

Functional Equivalence of Sleep Loss and Time on Task Effects in Sustained Attention

Bella Z. Veksler

Glenn Gunzelmann

Air Force Research Laboratory

Wright-Patterson Air Force Base

Author Note

Bella Z. Veksler, Cognitive Models and Agents Branch, US Air Force Research Laboratory; Glenn Gunzelmann, Cognitive Models and Agents Branch, US Air Force Research Laboratory.

This work was performed while the first author held a National Research Council Research Associateship Award at Air Force Research Lab. This work was supported by the Air Force Office of Scientific Research (AFOSR) [grant number 09RH06COR].

Correspondence concerning this article should be addressed to Bella Z. Veksler, Air Force Research Laboratory, 2620 Q Street (Bldg 852), Wright-Patterson AFB, OH 45431. Contact: bellav717@gmail.com

Abstract

Research on sleep loss and vigilance both focus on declines in cognitive performance, but the literatures have been largely separate in the past. The current work explores whether the same mechanisms can account for the effects of both sleep loss and time-on-task on performance. A classic task used in the sleep deprivation literature, the Psychomotor Vigilance Test (PVT), was extended from a traditional 10 minute duration to 35 minutes, in order to make the task similar in duration to typical vigilance tasks. A computational cognitive model demonstrated that the effects of time-on-task in the PVT were functionally equivalent to those observed with sleep loss. Furthermore, the same mechanisms were able to generalize to a more traditional vigilance task - the Mackworth Clock Task. This supports the hypothesis that these different types of fatigue may produce functionally equivalent declines in performance.

Keywords: vigilance, fatigue, ACT-R, cognitive modeling, time-on-task, psychomotor vigilance.

Functional Equivalence of Sleep Loss and Time on Task Effects in Sustained Attention

Introduction

Sustaining attention in prolonged tasks is crucial in many areas of the workplace and in school, whether you're driving a car, working as a baggage screening operator, or just attempting to pay attention to a lecture. These kinds of tasks - often called sustained attention, or vigilance, tasks are well-known for leading to a performance decrement as time-on-task increases (e.g., Davies & Parasuraman, 1982). While there have been many theories put forth to account for the vigilance decrement, computational accounts are lacking. The goal of this work is to test the hypothesis that cognitive declines associated with the vigilance decrement are functionally equivalent to fluctuations related to time awake and circadian rhythms. We do this by evaluating the capacity of a computational theory of sleep loss to account for the vigilance decrement, and whether this account can explain the underlying changes in cognitive processing that are responsible for the phenomenon.

Mackworth (1948) describes vigilance as the “psychological readiness to perceive and respond” (p. 133). Being vigilant, therefore, manifests itself behaviorally in fast and accurate responses to stimuli. Traditionally, vigilance tasks require participants to monitor a stream of stimuli and respond when a rare signal appears, typically in the context of visual detection. Human ability to detect the signal declines with increasing time-on-task. In particular, there is a decline in performance that unfolds over the first 30-45 minutes of the task, where performance is defined as both time to respond and detection accuracy (N. H. Mackworth, 1948). In typical vigilance tasks, latency of response increases while number of correct responses (accuracy) decreases with prolonged time of performing a task.

In order to successfully model the vigilance decrement, we leverage an existing theory

and set of mechanisms developed to explain fluctuations in performance related to sleep loss and circadian rhythms. Whereas theories about the impact of sleep loss and theories of the vigilance decrement are largely distinct (but see Gunzelmann, Moore, Gluck, Van Dongen, & Dinges, 2010; Krueger et al., 2008), it is our goal to highlight connections between these areas of study and propose that the same mechanisms can account for the changes in performance that result from both sleep loss and prolonged time on task. Our proposal is that, although the degradations operate at different time scales (hours and days in the case of sleep deprivation and minutes to hours in the case of vigilance studies), they are functionally equivalent in terms of their impact on cognitive processing. This is analogous to the finding that sleep occurs at both the neuronal assembly and behavioral levels, and may be use dependent (Krueger et al., 2008; Van Dongen, Belenky, & Krueger, 2011).

Background

Over the past century, there has been considerable evidence in the literature that perceptual-motor detection and discrimination tasks suffer from a vigilance decrement (Bakan, 1955; N. H. Mackworth, 1948; Parasuraman, 1979). A classic study by Mackworth (1948) illustrates the type of task typically used to study the vigilance decrement. In the Mackworth Clock Task, participants monitor the forward ticks of a clock hand to determine when the tick is twice the usual magnitude. Vigilance decrements have also been studied with tasks such as determining when a stimulus on a radar display changes (Adams & Boulter, 1960), detecting luminance changes (Bakan, 1955), or identifying an imminent collision on simulated radar displays (Hitchcock et al., 1999). The common thread through all of these tasks is that the signal requiring a response is rare and experimental results indicate that individuals find it increasingly difficult to detect signals as time on task increases.

Currently the most widely accepted theory associates the vigilance decrement with the depletion of information-processing resources that can not be replenished adequately during task performance (Parasuraman, 1979; Warm, Parasuraman, & Matthews, 2008). Support for this view comes from studies in which workload during a vigilance task is measured using the NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988). Workload scores increase with time-on-task mirroring performance declines, especially in tasks where memory is taxed, suggesting that *Vigilance Requires Hard Mental Work and Is Stressful* (Warm et al., 2008). The resource account is useful in describing some of the behavioral changes observed in vigilance tasks. In this paper, we describe a computational theory of the vigilance decrement that instantiates a resource account and provides a detailed explanation of the underlying changes in cognitive processing that lead to declines in performance during vigilance and sustained attention tasks.

