown in Figure 5.3.

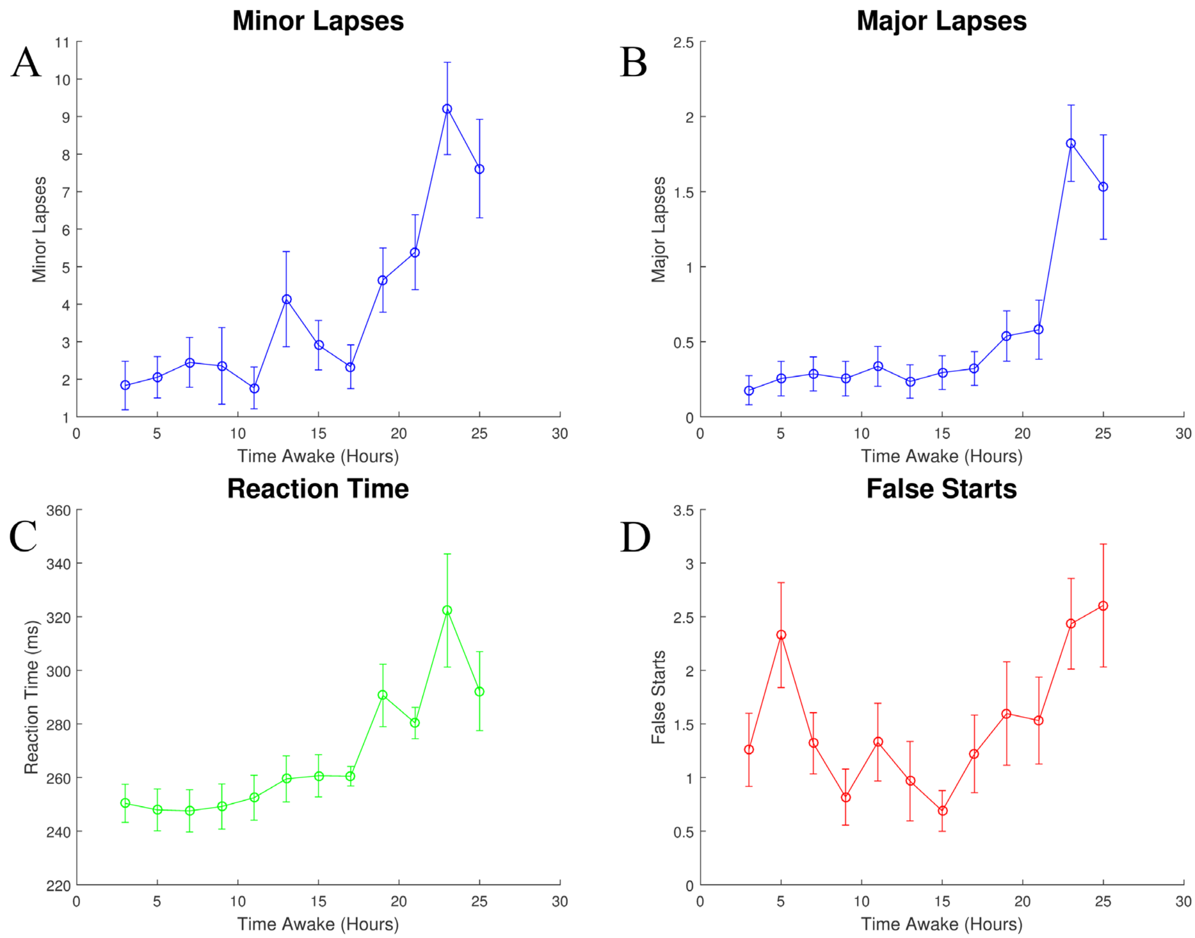

Figure 5.3. Human Performance during the PVT.

(A) Minor Lapses (±SEM) against time awake. (B) Major lapses (±SEM) against time awake. (C) Reaction time (±SEM) against time awake. (D) False starts (±SEM) against time awake.

When examining the reaction time, there was a statistically significant increase, of 74ms (30%), between reaction times at 23 hours of sleep deprivation compared to reaction times at earlier hours, $F(11,107) = 4.99$, $p < .01$. Reaction time also dropped after 25 hours awake, which was no longer significantly different from trials prior to being awake for 23 hours. In terms of the number of major lapses, there was a significant increase after 23 hours awake compared to the earlier hours, with a maximum difference of 7.2 (920%) major lapses, $F(11,107) = 7.86$, $p < .01$, and the number of major lapses continued to be significantly increased at 25 hours awake. The number of minor lapses had a

significant increase after 23 hours awake, by as much as 7.4 (420%) minor lapses, $F(11,107) = 6.48$, $p < .01$. It was found that for false starts there was a significant change in their distributions over the experimental period $F(11,164) = 2.36$, $p < .01$, and there was a measured difference of up to 1.9 (275%) false starts on average, but it was not significantly different, $p = .065$.

5.3.1.2 N-Back

During the n-back test, working memory was examined by measuring accuracy and misses for all three tests. Accuracy was determined by measuring how many tones were responded to correctly by the participant. Misses were determined by the number of tones a participant did not respond to before the next tone played, and were counted as incorrect in terms of accuracy. Once again, for all measures, we examined the average of normalized values.

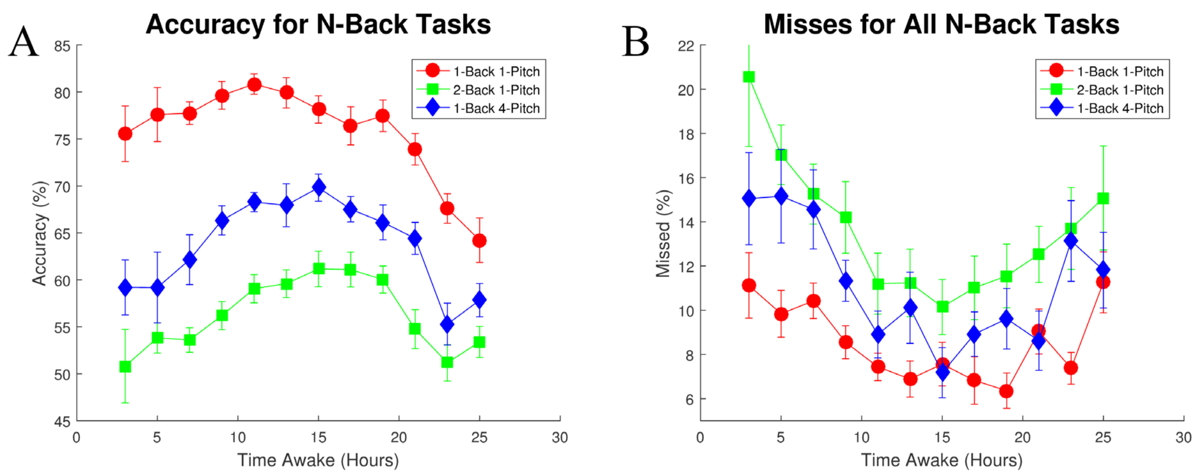

Figure 5.4. Human Performance for N-Back.

(A) N-back accuracy (±SEM) against time awake. (B) N-back misses (±SEM) against time awake.

For all n-back task measures, power law of learning effects were present despite training, and were most significant during the first four sessions. The 1-back 1-pitch task average accuracy had a significant decrease after 23 hours awake $F(11,163) = 6.93$, $p < .01$, with no increase in accuracy after that time. Misses for 1-back 1-pitch had an increased value after 25 hours awake $F(11,163) = 3.04$, $p < .01$, though just as many as at the beginning of the test. During the 2-back 1-pitch task, average accuracy

had a significant decrease after 23 hours awake, $F(11,160) = 3.7$, $p < .01$, with an increase in accuracy after 25 hours awake. Misses for 2-back 1-pitch had no significant increase over the extent of the test, $F(11,160) = 2.77$, $p < .01$. For the 1-back 4-pitch task, average accuracy had a significant decrease after 23 hours awake, $F(11,161) = 4.82$, $p < .01$, with an increase in accuracy after 25 hours awake. However, the increase was still significantly different from the prior 23 hours. Misses for 1-back 4-pitch had no significant increase during any session $F(7,111) = 1.02$, $p = .42$.

In addition, we examined the change in accuracies and miss rates across the different n-back tasks during hours awake. Regarding accuracy, 50% was used as the baseline, as that is the expectation from random choice. With this baseline, we noticed an accuracy decrease of 53% during 1-back 1-pitch, 92% during 2-back 1-pitch, and 73% during 1-back 4-pitch. Finally, there is an increase in miss rate of 77% during 1-back 1-pitch, 102% during 2-back 1-pitch, and 111% during 1-back 4-pitch. We saw the greatest increase in miss rate in the task with the most distractors.

5.3.1.3 Visual Search

During visual search, we examined three primary measures: time to detecting targets after appearance on the screen, number of targets missed, and number of non-existing identified targets (i.e., false positives) reported. The averaged normalized values of these measures are shown in Figure 5.5.

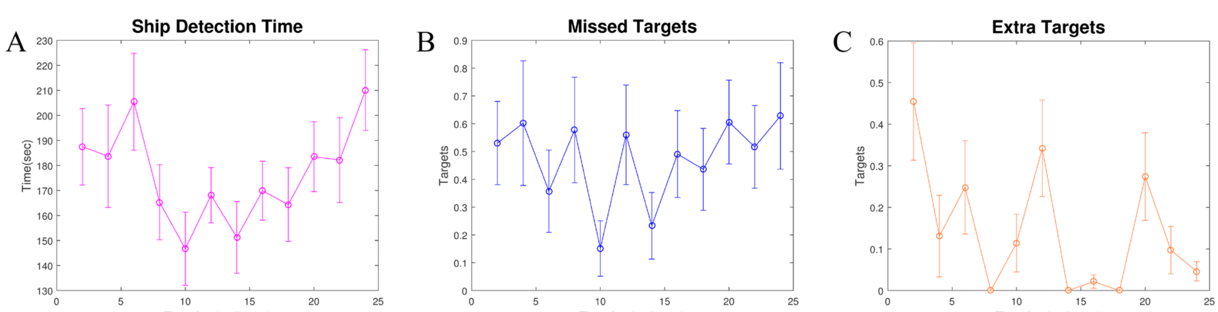

Figure 5.5. Human Performance during Visual Search.

(A) Ship detection time (±SEM) against time awake. (B) Missed targets (±SEM) against time awake. (C) Extra targets reported (±SEM) against time awake.

Similar to the n-back tasks, learning effects were visible for ship detection time despite training. The target detection time had a difference of up to 63 sec (43%), but detection times were not significantly different throughout the experimental procedure, $F(11,178) = 1.44$, $p = .16$. The number of missed targets had a difference of up to 0.45 targets on average, but also had no significant difference during the 24 hour period, $F(11,180) = 0.93$, $p = .51$. The number of extra targets, with an average increase of 0.48, showed no significant difference during the 24 hour period either, $F(11,180) = 2.04$, $p = .02$.

5.3.1.4 Eye-Tracking

During the visual search task an eye tracking camera was used to monitor each participant's eyes and record eye positions along with information on blinking, fixations, and pupil diameters. After performing one-way ANOVA in conjunction with Dunn-Sidak post-hoc analysis of the measures examined, no statistical differences with respect to hours awake were found in average pupil diameter, standard deviation of pupil diameter, fixation frequency, inter-fixation degree, or inter-fixation duration. However, statistically significant changes in blink duration, blink frequency, and fixation duration were found. The average normalized results are shown in Figure 5.6.

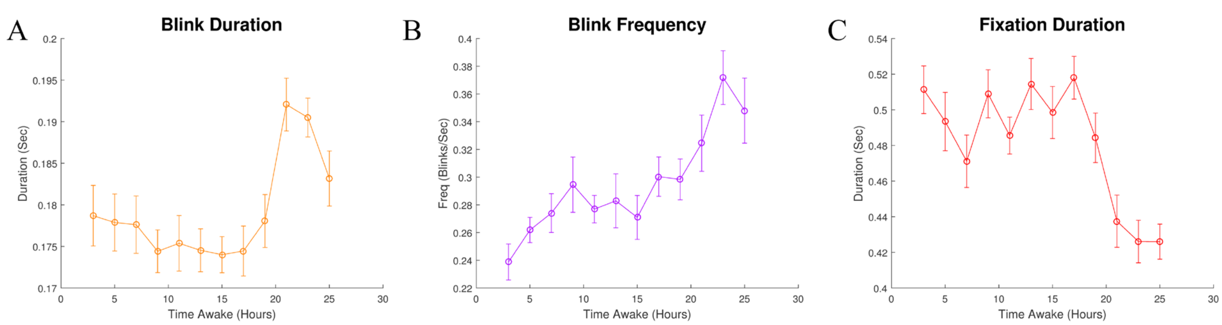

Figure 5.6. Human Eye Tracking Results.

(A) Blink Duration (±SEM) against time awake. (B) Blink frequency (±SEM) against time awake. (C) Fixation duration (±SEM) against time awake.

When examining blink duration, we found that there was a statistically significant difference after 21 hours of sleep deprivation, $F(11,178) = 4.14$, $p < .01$. Blink duration also dropped after 25 hours awake, which was no longer significantly different from trials prior to 21 hours awake. Blink frequency showed a significant increase after 23 hours awake, $F(11,178) = 4.99$, $p < .01$, and a decrease at 25 hours that was not significantly different than prior to 23 hours awake. Average fixation duration had a significant decrease after 21 hours awake, $F(11,176) = 6.41$, $p < .01$, but did not have a significant change after 21 hours awake.

5.3.1.5 Comparative Measures

We compared results from the n-back test and visual search tasks against the results of PVT to examine how well they correlated. Since the number of minor lapses was used as the primary independent variable for the psychomotor vigilance tasks, it was used as the dependent variable for correlation (Lim & Dinges, 2008). The comparisons are shown in Figure 5.7, and the correlation (*r*), 95% confidence interval (CI), significance (*p)* values, and degrees of freedom (*df*) are shown in Table 5.1.

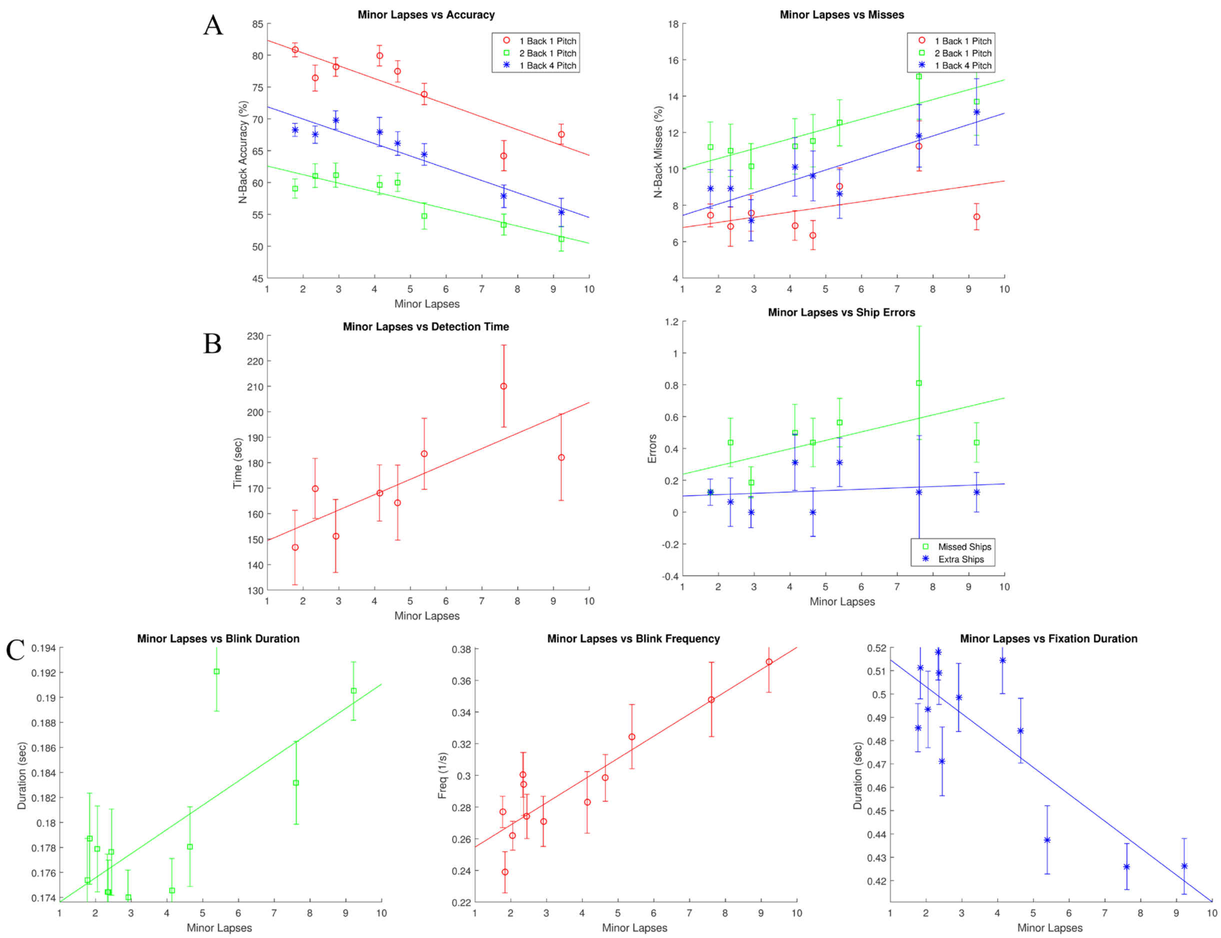

Figure 5.7. Correlation of the PVT Minor Lapses against N-Back and Visual Search Results. (A) Minor lapses versus n-back accuracy (±SEM) and misses (±SEM), respectively. (B) Minor lapses versus visual search detection time (±SEM), misses (±SEM), and extra ships (±SEM), respectively. (C) Minor lapses versus blink duration (±SEM), blink frequency (±SEM), and fixation duration (±SEM), respectively.