There has been descriptive mathematical modeling of the vigilance decrement, characterizing how performance declines with increasing time on task (Fisk & Schneider, 1981; Giambra & Quilter, 1987; J. F. Mackworth, 1964; Teichner, 1974). These models have found that the vigilance decrement is best fit by an exponential curve with a quick initial drop and subsequent asymptote. It is the goal of the current work to associate those dynamics with underlying changes to information processing in the cognitive system that lead to the performance decrement, and to evaluate whether those mechanisms match those that have been previously validated in the context of sleep loss (e.g., Gunzelmann et al., 2009). This will provide a more detailed theoretical account, which has the potential to generalize to other tasks and dependent measures.

Our approach is to use a computational cognitive architecture, which combines many

cognitive components (e.g. central cognition, memory, vision, motor control) in a running simulation, which collectively represent a theory of the human cognitive system. Utilizing a cognitive architecture, Gunzelmann, Gross, Gluck, and Dinges (2009) developed a mechanistic account of the effect of fatigue on information processing. According to the model, performance decrements observed in sleep deprived individuals are a result of disruptions to goal-directed processing brought on by a breakdown in the process of selecting and executing cognitive actions within the central cognitive component of the architecture. A primary goal of the current work is to show that the same theoretical mechanisms can account for the vigilance decrement. The value in using a computational cognitive model embedded in a running simulation is that it provides more than a description of existing empirical results. The cognitive account emerges out of the interaction of multiple cognitive components in the architecture, providing a process level account of the effect of fatigue on various components of cognition.

In the remainder of the paper we will demonstrate that the fatigue mechanism developed in the context of sleep deprivation is capable of accounting for the vigilance decrement. This will be done by comparing the model to human performance data on an extended version of a task typically used in sleep research and by testing the mechanism on a classic vigilance task.

Experiment 1

The sleep research community has been using the Psychomotor Vigilance Task (PVT) for several decades to explore the effects of sleep restriction/deprivation on performance in a stimulus-response task (Dinges & Powell, 1985; Van Dongen & Dinges, 2005). Participants in this task must respond as quickly as possible when an incrementing counter appears on the screen. Typically, the PVT lasts for 10 minutes (Dinges & Powell, 1985). Gunzelmann et al.

(2010) show how mechanisms developed to account for the effects of sleep loss on PVT performance can also capture changes that unfold during a 10-minute session. Although this model captures degradations in performance over the course of 10 minutes, it has never been tested in a longer task setting. Traditional vigilance tasks typically span more than 30 minutes (Davies & Parasuraman, 1982). To assess the model's ability to capture changes over a longer period of time, we presented participants with a 35-minute PVT task to align the task duration with traditional vigilance tasks as an initial step.

Participants

A total of 20 University of Dayton undergraduate and local area participants completed the study (12 Female) with a mean age of 22.35 ($SD=2.45$). This number of participants was selected because it is commensurate with previous vigilance and sleep research studies. Participants were paid \$15 for their participation. All participants had normal or corrected to normal vision. The protocol was approved by the Institutional Review Board of the Air Force Research Laboratory. All participants provided written informed consent.

Materials

A 17-inch monitor with the resolution set to 1280x1024 pixels displayed a grey background. Participants continuously monitored the computer screen for the appearance of an incrementing counter and responded by pressing a button on a button box as quickly as possible. The counter was a four-digit number indicating the time in milliseconds since its appearance. It was drawn in 80-point black Arial font in the middle of the screen and appeared after a uniformly random delay between 2-10s from the previous trial's response. The response had to be made within 30 seconds of the counter appearing to be counted as a valid response, otherwise the trial timed out (termed a *sleep attack* or *non-response* in previous PVT

literature). There were no *non-responses* observed in this study. Once a response was made, the counter stopped incrementing and displayed the participant's reaction time as feedback for 500ms. The dependent measure of response time was recorded in milliseconds. Participants completed 340 trials. The entire task took about 35 minutes to complete.

Because PVT performance is sensitive to sleep loss, we asked participants about their sleep history prior to beginning the task. The number of hours slept the night before and the time they woke up that morning were recorded prior to performing the task. In general, participants had between 5-10 hours of sleep the night before ($M=7.63$, $SD=1.31$) and woke up between 7:30am and 12pm ($M=8:57AM$, $SD=1.21$). Caffeine has been shown to alter performance on attention intensive tasks (Kilpeläinen, Huttunen, Lohi, & Lyytinen, 2010; Temple et al., 2000; Van Dongen et al., 2001), although these effects are generally not defined quantitatively in existing models of alertness (but see Benitez, Kamimori, Balkin, Greene, & Johnson, 2009; Ramakrishnan et al., 2013). Out of our sample of 20, 5 participants reported consuming coffee on the day of the study. We did not find a significant effect of coffee on the trend over blocks in median response time, lapses, or false starts and so it is not considered further in the analyses that follow. Participants were tested between the hours of 9am and 5pm, Monday through Friday.

Results

PVT performance is characterized by 4 measures: false starts, alert responses, lapses, and sleep attacks (non-responses). Participants who are sleep deprived show an increase in *lapses* (responses longer than 500ms during individual sessions), an increase in *false starts* (responses faster than 150ms), and the response time distribution shifts to the right, with the slope of the decline increasing with fatigue. Sleep attacks are typically found only after

significant sleep deprivation (Dinges & Powell, 1985; Doran, Van Dongen, & Dinges, 2001) and no sleep attacks were observed during the current study. Therefore, only *false starts*, *alert responses*, and *lapses* are considered in subsequent analyses.