Table 5.1

Correlation of Minor Lapses from the PVT against all other Measures

	Accuracy 1-Back 1-Pitch	**Accuracy 2-Back 1-Pitch**	**Accuracy 1-Back 4-Pitch**	**Misses 1-Back 1-Pitch**	**Misses 2-Back 1-Pitch**	**Misses 1-Back 4-Pitch**
r	-.87*	-.91*	-.95*	.46	.86*	.85*
CI	(-.89, -.85)	(-.92, -.89)	(-.96, -.94)	(.38, .53)	(.83, .88)	(.82, .87)
p (df)	< .01 (115)	< .01 (116)	< .01 (115)	.25 (115)	< .01 (116)	< .01 (115)
	Search: Target Detection Time	**Search: Missed Ships**	**Search: Extra Ships**	**Eye Tracking: Blink Duration**	**Eye Tracking: Blink Frequency**	**Eye Tracking: Fixation Duration**
r	.78*	.64	.18	.78*	.93*	-.84*
CI	(.74, .81)	(.58, .69)	(.09, .27)	(.75, .81)	(.92, .94)	(-.86, -.82)
p (df)	.02 (116)	.08 (117)	.67 (117)	.02 (172)	< .01 (172)	< .01 (170)

Note. * Represents statistically significant correlation. Confidence intervals (CI) are presented as 95% limits of agreement.

For comparison of the PVT against n-back and visual search tasks, the final 8 sessions were compared, due to the presence of learning effects during the first four sets of trials. During the n-back tasks, all accuracy measures had high correlation with minor lapses during the PVT. However, misses for the n-back task only had significant correlation for the two harder tasks of 2-back 1-pitch and 1-back 4-pitch. For the search performance task, there was a significant correlation between target detection time and number of minor lapses. Finally, regarding the eye tracking measures, there was a significant correlation in blink frequency, blink duration, and fixation duration with minor lapses in the PVT.

Next, eye tracking measures were compared with memory and search performance, and once again the last 8 sessions were used due to the presence of learning effects. This analysis was included in order to investigate whether these physiological measures could predict task performance for tasks other

than the PVT. We specifically focused on n-back accuracy and the target detection time, due to their higher correlation values with PVT compared to other measures. The correlation between these task performances and eye tracking measures are shown in Figure 5.8, while the correlation values, confidence intervals, *p* values and degrees of freedom are shown in Table 5.2.

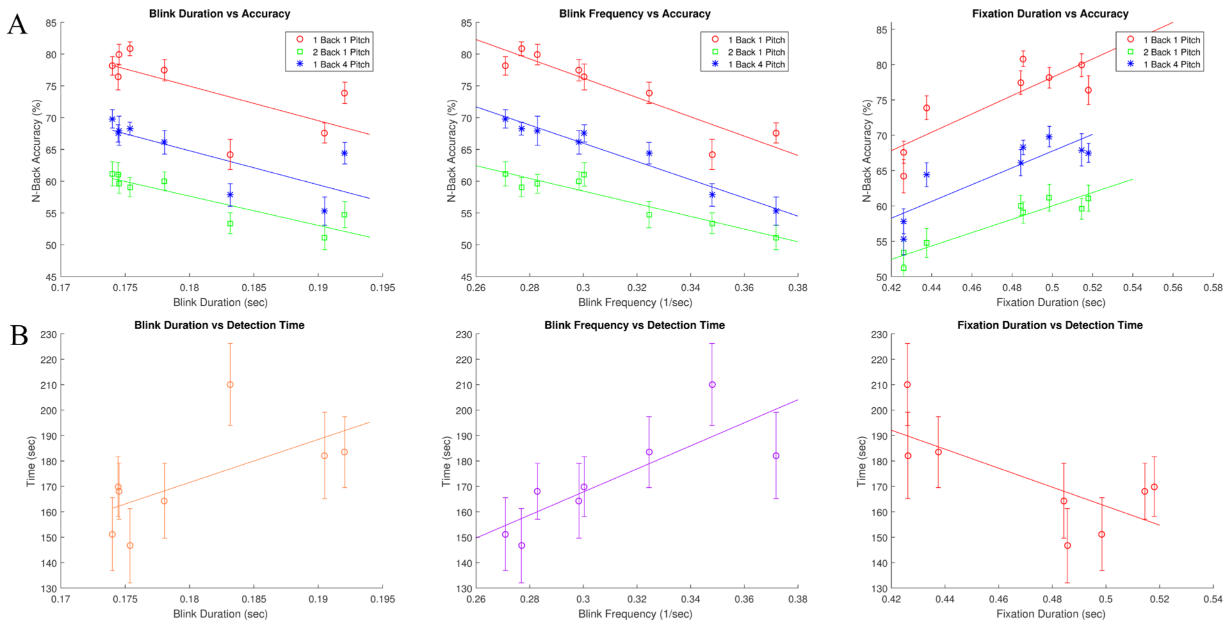

Figure 5.8. Correlation of Eye Tracking Measures against N-Back and Search Task Results.

(A) Blink duration, blink frequency, and fixation duration against n-back accuracy (±SEM), respectively.

(B) Blink duration, blink frequency, and fixation duration against detection time (±SEM), respectively.

Table 5.2

Comparison of Eye Tracking Measures against N-Back Accuracy and Time to Detection

		Accuracy 1-Back 1-Pitch	Accuracy 2-Back 1-Pitch	Accuracy 1-Back 4-Pitch	Time to Detection
Blink Duration	**r**	-.68	-.89*	-.75*	.81*
	CI	(-.73, -.63)	(-.91, -.87)	(-.79, -.71)	(.78, .84)
	p (df)	.06 (123)	< .01 (124)	.03 (123)	.01 (124)
Blink Frequency	**r**	-.92*	-.93*	-.98*	.62
	CI	(-.93, .90)	(-.94, -.92)	(-.98, -.98)	(.56, .67)
	p (df)	< .01 (123)	< .01 (124)	< .01 (123)	.1 (124)
Fixation Duration	**R**	.84*	.94*	.86*	-.71*
	CI	(.81, .86)	(.93, .95)	(.83, .88)	(-.75, -.66)
	p (df)	.01 (123)	< .01 (124)	< .01 (123)	.05 (124)

Note. * Represents statistically significant correlation. Confidence intervals (CI) are presented as 95% limits of agreement.

We found that accuracy for all n-back tasks had high correlation with blink frequency and fixation duration, and that accuracy for 2-back 1-pitch and 1-back 4-pitch correlated highly with blink duration. For the search performance task, we noticed significant correlation only with fixation duration, and blink duration.

5.3.2 Model Comparison

After collecting empirical data from the participants during this experiment, the next step was to take the results of the model runs and compare the two data sets. For all three tasks, the data was directly compared and correlations between model data and human data for each measure were calculated.

5.3.2.1 PVT

Once again for PVT, the primary measures are reaction time, major lapses, minor lapses, and false starts. During the model run there were no major lapses predicted, so a comparison of that measure was not performed. The data from both the model and human experiment were overlaid to allow for better comparison. The overlaid data, along with the correlation plots are shown in Figure 5.9.

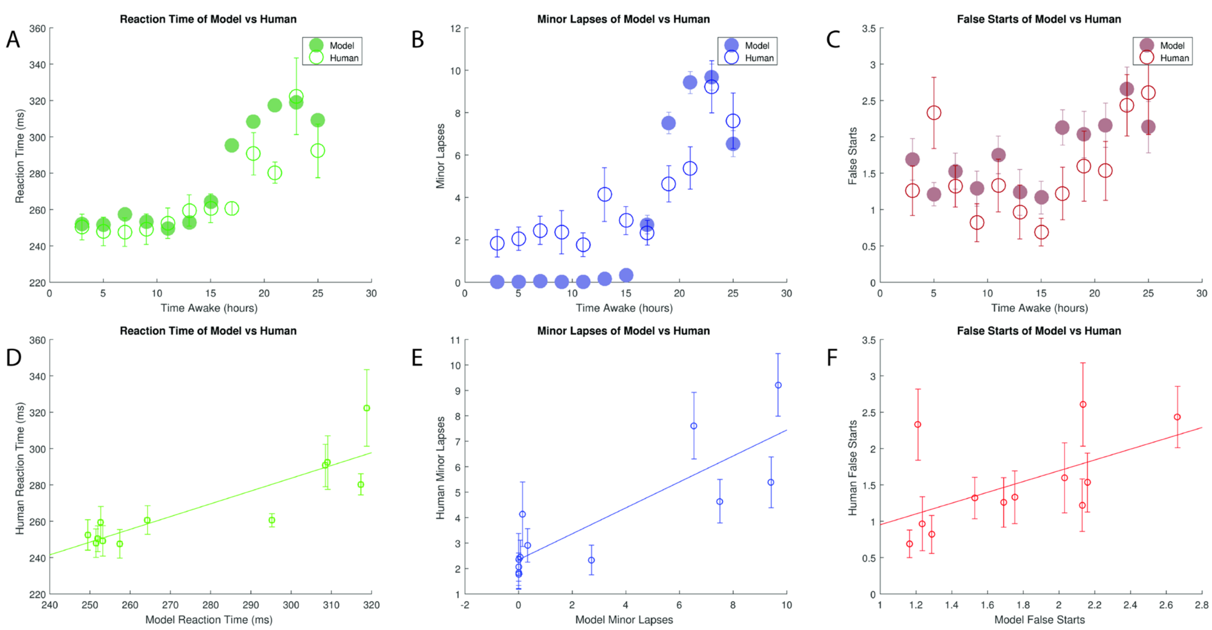

Figure 5.9. Comparison of PVT Performance between Model and Human.

(A) Reaction time (±SEM). (B) Minor Lapses (±SEM). (C) False Starts (±SEM). (D) Reaction Time Correlation. (E) Minor Lapses Correlation. (F) False Starts Correlation.

For reaction time and minor lapses, there were significant correlations of .88 and .82 respectively. For false starts, the correlation for the final 8 trials was examined, due to an evidence of some type of learning effect, and was found to be 0.8, showing significant correlation. A collection of the correlation values as well as supporting statistics are displayed in Table 5.3.

Table 5.3

Correlation of PVT Performance between Model and Humans

	Reaction Time	**Minor Lapses**	**False Starts (Last 8)**
r	.88*	.84*	.80*
CI	(.86, .89)	(.82, .86)	(.77, .83)
p (df)	< .01 (186)	< .01 (174)	0.02 (174)

Note. * Represents statistically significant correlation. Confidence intervals (CI) are presented as 95% limits of agreement.

5.3.2.2 N-Back Task

For examining the n-back back, the accuracy and misses for the all three different tasks were compared between model and human data. Due to the power law of learning effect during the first four trials of testing, correlation was only performed for the final 8 trials, as the models did not exhibit the same learning effects. A comparison of the accuracy measures for the 1-back 1-pitch, 2-back 1-pitch, and 1-back 4-pitch tasks was done, and overlays of the model versus human data, as well as correlation plots, are shown in Figure 5.10. It was found that the correlations for accuracy during the three tasks were .67, .30, and .49. For these measures, it was found that there was no significant correlation at 95% confidence. The correlation values and related statistics are shown in Table 5.4.

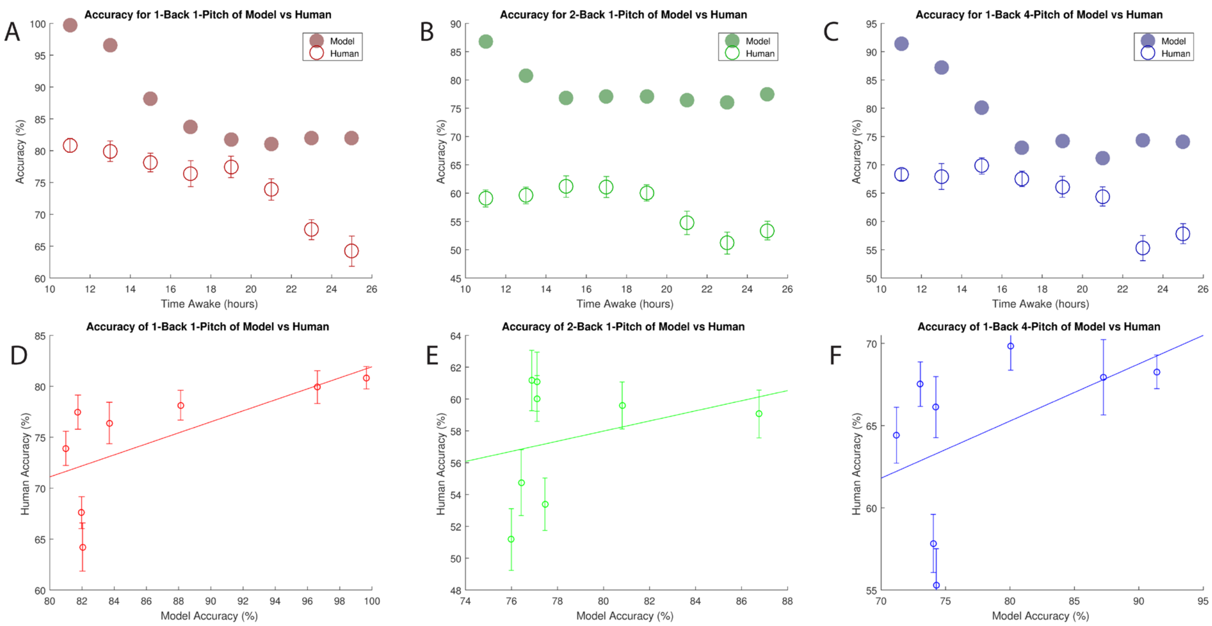

Figure 5.10. Comparison of N-Back Accuracy Performance between Model and Human.

(A) 1-Back 1-Pitch (±SEM) against time awake. (B) 2-Back 1-Pitch (±SEM) against time awake. (C) 1-Back 4-Pitch (±SEM) against time awake. (D) 1-Back 1-Pitch Correlation. (E) 1-Back 2-Pitch Correlation. (F) 1-Back 4-Pitch Correlation.

The second set examined data for the n-back tasks was the miss percentage. Data overlays between models and humans as well as correlation plots are presented in Figure 5.11. Regarding misses on 1-back 1-pitch, 2-back 1-pitch, and 1-back 4-pitch tasks, the correlation values were 0.29, 0.32, and 0.26 respectively. Once again, none of these correlations were determined to be significant, and the correlation statistics can be found in Table 5.4.

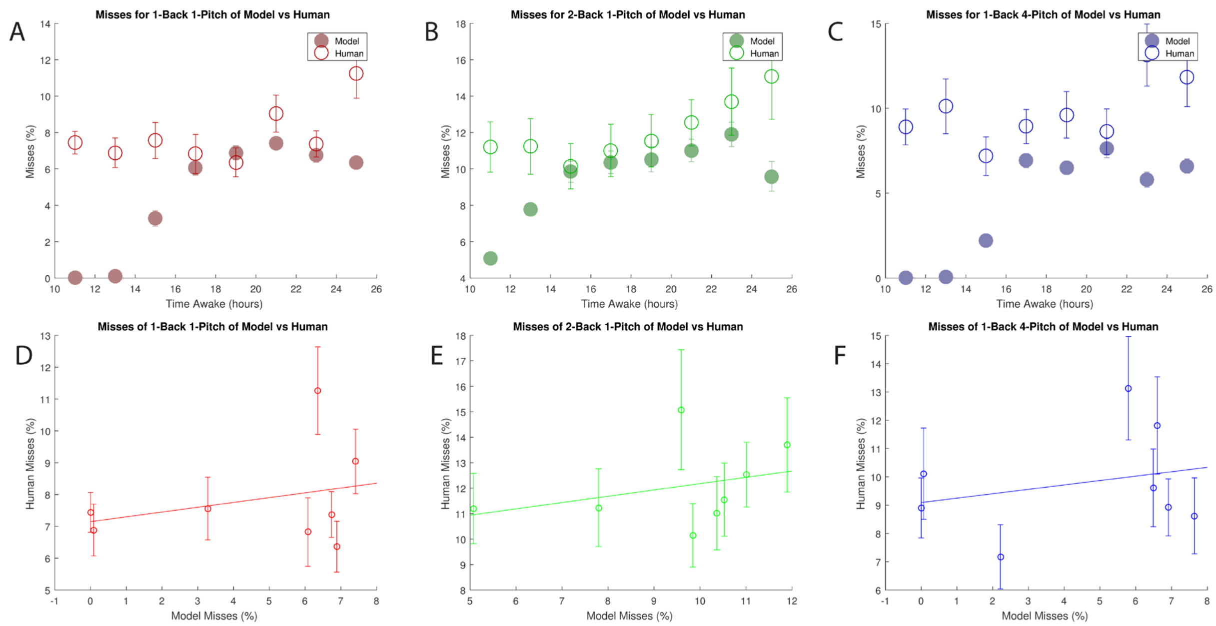

Figure 5.11. Comparison of N-Back Miss Performance between Model and Human.

(A) 1-Back 1-Pitch (±SEM) against time awake. (B) 2-Back 1-Pitch (±SEM) against time awake. (C) 1-Back 4-Pitch (±SEM) against time awake. (D) 1-Back 1-Pitch Correlation. (E) 1-Back 2-Pitch Correlation. (F) 1-Back 4-Pitch Correlation.

Table 5.4

Correlation of N-Back Performance between Model and Humans

	Accuracy 1-Back 1-Pitch	**Accuracy 2-Back 1-Pitch**	**Accuracy 1-Back 4-Pitch**	**Misses 1-Back 1-Pitch**	**Misses 2-Back 1-Pitch**	**Misses 1-Back 4-Pitch**
r	.67	.30	.49	.29	.32	.26
CI	(.61, .71)	(.21, .38)	(.42, .55)	(.21, .37)	(.25, .41)	(.17, .34)
p (df)	.07 (124)	.47 (125)	.22 (124)	.48 (124)	.43 (125)	.53 (124)

Note. * Represents statistically significant correlation. Confidence intervals (CI) are presented as 95% limits of agreement.

5.3.2.3 Visual Search

Finally, for visual search, the variables for comparison included target detection time, missed targets, and false alarms. Similarly, to the other two tasks, the overlaid comparisons and correlation plots between models and humans for each of these measures are shown in Figure 5.12. It was found that the correlation values for target detection time, missed targets, and false alarms were 0.63, 0.49, and -0.21, respectively, once again showing no significant correlation for any of the measurements. The remainder of the correlation statistics can be found in Table 5.5.