Median Response Time. The data were divided into seven 5-min blocks. For each participant the median of all response times within each block, excluding false starts, was used in the analysis. Using this measure, there was a significant increase in median response time as time-on-task increased, $F(6, 114) = 8.439$; $p < .001$. The median response times for the seven 5-min blocks are presented in **Table 1**. We compared our results to PVT results from a 62 hour sleep deprivation study (Tucker, Whitney, Belenky, Hinson, & Van Dongen, 2010) and found that the median RT for the last 5-min block in our study was commensurate with the median RT of participants halfway through a 10 minute PVT after a full night of sleep deprivation.

Percent Lapses. The percentage of responses characterized as lapses (response times longer than 500ms) also increased across blocks, $F(6, 114) = 5.435$, $p < .001$. **Table 1** contains the percentage of lapses for the seven 5-min blocks. For comparison, the percentage of lapses for participants in a recent sleep deprivation study was about 1.4% at baseline across the 10 minutes whereas the percentage of lapses after a full night of sleep deprivation started at 4.7% in the first minute and rose to 11.9% in the last minute of the 10 minute PVT (Tucker et al., 2010).

Table 1.

PVT Results with 95% Confidence Intervals

5 Minute Block	Response Time (ms)	% Lapses		% False Starts	
	Median	M (SD)	95% CI	M (SD)	95% CI
1	285	1.18 (1.72)	[0.42,1.93]	0.63 (1.00)	[0.2,1.07]
2	290	1.67 (2.08)	[0.76,2.58]	0.31 (0.75)	[-0.02,0.63]
3	293	1.85 (3.25)	[0.43,3.27]	1.26 (2.98)	[-0.04,2.56]
4	302	2.46 (2.54)	[1.34,3.57]	1.53 (2.79)	[0.31,2.75]
5	307	3.78 (4.71)	[1.71,5.84]	1.74 (3.49)	[0.21,3.27]
6	308	2.93 (2.80)	[1.7,4.16]	1.53 (2.98)	[0.22,2.83]
7	312	5.29 (4.83)	[3.18,7.41]	2.07 (4.52)	[0.09,4.06]

Percent False Starts. There was not a significant change in mean percent false starts by 5-min blocks across time-on-task, $F(6, 114) = 1.459$, $p = .19$. While not significant, there appears to be a trend toward an increasing percentage of false starts over the seven 5-min blocks (see **Table 1**). For comparison, the percentage of false starts at baseline in Tucker et al. (2010) was about 1.5% and rose to 2.6% after one night of sleep deprivation.

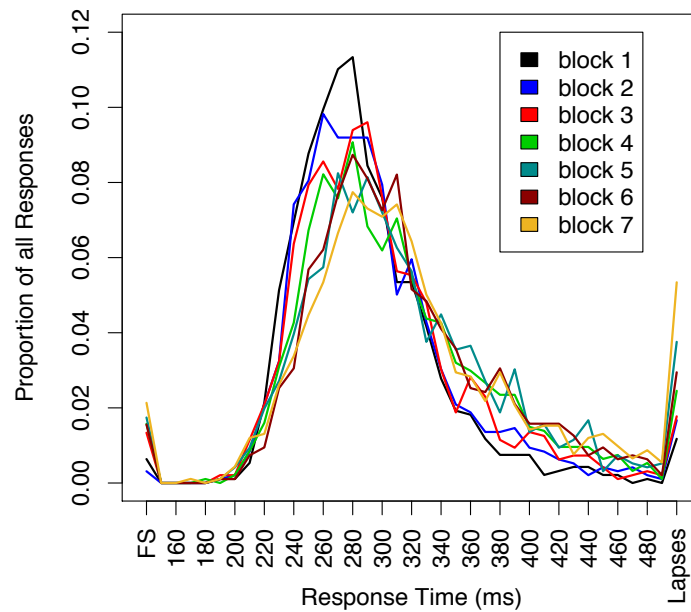


Fig. 1.

PVT human response time distributions from each 5 min block.

RT Distribution. As a measure of central tendency, median response times provide a concise characterization of the changes with time-on-task. However, a response time distribution gives us a more precise picture of how performance changes. Fig. 1 presents probability density functions (PDFs) for each of the 5-minute blocks in the data. As time-on-task increases, the response time distribution gradually stretches to the right, reflecting the slowing of responses and increase in lapses.

Whereas PDFs provide a helpful way of viewing the data, cumulative distribution functions (CDFs) offer some advantages for comparing human performance to model results. CDFs have been used in the past for fitting response time distributions helping to account for the positive skew characteristic of response times (Van Zandt, 2000). The CDF specifies how much of the data in the distribution falls below each predetermined boundary. For purposes of

these data, we used the data from all participants for the first 5-minute block to determine the 5% quantile boundaries of the response time distribution for all responses made within 150ms to 30s (**Table 2** lists the specific values). These boundaries were then used to determine how much of the distribution fell below each boundary in each subsequent block. We incorporated *false starts* into the CDF by including the percentage of *false starts* as an additional quantile. Fig. 2 plots the Cumulative Distribution Functions of all 7 blocks of the human data by using the quantile boundaries from the first block. Note that as time on task increases (block number increases), the cumulative percentage of RTs falling below each boundary decreases, illustrating the shift of the distribution. Both the PDF and the CDF provide similar information about the response time distribution. However, the CDF provides a concise description, is monotonic, and captures the full distribution which makes comparison across blocks and between model and human data much more straightforward.