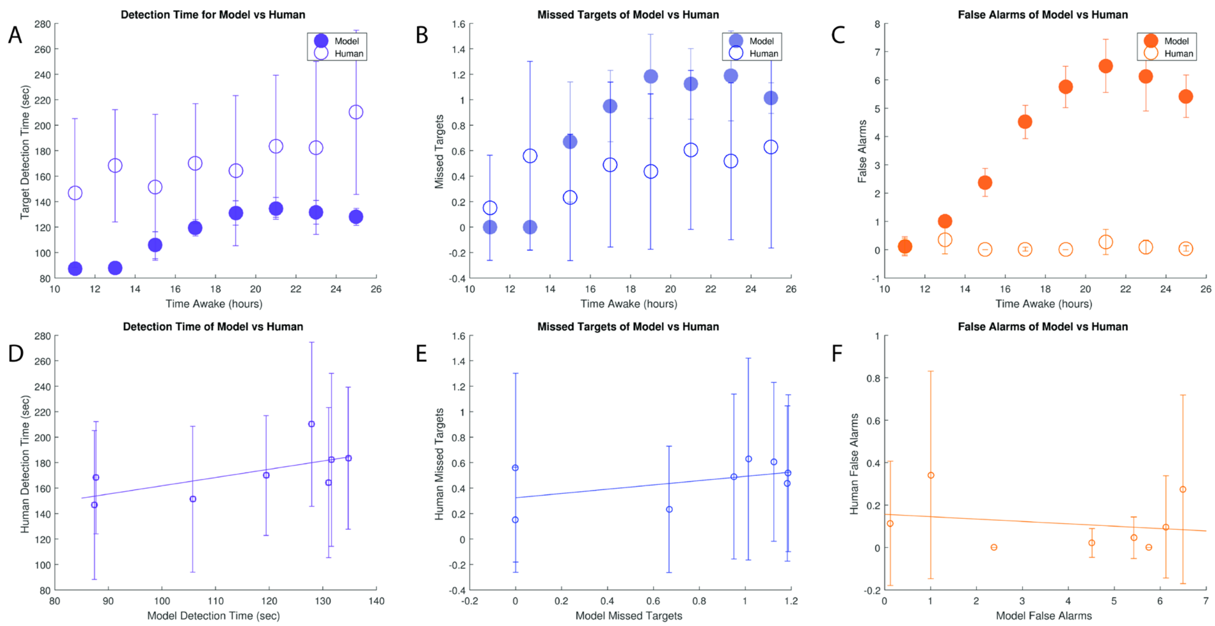

Figure 5.12. Comparison of Visual Search Performance between Model and Humans. (A) Target Detection Time (±SEM) against time awake. (B) Missed Targets (±SEM) against time awake. (C) False Alarms (±SEM) against time awake. (D) Target Detection Time Correlation. (E) Missed Targets Correlation. (F) False Alarms Correlation.

Table 5.5

Correlation of Visual Search Performance between Model and Humans

	Detection Time	**Missed Ships**	**False Alarms**
r	.63	.49	-.21
CI	(.57, .68)	(.41, .55)	(-.30, -.12)
p (df)	.09 (125)	.22 (126)	.61 (126)

5.4 Discussion

5.4.1 Experimental

The first analysis of this experiment concerned psychomotor vigilance, which was used as a baseline. Many other studies in the past have used psychomotor vigilance and have proven its robust nature as a measure of sleep deprivation (Lim & Dinges, 2008). We found that psychomotor vigilance did in fact have a response to sleep deprivation, and participants showed decreased performance in all measured parameters (minor lapses, major lapses, reaction time, false starts), as expected, and shown with previous work (Lim & Dinges, 2008; Lo et al., 2012). This allowed us to compare the results from the other tasks and measurements to those from the PVT, under the assumption that participants were affected by sleep deprivation as expected.

We utilized three different variations of the n-back task to examine working memory. First, when examining the 1-pitch 1-back test, a trend in dropping performance began after approximately 19 hours awake, with a significant decrease in performance starting at 23 hours awake. Unlike other measures, 1-back 1-pitch performance did not recover at the 25-hour mark with the circadian rhythm cycle. This implies that there is diversity between specific tasks, even between different n-back variations. For the 2-back 1-pitch we found there was a downward performance trend after 19 hours with a significant decrement at approximately 23 hours of wakefulness. However, unlike the 1-back 1-pitch, we did see some effect on performance due to circadian rhythm, after being awake for 25 hours. Finally, during the 1-back 4-pitch, which included distractors, we noticed similar results as with the 2-back 1-pitch. This is not too surprising as they are both more difficult than the 1-back 1-pitch scenario. However, the decrease in accuracy was highest in the 2-back memory protocol compared to the other two, while the miss percentage was higher during 1-back 4-pitch, suggesting that more difficult tasks and those which introduce distractors are more affected by sleep deprivation.

When comparing working memory task results to those from the PVT, we found that all accuracy measures had high correlation coefficients, all with high significance. These results suggest that minor

lapses from the PVT could be used as a predictor for accuracy during an n-back test, across all difficulties examined in this study, during sleep deprivation. The miss percentage for 2-back 1-pitch and 1-back 4-pitch correlated with minor lapses during the PVT, however miss percentage for 1-back 1-pitch was not correlated, which once again suggests that easier working memory tasks are affected less by sleep deprivation.

When examining visual search, we found that the measures of target detection time, missed targets, and reporting extra targets (false positives) were not statistically significantly linked to hours awake. However, target detection time was the only measure to show significant correlation with minor lapses from the PVT. Since there was a significant correlation, but not enough statistical difference, we suggest that visual search is indeed affected by sleep deprivation and some performance components can be predicted using PVT. However, at least in the visual search task used in this study, the variance in performance was larger than expected due to the limited number of targets, or perhaps the short time period for the visual search. The effect size in this task was smaller than expected and, while there could be increased significance with more participants, it would be more worthwhile to pursue higher complexity tasks that are more ecologically relevant. Classic thinking would suggest that tasks that are less engaging are more vulnerable to performance decrement due to sleep deprivation, but this seemingly contradictory result may be influenced by the high variance seen in this study (Pilcher et al., 2007).

Our comparison of eye tracking measures against human performance during the n-back and visual search tasks provided insight into how well physiological measures, in this case those derived from eye tracking, can potentially predict task performance. As stated earlier, we focused on three eye tracking measurements: blink duration, blink rate, and fixation duration. We found that most accuracy measurements during the n-back task were correlated with all three eye tracking measures during the visual search task. In addition, we found that time to ship detection was correlated with fixation duration and blink duration. These results suggest that these eye tracking measures can potentially be used as predictors for performance during working memory and visual search tasks.

5.4.2 Model Comparison

We now shift the discussion to comparing the results of the model to those of the experiment described in this Chapter. Comparing the PVT predictions from the model to collected human data, there was a correlation of 0.87, 0.84, and 0.8 with reaction time, minor lapses, and false starts. These were the only three measures of PVT that significantly correlated between the human and the model. In this particular case, these correlations were not surprising, especially the correlations of reaction time and minor lapses, as they were the primary parameters used when determining effectiveness in the SAFTE function (Hursh et al., 2004). However, the lack of a significant relation between the model and the human, when it came to false starts and major lapses was an unexpected result.

When examining major lapses, the model predicted no major lapses at any point during the period of 24 hours awake. The take away for this is in fact two sided. One possible explanation for this result is that the functionality included in the model, while reasonably predicting minor lapses, did not truly encompass the overall mechanism for lapses, or major lapses. With that being said, when examining the data, it was found that four of the sixteen participants experienced no major lapses, during the timeline of the experiment. For that subset of participants, the model did predict the number of major lapses, to some degree. Even so, the average number of major lapses peaked at approximately two, making them a rare occurrence for this experimental setup. These outcomes suggest one of two possibilities when it comes to major lapses: either the mechanism was not captured, or the models need to be tuned to individuals for accurate prediction of major lapses.

The next comparison of the n-back task showed no significant correlation between model and human data among all accuracy measures during the n-back task. While this was surprising at first, upon further examination if we compare when the performance begins to significantly decrease, we notice a consistent difference between the model and the human. Regarding the accuracy measures of the model, a significant change in performance begins at approximately 15 hours awake, which matches the results of similar working memory models developed in the past (Turner, Drummond, Salamat, & Brown, 2007).

However, in the human results, a significant change in performance starts occurring at approximately 23 hours. This dissidence suggests that the performance prediction from the n-back task would stray from that predicted by the SAFTE model. The comparison of misses between the model and human data during the n-back task also showed low correlation values. Once again, the biggest factor seen here was when the decrement in performance was first significant, that is, 15 hours for the model and 23 hours for people.

What is fascinating in this comparison is that during the experiment, minor lapses measured during the PVT correlated with performance measures from the n-back, implying that current methodology for the implementation of the working memory model for the purposes of sleep deprivation is partially flawed. In addition, the model and human data become even less correlated with increased difficulty, pointing to the fact that the model or current architecture does not support modeling performance with increasing difficulty.

Finally, during visual search, there were no significant correlations between human and model data for any values in visual search. Looking at detection time, the most likely contributor to the lower level of correlation is simply low significance in change in the detection time for humans, versus a significant change in detection time for the model. This particular phenomenon is even more prevalent in missed ships as well as false positives, neither of which show any upward trend for the human data. The data presented during the search task provide more evidence that performance on certain tasks, such as visual search, are not necessary predicted by the SAFTE model, and there are other, more complex factors involved that are not taken into account. In addition, there is more dissidence for the false positive results, as ACT-R currently does not have a built-in module for errors of commission, and thus the implementation based on the SAFTE model is not necessarily the correct way to go about exploring performance in this area.

Chapter 6: Conclusions

This dissertation sets out to examine three primary aims: to examine the current state of modeling with respect to sleep deprivation, to explore the effects of sleep deprivation on motor control tasks, and to look at the effects of sleep deprivation on various cognitive tasks. Throughout the last five chapters all of these aims have been addressed, and the results of all findings have been presented. In this final Chapter, some conclusions based on the findings are presented. In addition, this Chapter includes a discussion of future work regarding some of the major areas lacking in information.

6.1 Cognitive Modeling

In Chapter 3, four separate models were presented to predict human performance of various tasks while under the effects of sleep deprivation. This was primarily done through the use of cognitive architectures, which provided a framework for modeling cognition and predicting human behavior (Duch et al., 2008). With their collection of physiological and psychological measures from a breadth of empirical studies, they provided a new level of performance analysis. A mathematical model for task performance during sleep deprivation, specifically SAFTE, was integrated into cognitive models to accomplish this goal. This resulted in a more complete picture for predicting human performance than simply using mathematical constructs as standalones. The goal of these models was to serve as a precursor, and to predict task performance without prior collection of data.

The first model presented was that of a motor control task. With the capacities available to the EPIC cognitive architecture, the variables that control manually aimed movement for a task were examined in conjunction with the SAFTE model. By incorporating the SAFTE model as a modulator for the motor control coefficients within the architecture, it was expected that it would be possible to predict the performance of a person during a tilt-based motor control task under the effects of sleep deprivation. The output of the model yielded results that predicted decrement of performance based on fatigue caused by sleep deprivation.

A model of PVT was developed in ACT-R, a cognitive architecture that is most suited for the analysis of primarily cognitive tasks (Anderson et al., 2004). Utilizing prior methodology, the SAFTE model was used to produce lapses in the cognitive model in order to mimic the observed phenomena found with people during sleep deprivation (Gunzelmann et al., 2009; Lim & Dinges, 2008). Performance measures for reaction time, lapses, and false starts were predicted with increased sleep deprivation, and showed performance decrement.

Working memory was also modeled in ACT-R, as it is another primarily cognitive task. An auditory n-back task with different lengths and frequencies of tones was designed, as n-back procedures are the most commonly used tools for examining working memory (Jaeggi et al., 2010). Utilizing the declarative memory capacity of the ACT-R architecture, the SAFTE model was integrated to both modulate memory recollection as well as produce confusion to simulate the model succumbing to distractions during the task. With this, the model produced predictions of performance for 1-back 1-pitch, 2-back 1-pitch, and 1-back 4-pitch tasks during sleep deprivations, showing decreased accuracy and increased miss rate with time awake.

Finally, a visual search model produced in ACT-R, was developed to perform a target detection task. Like the other three models, the SAFTE model was included for predicting the effects of fatigue on performance. In this model, the primary areas for modulating performance were in visual acuity, or the ability for the model to properly detect a target, as well as the detection of false positives. Once again, the model predicted performance decrements with sleep deprivation. These decrements were measured with target detection time, false positive rate, and missed targets.

As mentioned earlier, the purpose of these models was to act as precursors. The goal was to examine the possibilities of producing predictions for sleep deprivation in cognitive models, a task which has been approached by some, but is far from completion (Gunzelmann et al., 2009). All predictions and designs were based off of information available in literature to discover how well current models would

predict performance during sleep deprivation, and what potential information is necessary to move towards advancing their predictive capabilities.

6.2 Motor Control

Toward the completion of Aim 2, a study was run to examine the change in users' performance and accuracy while using a tilt-based control device over a period of 24 hours of sleep deprivation. The observed performance parameters of movement time, throughput, and average intercept did not significantly change over the duration of the experiment. In addition, the accuracy parameters of movement error, movement variability, and number of reentries also did not change over the experimental period. These findings highly suggest that performance during a motor control task, such as the one presented during this experiment, is fairly robust.

6.2.1 Countered Expectations

The sustained performance and accuracy over a 25-hour period for motor control does not follow previously found parameters from PVTs. The findings presented here undermine the notion that fatigue affects all performance tasks equally, as is currently predicted. This may have been caused by a number of factors, one possibility being the lack of vigilance present in this procedure versus other psychomotor-based tasks. These conclusions support the idea that the tilt-based motor task can be better explained by the controlled attention hypothesis. Looking at this task as a more "active" or "engaging" task in this framework leads to the fact that it is not as affected by sleep deprivation (Pilcher et al., 2007).

Future explorations in this area could fine tune the influences on vigilance and other variables by varying pieces of the task, such as the time between targets or active start button, to determine how each change affects performance. Finally, while decrements are seen in other tasks within 24 hours, the time span may not have been long enough to see performance degradation caused by sleep deprivation on this task, and future studies may look to increase this time, as past studies have shown that time awake is one of the most significant factors when examining between-study variability (Lim & Dinges, 2010).

6.2.2 Tilt vs EPIC

The prediction of the motor task, as presented by the EPIC cognitive architecture, showed a significant decrement in performance over a 24-hour period. However, as stated before the results from the human experiment did not show such a change. This unexpected outcome could have many potential explanations, however the two most likely, in terms of this work, focus on the SAFTE model and the architectural influences.

One potential explanation points to an additional transform to the SAFTE model, if not a different model all together, necessary to calculate a task effectiveness. It has been suggested that the SAFTE model can be modified in order to fit the data of a specific task. However, the data for the task would have been needed to be collected prior to the design of the model, which counters the design of this dissertation.

6.3 Cognitive Tasks

The goal regarding the cognitive experiment described in Chapter 5 was to examine how the PVT and physiological measurements work as predictors for performance during sleep deprivation, and how well the model developed prior predicted human performance. Here, the results and conclusions from this study will be discussed, with the goal of discussing the meaning behind them.

6.3.1 Predictive Power of PVT and Eye Tracking

The 24-hour study described in Chapter 5 provided a unique opportunity to compare the results of the PVT as well as the measures taken from the eye tracking camera, to see how well these metrics predicted the performance of other tasks. From the memory and visual search tasks examined, it was found that PVT measures correlated with the results of the n-back task, the target detection time of the visual search task, and the three eye tracking metrics presented. These outcomes suggest that the PVT can potentially be used as a predictor for alternative and more complex tasks, which require working memory and visual search. It was also found that the eye tracking measures, physiological metrics, can potentially

predict the performance of all three tasks as well. This information provides valuable relationships that would allow more complex tasks to be predicted by the performance of simpler tasks or even physiological measures. Thus, such metrics could be used to predict operationally relevant indicators for environmentally relevant efforts.

6.3.2 PVT vs ACT-R

When comparing human performance of the PVT with the ACT-R model, it was found that there was significant correlation between the two datasets with respect to reaction time, minor lapses, and false starts. As stated before, this was not too surprising, although it reinforced that for these predicted variables the mechanisms that were examined within the ACT-R model looked as though they led in the correct direction. In addition, the SAFTE model integration seemed to be an adequate fit for this particular data set.

For the other metrics, the model predicted no major lapses and did not have significant correlation with the human data. The result for major lapses was difficult to reconcile, as a number of participants showed no major lapses during the experiment. Thus, the two most likely conclusions were that the model needs to be tuned to individuals to adequately measure major lapses, or that it does not encompass the mechanism for them. This particular debate would require a different level of analysis and psychological approach when creating the model, though it may be valuable to examine in a future step.

6.3.3 N-back vs ACT-R

When comparing the results of the working memory task between the model and human data, the prediction capacity of the model was less than expected. Looking at accuracy measures for the three tasks, it was found that there was no significant correlation between human and model data for any of the tasks. Taking a closer look, one of the major differences between the human and the model was the timing at which performance of the task was significantly decreased. While there may be a number of contributing

factors to this separation, there are two primary explanations as to why the prediction of the model did not exactly fit.

The first possibility of the inconsistency in performance prediction for the n-back task is encompassed by the idea that its performance is not captured by the SAFTE model. While there have been modifications of SAFTE and similar models in the past to accommodate for specific tasks, those modifications required prior collection of experimental data for those tasks (Hursh et al., 2004). However, as the point of this research was to test how well models based on literature could predict human performance, the information available was taken and modeled prior. In the future, the human data presented in this dissertation can be used to modify the model for working memory based task. The other possible explanation for the dissidence between the model and human data is a focus and modification of the incorrect parts of the cognitive architecture. While the model was built around the declarative module of ACT-R, if other cognitive influences from sleep deprivation exist within this context, then the implemented model is not yet complete. Another unexpected factor in these comparisons is that the correlations between human and model data for the easier tasks are higher than those of the more difficult ones. This implies that there is another underlying mechanism with unexpected interactions with tasks that are more difficult or have more distractors.