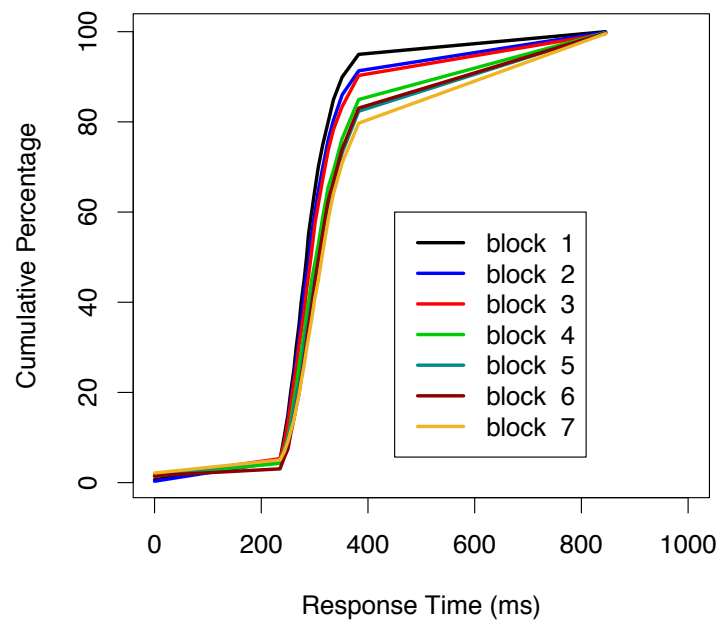


Fig. 2.

PVT human cumulative distribution function. Blocks 2-7 use the quantile boundaries from block 1.

Table 2.
Quantile Boundaries of Response Time Distribution in PVT (in ms) from first block of Human Data.

Cumulative %	Boundary (ms)
5%	235
10%	243
15%	250
20%	255
25%	261
30%	266
35%	271
40%	275
45%	280
50%	285
55%	288
60%	294
65%	301
70%	307
75%	315
80%	325
85%	335
90%	352
95%	383
100%	845

Discussion

The 10-minute Psychomotor Vigilance Task (PVT) has been used extensively in the sleep research community to assess the effect that sleep deprivation has on cognitive performance. Within as little as one night of sleep deprivation, the percent of lapses and false starts increases dramatically (Doran et al., 2001). To bridge the gap between the sleep research and vigilance research communities, we extended the PVT to run for 35 minutes, a duration similar to most vigilance studies. The results from the 35-minute PVT demonstrate that performance degrades to a degree similar to declines seen in the 10-minute version of the PVT after a full night of sleep deprivation. This suggests that extended time on task may produce similar degradations in cognitive function to those seen with sleep loss. To further support this claim, the next section describes a model that performs the 35 minute PVT using the same modeling framework and mechanism as that previously applied to the 10 minute PVT in sleep deprived participants (Gunzelmann et al., 2009).

Modeling Framework

Cognitive architectures based on decades of psychological theory have successfully reproduced many effects found in the psychology literature (Anderson et al., 2004). To model human performance on the PVT, we used the adaptive control of thought-rational (ACT-R) cognitive architecture (Anderson et al., 2004). ACT-R is a theory of the mechanisms of human cognition, implemented as a running simulation. Hundreds of models have been developed in ACT-R, which provide quantitative accounts for a wide variety of tasks (e.g., Anderson et al., 2004)¹. ACT-R is a modularized production system with visual and motor modules to embed

¹ Models and their corresponding publications can be found at <http://act-r.psy.cmu.edu/publication/>

it in a task environment. Its central production system interfaces with other components of the cognitive system through limited capacity buffers that allow information to be passed in both directions. ACT-R's ability to interface with the same environment as humans while providing behavioral predictions based on cognitive constraints make it an excellent framework for modeling many phenomena, including the vigilance decrement.

The production system representing central cognition is driven by procedural knowledge that is represented by productions, or if-then rules, in ACT-R. Central cognition operates through the selection and execution of these productions based on utility values. During the production selection process, a utility value is computed for each production based upon its history of use and match to the current context. All productions whose preconditions match the state of the cognitive system (i. e., contents of buffers) are placed in a *conflict set*. Each production's utility is then calculated based on its previous utility value and reinforcement learning. To be selected, a production must have a utility value that exceeds the utility threshold (UT) and is greater than any other production's utility in the *conflict set*. Typically, the utility threshold remains static throughout model execution while production utilities are recalculated during each production cycle and compared both to each other and to the UT.

Production Partial Matching

Before we discuss the fatigue mechanism, there is one ACT-R mechanism that influences production selection, which plays a critical role in the current research. The production partial matching mechanism allows productions to be selected even if the values in the buffers are not exactly those specified by the production. This allows the model to make errors of commission, and generate *false starts*.

A scaling parameter, *ppm* dictates the degree of discounting of the utility when the

current contents of the buffers don't match the production. The equation that utilizes this parameter is:

$$U_i(t)' = U_i(t) + \sum_j^N ppm * similarity(d_j^i, v_j^i) \quad (1)$$

where $U_i(t)$ is production i 's current utility value prior to production partial matching, ppm is the scaling parameter, d_j^i is the desired value for slot j in production i , and v_j^i is the actual value for slot j in production i . The function computing the similarity between d_j^i and v_j^i is assumed to be 0 when the expected and actual values are identical, and a negative value when they differ. This newly computed utility value is then modulated by the fatigue mechanisms described in the next section.

Fatigue Mechanisms

One of the benefits of using ACT-R as the modeling framework is that it is simple to extend its functionality by adding additional modules to the architecture to extend its explanatory power. Previous work within the ACT-R architecture has resulted in the addition of a fatigue module to make predictions about performance degradations as a result of sleep deprivation (Gunzelmann et al., 2009). The module modifies three key components of the architecture: the behavior of the model when no productions have utilities above threshold, utility calculations, and the value of the utility threshold.

ACT-R's default behavior is to halt model execution if no productions have utilities that rise above the utility threshold. The fatigue module allows model execution to continue after a production execution cycle during which no productions are executed. These gaps in cognitive processing are referred to as *microlapses* (Gunzelmann et al., 2009). They are brief (approximately 50 ms), but can accumulate to produce substantial changes in cognitive

performance. *Microlapses* result in delays in information processing, manifesting behaviorally as slowed response times and, for time-critical tasks, missed responses.

The probability of *microlapses* is influenced by the utility values of the productions and the utility threshold. The fatigue module diminishes both values as alertness declines. Lower utility values lead to more *microlapses*. A simultaneous compensatory mechanism lowers the utility threshold, which reduces the probability of *microlapses*. At the same time, this compensation increases the probability of executing lower-utility options as stochasticity plays a larger role, increasing the likelihood of incorrect responses.

As per Gunzelmann et al. (2010), the fatigue mechanism modifies the production's utility as follows:

$$Utility(t) = FP(t) * U_i(t) + \epsilon \quad (2)$$

where $U_i(t)$ is the non-fatigued utility value at time t and $FP(t)$ is the fatigue modulation value at time t . Logistically distributed noise is subsequently added to produce stochasticity in model performance, denoted as ϵ .

In Gunzelmann et al. (2010), the fatigue module was validated within participants whose sleep histories were closely controlled and who performed the same tasks across numerous sessions over the course of several sleep deprived or sleep restricted days. However, participants in the current study only performed the PVT during one session and self-reported their sleep histories. Therefore, we omitted the component of the FP calculation that relies on sleep history and circadian rhythm information. Variability in initial performance stemming from sleep history, motivation, and other factors is absorbed into estimates of the Initial Utility ($U_i(0)$) for each individual. For purposes of this paper, $FP(t)$ is composed solely of the time-on-task component and calculated in the fatigue module as follows:

$$FP(t) = t^{PU_{TOT}} \quad (3)$$

where t is the amount of time (in seconds) since the task began and PU_{TOT} is a parameter that reflects the production utility time-on-task decline. At the beginning of task execution, FP is 1 and gradually declines as time on task elapses, which causes production utilities to decrease. The structure of the equation is informed by research in the vigilance literature that has suggested performance declines exponentially with time-on-task (Giambra & Quilter, 1987).

In addition to the changes in production utility, the fatigue module also alters the utility threshold. The utility threshold dictates the minimum utility value a production must have in order to be selected. Without the fatigue module, this is set as a constant parameter and does not change with model execution. To represent increased effort to compensate for feeling tired, the fatigue module modifies the utility threshold. This is done in a manner similar to the production utility modulations as follows:

$$FT(t) = t^{UT_{TOT}} \quad (4)$$

where UT_{TOT} is the utility threshold time-on-task decline parameter. This value is then used to modulate the initial Utility Threshold set by the modeler, $UT(0)$. The setting of $UT(0)$ is intended to capture any sleep or motivational factors.

$$UtilityThreshold(t) = FT(t) * UT(0) \quad (5)$$

The fatigue mechanisms, together with the production partial matching mechanism, modify model performance across time-on-task in accordance with the changes seen in sleep studies. We next demonstrate its efficacy by modeling both the extended PVT and a more traditional vigilance task, the Mackworth Clock Task.

Model of the PVT

The model described here varies in some respects from the model in Gunzelmann et al.

(2009). Changes were made to maintain currency with the evolution of the ACT-R architecture. The mechanisms remain functionally equivalent and produce identical fits to human data. The original model, which ran for 10 minutes, was extended to run for 35 minutes to reflect the duration of the task performed by human participants in the current study.

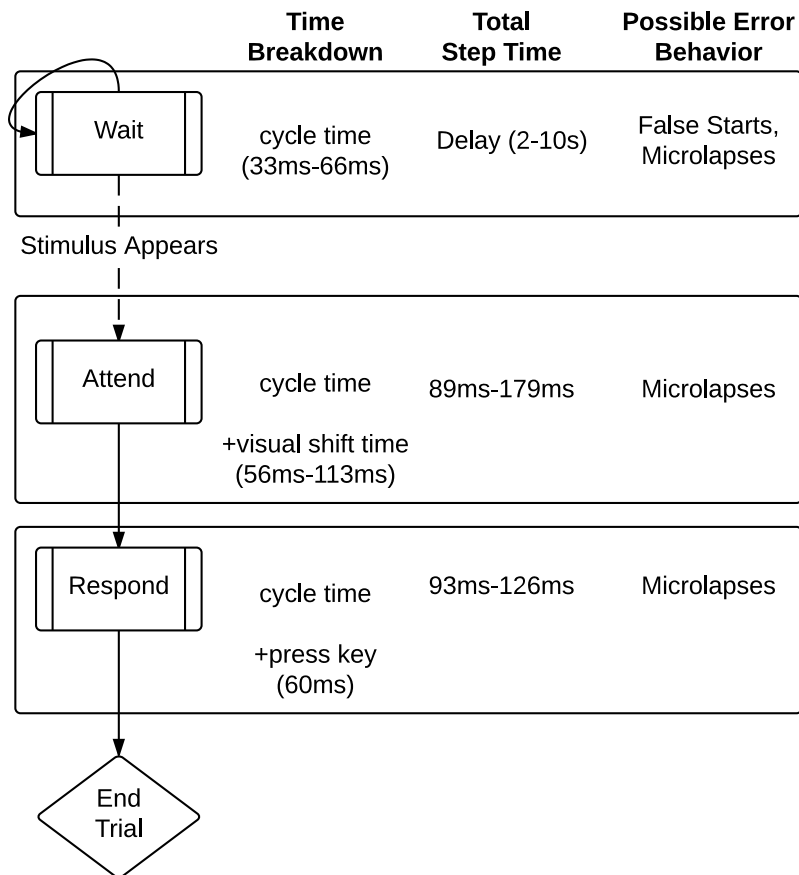


Fig. 3.

Model of the PVT. Indicates the amount of time each production takes to execute and the possible implications of the fatigue module at each step.