The percentage of missed stimuli for the n-back task also did not correlate between the human and model data. This measure can potentially be explained by the same phenomena as the accuracy data, however there is one more factor to consider for this measure. The misses in this particular task are based on errors of omission, which are handled somewhat differently in the PVT model, as they are determined based on the lack of ability to recall a tone. Though in the end, the complex interactions between the SAFTE function and the cognitive models make it difficult to disentangle which model component is lacking the information for higher accuracy prediction.

6.3.4 Visual Search vs ACT-R

Finally, the resolution of the visual search model, similar to that of the n-back model, had a low level of correlation between the model and the human data. For target detection time, while both the human and model have a general increasing trend with time awake, their correlation is not particularly high. This leads to similar conclusions to those reached from the n-back task: the SAFTE model could not fit exactly with this type of task, or the mechanisms examined in the model may not create a full representation of the effects of sleep deprivation.

While the results for the comparisons of false positives and missed targets also point to the same outcome, there is another factor to take into consideration here. The human data for these two measures showed no upward trends with time awake, thus, implying that these two performance metrics, specifically errors of omission and errors of commission, may be more robust to the effects of sleep deprivation than expected, similarly to the performance analysis of the tilt-task.

6.4 Future Work

The modeling, data collection, and comparisons performed and discussed over the past five chapters have provided insight as well as more questions. A delve into cognitive models has shown the current state of these models, as well as how approaches have been made to allow them to predict human performance of tasks during sleep deprivation. Furthermore, the results that have been described here have also illuminated many future paths of work that can be pursued to advance the field.

6.4.1 Performance Measurements

For most of the performance measurements, the inconsistencies for predictions provide a vast number of factors to be examined in the future. For the tilt-based task, the next steps would be to examine varied versions of motor control to determine which aspects are more or less affected by sleep deprivation. Some examples of these variations include using a wider range of task difficulty (e.g.,

smaller targets) to see if differences can be found at higher levels of difficulty, or changing the user interface to introduce more vigilance to the task.

When it comes to the cognitive tasks, a lot of information was provided by the work done within this dissertation. The next step would be to work with the modules within ACT-R for a deeper examination of the current cognitive models. With this type of work, it is possible to test different mechanisms and determine how to potentially create a better framework with the available data.

A deeper look into distractors and task difficulty would provide greater insight into how these two factors affect performance. While the work presented here provided some preliminary observations on task difficulty, designing tasks with varied numbers of distractors could illustrate a more in-depth relationship between the number of distracting elements and the ability to perform a task. Finally, since the simple visual search task had such a small effect size relative to variance, designing a study with a longer search period accompanied by an increased number of targets and distractors could provide a clearer relationship between visual search performance and sleep deprivation.

6.4.2 Accuracy and Eye Tracking in Cognitive Architectures

During the motor control task, accuracy measures were collected, including movement variability, movement error, and target reentries. It was mentioned that the EPIC architecture did not have the capacity to model those types of measures, and it is not the only one (Kieras & Meyer, 1996). Currently, cognitive architectures motor capacity can be described as direct to target paths, with noise and movement calculation influenced only by movement time. With the data presented in this work, and possibly in other in-depth movement human motor control studies, integrating human error in movement into cognitive architectures can be a possibility.

Another missing construct from cognitive architectures includes eye tracking metrics, which were collected during the visual search task. These data provide additional information about how people went about searching for targets, and possible influences on their performance. However, as of this moment,

the eye movement mechanics of cognitive architectures are fairly simplistic (Bothell, 2004; Kieras & Meyer, 1996). Beyond simple eye movements, blinks and eye closure are currently factors that models do not take into account. With the information found during the cognitive task study, the eye tracking metrics can be utilized to advance models and assist in improving their predictive capacities.

6.5 Final Thoughts

This dissertation set out to complete three primary objectives, which included evaluating the state of cognitive models and their capacity to predict performance during sleep deprivation, by determining the effects of sleep deprivation on motor control tasks, and examining the effects of sleep deprivation on cognitive tasks. All of this work was done primarily to both advance the field in knowledge and create the stepping stones and methodology for advancing cognitive models to better predict human behavior during sleep deprivation. This work is invaluable to countless industries and service organizations in which performance decrement due to fatigue can cause serious damage.

By both building models and collecting data, the information and insight gained in this dissertation is much greater than the sum of its parts. Beyond just information on how sleep deprivation affects individual tasks, the relationships between the performance decrement for each task were found. This provided a basis beyond just modeling, but into the prediction of performance based on other measures, such as vigilance or eye tracking metrics. With these comparisons, we can start examining how to predict performance during more complex tasks and other operationally relevant indictors.

While significant information was gained in the creation and building of these models, more questions and theoretical issues have come to the surface. It was found that for basic psychological vigilance tasks, the current state of models and knowledge had a reasonable level of prediction. However, for many of the other tasks, the models and information available to build them did not quite match recorded outcomes. The work done within this dissertation has shown that the differences between tasks can cause major discrepancies in performance and how they are affected by sleep deprivation. While

models can be adjusted to fit data, that idea does not serve to support prediction and risk mitigation. Thus, data such as that collected in this dissertation is imperative to advancing this field and supporting generalized risk management. With better data of simple, compartmentalized tasks, it is possible to construct models of higher complexity and environmental validity. And with that, this dissertation accomplished its original intent, which was to examine the effects of sleep deprivation on cognitive and motor performance of simple tasks with the goal of contributing to quantitative predictive models of complex real-world tasks.

6.6 Acknowledgements

The work presented in this dissertation would have not been possible without the help and support of countless people. First and foremost, I would like to thank my advisors and mentors, Dr. Michael Qin and Dr. Ki Chon, for their countless hours of advice, review, and support throughout this process. I would also like to thank all those that have helped and worked with me in the Biomedical engineering lab at UCONN, with a special thanks to Dr. Natasa Reljin, Dr. Hugo Posada, Dr. Bersain Reyes, and Stefan Sumsky. Next, I could not have done all this without the support of the people at the Naval Submarine Medical Research Lab (NSMRL), with a special thanks to Matthew Babina, David Gever, and Dr. Ethan Hada. Additionally, I would like to thank my friends and family for all their support, with a special thanks to Alex Strupinski and Lia Bonacci. Finally, a thank you to NSMRL, the Office of Naval Research, the University of Connecticut, and Worcester Polytechnic Institute for the financial support which made this all a reality.

References

Achermann, P., & Borbély, A. A. (1994). Simulation of daytime vigilance by the additive interaction of a homeostatic and a circadian process. *Biological Cybernetics*, *71*(2), 115–121.

Alhola, P., & Polo-Kantola, P. (2007). Sleep deprivation: Impact on cognitive performance. *Neuropsychiatric Disease and Treatment*, *3*(5), 553–567.

Anderson, J. R. (2013). *The architecture of cognition*. Psychology Press.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036.

Balkin, T. J., Rupp, T., Picchioni, D., & Wesensten, N. J. (2008). Sleep loss and sleepiness: current issues. *Chest Journal*, *134*(3), 653–660.

Binks, P. G., Waters, W. F., & Hurry, M. (1999). Short-term total sleep deprivations does not selectively impair higher cortical functioning. *Sleep: Journal of Sleep Research & Sleep Medicine*.

Borb, A. A., & Achermann, P. (1999). Sleep homeostasis and models of sleep regulation. *Journal of Biological Rhythms*, *14*(6), 559–570.

Borbély, A. A., Daan, S., Wirz-Justice, A., & Deboer, T. (2016). The two-process model of sleep regulation: a reappraisal. *Journal of Sleep Research*, *25*(2), 131–143. https://doi.org/10.1111/jsr.12371

Bothell, D. (2004). Act-r 6.0 reference manual. *Working Draft*.

Card, S. K., Moran, T. P., & Newell, A. (1986). The model human processor: An engineering model of human performance. *Handbook of Human Perception*, *2*(Journal Article).

Chee, M. W., & Choo, W. C. (2004). Functional imaging of working memory after 24 hr of total sleep deprivation. *The Journal of Neuroscience*, *24*(19), 4560–4567.

Collins, M. G., Juvina, I., & Gluck, K. A. (2016). Cognitive model of trust dynamics predicts human behavior within and between two games of strategic interaction with computerized confederate agents. *Frontiers in Psychology*, *7*, 49. https://doi.org/10.3389/fpsyg.2016.00049

Dawson, D., Darwent, D., & Roach, G. D. (2017). How should a bio-mathematical model be used within a fatigue risk management system to determine whether or not a working time arrangement is safe? *Accident Analysis & Prevention*, *99*, 469–473.

Dinges, D. F. (1995). An overview of sleepiness and accidents. *Journal of Sleep Research*, *4*(s2), 4–14.

Dinges, D. F., Pack, F., Williams, K., Gillen, K. A., Powell, J. W., Ott, G. E., … Pack, A. I. (1997). Cumulative sleepiness, mood disturbance and psychomotor vigilance performance decrements during aweek of sleep restricted to 4-5 hours per night. *Sleep: Journal of Sleep Research & Sleep Medicine*.

Doran, S., Van Dongen, H., & Dinges, D. F. (2001). Sustained attention performance during sleep deprivation: evidence of state instability. *Archives Italiennes de Biologie*, *139*(3), 253–267.

Dorrian, J., Rogers, N. L., & Dinges, D. F. (2005). Psychomotor vigilance performance: Neurocognitive assay sensitive to sleep loss.