The model is composed of three productions: *wait*, *attend*, and *respond* as illustrated in Fig. 3. Initially, while no stimulus is present on the screen, the model executes the *wait*

production. Once the stimulus appears, the model has to shift its attention to the stimulus source. This is achieved by executing the *attend* production which sends a request to the visual module to shift attention to the stimulus. Once the attention shift has occurred, the *respond* production executes and initiates a *press key* action in the motor module.

Fig. 3 illustrates the general flow of the model. As there is noise introduced into cycle time and attention shift time, the approximate amount of time each cognitive action requires to complete is depicted. Note that the upper bound on the total time does not incorporate time for *microlapses*, during which productions do not reach threshold and fail to execute on a given cycle. *Microlapses* can occur during the conflict resolution phase in any step of model execution, which would interfere with production execution. Such disruptions become more prevalent as time-on-task increases.

False starts have the potential to occur while the model is waiting for the stimulus to appear. This is a direct result of the partial matching mechanism, with the *ppm* parameter set to twice the Initial Production Utility (IU). The similarity value between a ‘wait’ state and a ‘respond’ state is set to -.5, which means that the *respond* production’s utility during the waiting period is $U_i(t) = IU - (2 * IU) * (.5) + \epsilon = \epsilon$. *False starts* are thus a direct result of stochasticity introduced to the utility calculation. This becomes more pronounced as the utility threshold decreases with time-on-task.

Model Evaluation

We explored the performance of the model by varying four parameters involved in calculating the utility and utility threshold, PU_{TOT} , UT_{TOT} , $UT(0)$ and IU , the ranges of which are shown in **Table 3**. Each parameter set was run 1000 times in order to achieve stable performance estimates. The MindModeling@Home <http://mindmodeling.org/> resource was

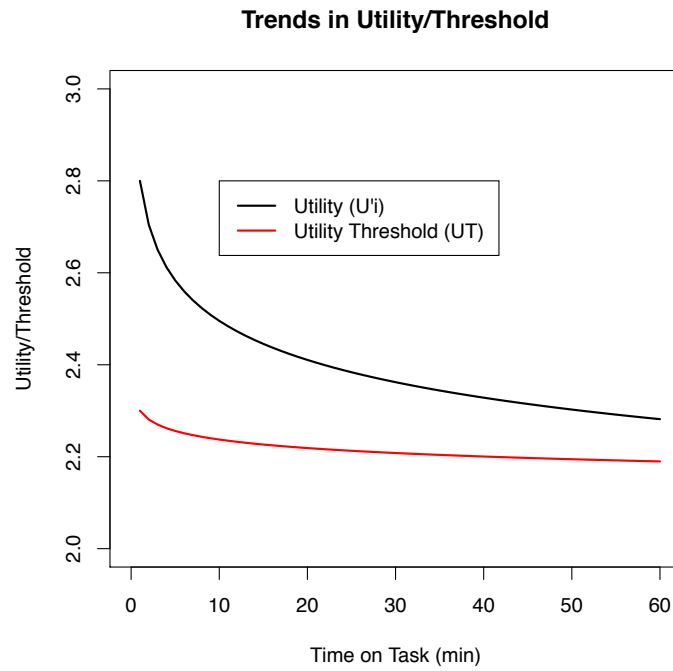
used in both the PVT and Mackworth model fitting (Harris, 2008).

Table 3.

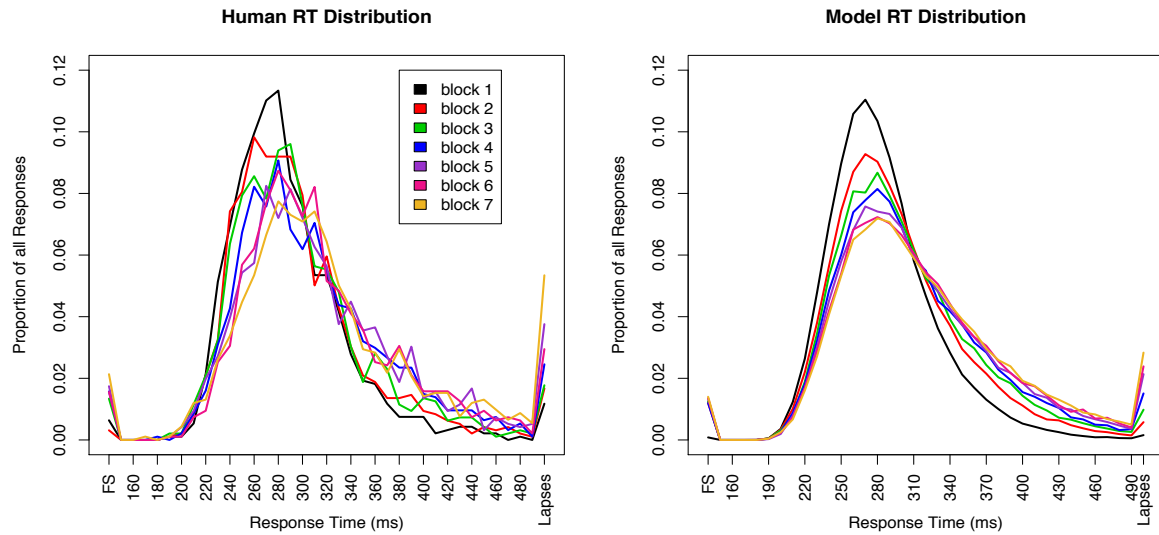
Parameter space run and resulting best fitting parameters for all human data in the 35-min PVT model.

Parameter	Range	Best Value
PU_{TOT}	-.01 to -1	-.05
UT_{TOT}	-.01 to -1	-.012
Initial Utility	1 to 3	2.8
Initial Utility Threshold	1 to 3	2.3
CDF RMSE (R^2)	0.001 (0.99)	
RT Distribution RMSE (R^2)	0.0054 (0.96)	

The model's performance was evaluated based on fit to the Cumulative Distribution Functions (CDFs) for each 5-min block with the quantile boundaries based on the first block of the human data. We compared human and model CDFs across all blocks and derived both Root Mean Square Error (RMSE) and R^2 values for each parameter set. The best fitting parameters are listed in **Table 3** and the graphical representation of the changes in both Utility Threshold and Utility as a function of time-on-task is depicted in Fig. 4. The most pertinent aspect is the relationship between Utility and Utility Threshold – as time on task increases the difference between the two values decreases, leading to lower probabilities of successfully executing appropriate actions and allowing for stochasticity in the model to drive behavior to a larger degree. As can be seen both qualitatively in Fig. 5 and quantitatively in **Table 3**, the resulting fits to the extended PVT data were very good.

**Fig. 4.**

Trends of production utilities and utility threshold of best fitting model parameters for PVT model (prior to adding noise).

**Fig. 5.**

Human and model RT distributions for all blocks.

Traditional Vigilance Tasks vs. PVT

We conducted an empirical study of a prolonged PVT to bridge the gap between the sleep deprivation and vigilance literatures. The degradations in performance in the prolonged PVT are akin to drops in performance seen after a night of sleep deprivation, suggesting the same mechanism may be responsible for both. Furthermore, the degradation in performance appears similar to results from other vigilance tasks and studies. However, while the 35 minute PVT is more commensurate with typical vigilance tasks in its duration, there remain critical differences relative to traditional vigilance tasks.

First, participants in traditional vigilance studies must discriminate between signals and non-signals. In contrast, stimulus discrimination is not required in the PVT as each time the counter appears on the screen, a response must be made.

Second, signals are rare in a traditional vigilance task, generally about 3% of stimuli. In the PVT each stimulus presentation is a signal, resulting in a high presentation rate (around 80 signals per 10-min session). The low signal rate in traditional vigilance tasks is the primary factor influencing the decline in performance over time (only about 10-20 signals within a 30-min session). Generally speaking, lower signal rates lead to larger decrements (Davies & Parasuraman, 1982; Parasuraman, 1979).

Third, although signals are rare in traditional vigilance tasks, stimuli are typically presented at a rate of about 1 per second. Combined with a low signal rate, the high presentation rate of non-signals exacerbates the decrement (Lanzetta, Dember, Warm, & Berch, 1987; Wiener, 1977). On the other hand, during the PVT, stimuli are presented every 2-10s, which is a much lower stimulus presentation rate.

Finally, response time is the only dependent measure provided by the PVT, as each

stimulus requires a response. There is no measure of accuracy as defined by correctly identifying a signal, though lapses are sometimes considered errors in the task (Lim & Dinges, 2008). However, traditional vigilance tasks often report response time and accuracy measures, which may be differentially affected by time-on-task (Adams, Stenson, & Humes, 1961).

In summary, there are 4 significant differences between the PVT and traditional vigilance tasks: 1) whether stimulus discrimination is required, 2) signal probability, 3) stimulus rate, and 4) dependent measures. These differences leave questions unanswered about the generalizability of the fatigue mechanism to account for the vigilance decrement. We thus turned to a classic vigilance task, the Mackworth Clock Task (N. H. Mackworth, 1948), which address these differences.

Mackworth Clock Task

The Mackworth Clock Task (MCT) is the seminal task in the vigilance literature, run originally in the 1940's (N. H. Mackworth, 1948) and replicated in numerous studies (e.g., Giambra & Quilter, 1987; Kaida, Akerstedt, Kecklund, Nilsson, & Axelsson, 2007; Kass, Vodanovich, Stanny, & Taylor, 2001; Lichstein, Riedel, & Richman, 2000). Recent instantiations of the MCT have been used to generate datasets that are better suited for model comparison (e.g., Giambra & Quilter, 1987; Lichstein et al., 2000). For instance, Mackworth originally only reported average accuracy of 25 participants for each 30-min block in a 2-hr session. Giambra and Quilter (1987), however, ran the study with 613 participants to obtain both accuracy and response time data for each of the 23 critical signals over the course of a 62 minute study. We compared our model's performance to the mean values of response times and hits from the human data reported by Giambra and Quilter (1987). The method and results are briefly described below.

Giambra and Quilter (1987)

Giambra and Quilter (1987) replicated Mackworth's Clock Task study using a physical metal box with a plain white face and a single black pointer which moved in 100 discrete steps around the face. Over the course of 62 minutes, participants had to watch the pointer and detect an infrequent double step (the critical signal). This occurred 23 times over the course of the study at 2, 4, 7, 11, 13, 15, 19, 21, 23, 26, 30, 32, 34, 37, 41, 43, 45, 49, 51, 53, 56, 60, and 62 minutes. This fixed timing of critical signals across all participants was an artifact of the original Mackworth study, which Giambra and Quilter (1987) replicated in their study design. One benefit of this consistency is that participants' data can be aggregated for each critical