Drummond, S. P., Bischoff-Grethe, A., Dinges, D. F., Ayalon, L., Mednick, S. C., & Meloy, M. (2005). The neural basis of the psychomotor vigilance task. *SLEEP-NEW YORK THEN WESTCHESTER-*, *28*(9), 1059.

Duch, W., Oentaryo, R. J., & Pasquier, M. (2008). Cognitive Architectures: Where do we go from here? In *Frontiers in Artificial Intelligence and Applications* (Vol. 171, pp. 122–136).

Durmer, J. S., & Dinges, D. F. (2005). Neurocognitive consequences of sleep deprivation. In *Seminars in Neurology* (Vol. 25, pp. 117–129).

Fitts, Paul M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *47*(6), 381.

Fitts, P.M., & Posner, M. I. (1967). *Human performance.* Oxford, England: Brooks/Cole.

Gunzelmann, G., Gross, J. B., Gluck, K. A., & Dinges, D. F. (2009). Sleep deprivation and sustained attention performance: Integrating mathematical and cognitive modeling. *Cognitive Science*, *33*(5), 880–910.

Harrison, B. L., Fishkin, K. P., Gujar, A., Mochon, C., & Want, R. (1998). Squeeze me, hold me, tilt me! An exploration of manipulative user interfaces. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 17–24). ACM Press/Addison-Wesley Publishing Co.

Hursh, S. R., Redmond, D. P., Johnson, M. L., Thorne, D. R., Belenky, G., Balkin, T. J., … Eddy, D. R. (2004). Fatigue models for applied research in warfighting. *Aviation, Space, and Environmental Medicine*, *75*(Supplement 1), A44–A53.

Ingre, M., Van Leeuwen, W., Klemets, T., Ullvetter, C., Hough, S., Kecklund, G., … Åkerstedt, T. (2014). Validating and extending the three process model of alertness in airline operations. *PloS One*, *9*(10), e108679.

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory*, *18*(4), 394–412.

Jagacinski, R. J., Repperger, D. W., Moran, M. S., Ward, S. L., & Glass, B. (1980). Fitts' law and the microstructure of rapid discrete movements. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(2), 309–320. https://doi.org/10.1037/0096-1523.6.2.309

Jewett, M. E., & Kronauer, R. E. (1999). Interactive mathematical models of subjective alertness and cognitive throughput in humans. *Journal of Biological Rhythms*, *14*(6), 588–597.

Jones, K., & Harrison, Y. (2001). Frontal lobe function, sleep loss and fragmented sleep. *Sleep Medicine Reviews*, *5*(6), 463–475.

Kane, M. J., Conway, A. R., Miura, T. K., & Colflesh, G. J. (2007). Working memory, attention control, and the N-back task: a question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 615.

Kendall, A. P., Kautz, M. A., Russo, M. B., & Killgore, W. D. (2006). Effects of sleep deprivation on lateral visual attention. *International Journal of Neuroscience*, *116*(10), 1125–1138.

Kerkhof, G., & Van Dongen, H. (2010). Effects of sleep deprivation on cognition. *Human Sleep and Cognition: Basic Research*, *185*, 105.

Khitrov, M. Y., Laxminarayan, S., Thorsley, D., Ramakrishnan, S., Rajaraman, S., Wesensten, N. J., & Reifman, J. (2014). PC-PVT: a platform for psychomotor vigilance task testing, analysis, and prediction. *Behavior Research Methods*, *46*(1), 140–147.

Kieras, D. E., & Meyer, D. E. (1996). The EPIC architecture: Principles of operation. *Unpublished Manuscript from Ftp://Ftp.eecs.umich.edu/People/Kieras/EPICarch.ps*, (Journal Article).

Kieras, D. E., & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction*, *12*(4), 391–438.

Kieras, D. E., Wood, S. D., & Meyer, D. E. (1997). Predictive engineering models based on the EPIC architecture for a multimodal high-performance human-computer interaction task. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *4*(3), 230–275.

Kim, G. J. (2015). *Human–Computer Interaction: Fundamentals and Practice*. CRC Press.

Kingdon, R. (2008). *A review of cognitive architectures*. Tech. rep., ISO Project Report. MAC 2008-9.

Lane, N. D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., & Campbell, A. T. (2010). A survey of mobile phone sensing. *IEEE Communications Magazine*, *48*(9), 140–150.

Lim, J., & Dinges, D. F. (2008). Sleep deprivation and vigilant attention. *Annals of the New York Academy of Sciences*, *1129*(1), 305–322.

Lim, J., & Dinges, D. F. (2010). A meta-analysis of the impact of short-term sleep deprivation on cognitive variables. *Psychological Bulletin*, *136*(3), 375.

Lo, J. C., Groeger, J. A., Santhi, N., Arbon, E. L., Lazar, A. S., Hasan, S., … Dijk, D.-J. (2012). Effects of partial and acute total sleep deprivation on performance across cognitive domains, individuals and circadian phase. *PloS One*, *7*(9), e45987.

MacKenzie, I. S. (1992). Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction*, *7*(1), 91–139.

MacKenzie, I. S., Kauppinen, T., & Silfverberg, M. (2001). Accuracy measures for evaluating computer pointing devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 9–16). ACM.

MacKenzie, I. S., & Teather, R. J. (2012). FittsTilt: the application of Fitts' law to tilt-based interaction. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design* (pp. 568–577). ACM.

Magill, R. A., Waters, W. F., Bray, G. A., Volaufova, J., Smith, S. R., Lieberman, H. R., … Ryan, D. H. (2003). Effects of tyrosine, phentermine, caffeine D-amphetamine, and placebo on cognitive and motor performance deficits during sleep deprivation. *Nutritional Neuroscience*, *6*(4), 237–246.

McCauley, P., Kalachev, L. V., Mollicone, D. J., Banks, S., Dinges, D. F., & Van Dongen, H. (2013). Dynamic circadian modulation in a biomathematical model for the effects of sleep and sleep loss on waking neurobehavioral performance. *Sleep*, *36*(12), 1987–97.

Mottet, D., Guiard, Y., Ferrand, T., & Bootsma, R. J. (2001). Two-handed performance of a rhythmical Fitts task by individuals and dyads. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(6), 1275.

Murre, J. M. J., & Chessa, A. G. (2011). Power laws from individual differences in learning and forgetting: mathematical analyses. *Psychonomic Bulletin & Review*, *18*(3), 592–597. https://doi.org/10.3758/s13423-011-0076-y

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, *434*(7031), 387–391.

Nilsson, J. P., Söderström, M., Karlsson, A. U., Lekander, M., \AAkerstedt, T., Lindroth, N. E., & Axelsson, J. (2005). Less effective executive functioning after one night's sleep deprivation. *Journal of Sleep Research*, *14*(1), 1–6.

Nyamsuren, E., & Taatgen, N. A. (2013). Pre-attentive and attentive vision module. *Cognitive Systems Research*, *24*, 62–71.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, *25*(1), 46–59.

Perrott, D. R., Saberi, K., Brown, K., & Strybel, T. Z. (1990). Auditory psychomotor coordination and visual search performance. *Attention, Perception, & Psychophysics*, *48*(3), 214–226.

Pilcher, J. J., Band, D., Odle-Dusseau, H. N., & Muth, E. R. (2007). Human performance under sustained operations and acute sleep deprivation conditions: Toward a model of controlled attention. *Aviation, Space, and Environmental Medicine*, *78*(Supplement 1), B15–B24.

Remington, R. W. (1980). Attention and saccadic eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(4), 726.

Rodriguez-Jimenez, R., Avila, C., Garcia-Navarro, C., Bagney, A., de Aragon, A. M., Ventura-Campos, N., … others. (2009). Differential dorsolateral prefrontal cortex activation during a verbal n-back task according to sensory modality. *Behavioural Brain Research*, *205*(1), 299–302.

Rowland, L. M., Thomas, M. L., Thorne, D. R., Sing, H. C., Krichmar, J. L., Davis, H. Q., … Redmond, D. P. (2005). Oculomotor responses during partial and total sleep deprivation. *Aviation, Space, and Environmental Medicine*, *76*(Supplement 1), C104–C113.

Smith, A., & Maben, A. (1993). Effects of sleep deprivation, lunch, and personality on performance, mood, and cardiovascular function. *Physiology & Behavior*, *54*(5), 967–972.

Turner, T. H., Drummond, S., Salamat, J. S., & Brown, G. G. (2007). Effects of 42 hr of total sleep deprivation on component processes of verbal working memory. *Neuropsychology*, *21*(6), 787.

Van Dongen, H. P., Maislin, G., Mullington, J. M., & Dinges, D. F. (2003). The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *SLEEP-NEW YORK THEN WESTCHESTER-*, *26*(2), 117–129.

Williams, A., & Davids, K. (1998). Visual search strategy, selective attention, and expertise in soccer. *Research Quarterly for Exercise and Sport*, *69*(2), 111–128.

Williams, H. L., Lubin, A., & Goodnow, J. J. (1959). Impaired performance with acute sleep loss. *Psychological Monographs: General and Applied*, *73*(14), 1.

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238.

Zaal, F. T., & Thelen, E. (2005). The developmental roots of the speed-accuracy trade-off. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1266–1273.

Zar, J. H. (1999). *Biostatistical analysis*. Pearson Education India.