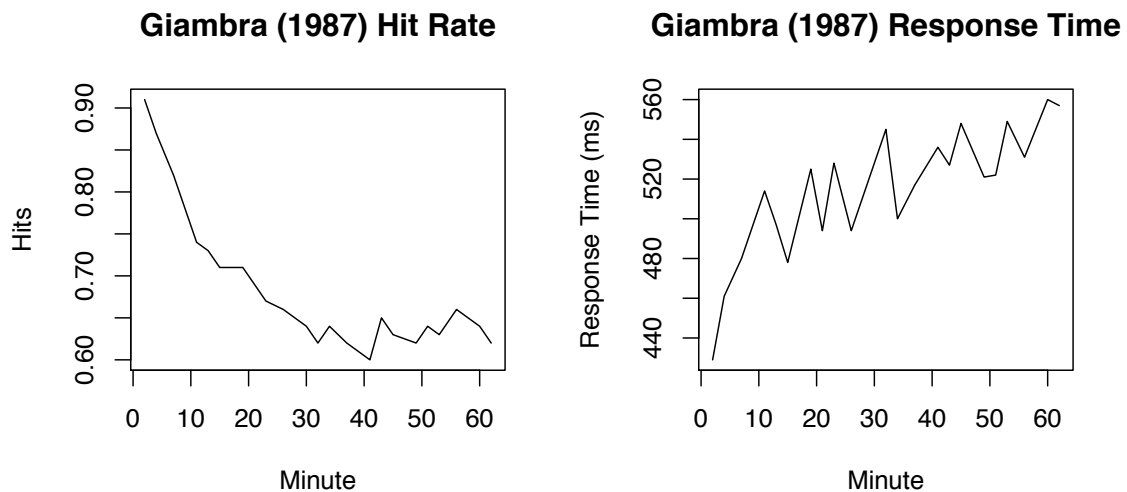


Fig. 6.

Human data from the Mackworth Clock Task (adapted from Giambra & Quilter, 1987).

signal. A second benefit is that we were able to present the model with identically timed signals. Human data from the study is presented in Fig. 6, based on the original publication (Giambra & Quilter, 1987). Giambra and Quilter (1987) reported that few subjects had false

alarms and hence mean values were unreliable.

8.2 Model

Similar to the PVT model, the computational cognitive model for the Mackworth Clock Task is composed of 3 productions: *wait*, *attend*, and *respond*. The *wait* production executes between screen updates. The *attend* production executes when there is a screen change - when the clock hand moves to a new location around the clock. Visual attention is then directed to the location to which the clock hand moved. The *respond* production executes when the critical signal is detected - in this case a hand movement of 2 ticks.

ACT-R's visual system does not actually sense the visual signal of the clock hand movement as a human eye would with the retina, so it was necessary to present the stimuli to ACT-R in noisy symbolic form. To reflect the noise in the sensing of the stimulus (the clock hand movement), we estimated sensitivity from the human data using signal detection theory (Green & Swets, 1966). While we realize that this was not a pure measure of visual sensitivity, it was the closest empirical measure possible in this context. The magnitude of clock hand movement of 2 ticks was drawn from the normal distribution with a mean of 2 and a standard deviation estimated from the human data at the beginning of the study. Giambra and Quilter (1987) did not report sensitivity information for participants performing the MCT, so we turned to Lichstein et al. (2000) who ran a computerized version of the Mackworth Clock Task and reported d' (sensitivity) for each half hour block, while also replicating Mackworth's accuracy results. Lichstein et al. (2000) indicated that in the first block of 30 minutes, the sensitivity value was 3.97, whereas on the second block it dropped down to 3.52. The relationship between sensitivity, d' , and the standard deviation, σ , of the signal distribution can be computed as:

$$d' = (\mu_s - \mu_n) / \sigma_n \quad (6)$$

where μ_s is the mean of the signal distribution (in this case, 2 ticks) and μ_n is the mean of the noise distribution (in this case, 1 tick). Therefore, in our case, $d' = 1/\sigma_n$.

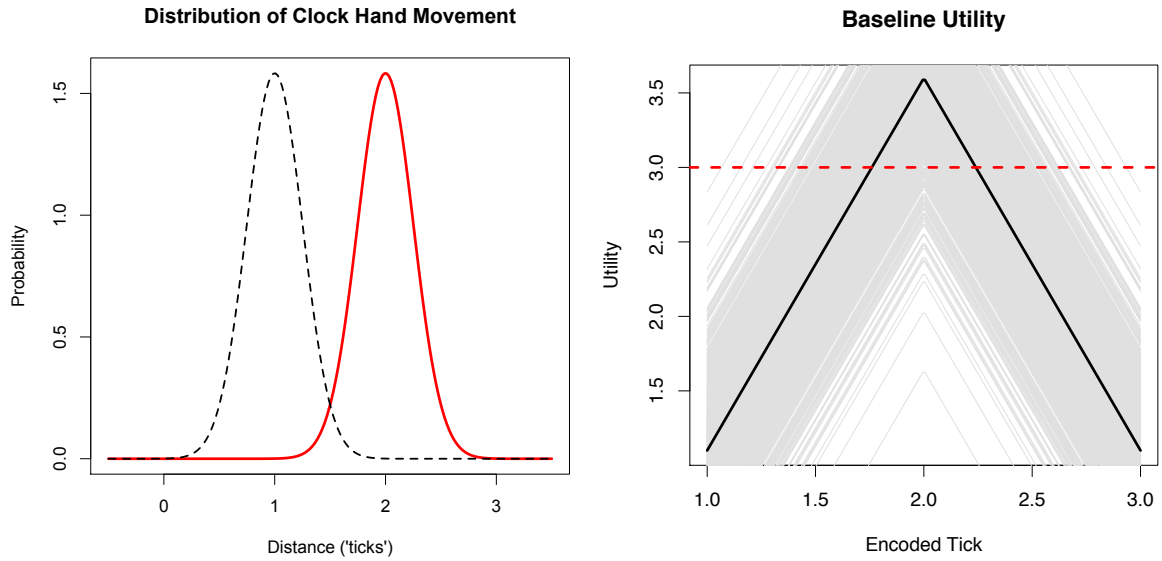


Fig. 7.

Left panel: Distribution of values encoded by model around actual clock hand movement ($\sigma = .252$). Right panel: Utility value as a function of encoded 'tick' value. Red dotted line indicates the initial utility threshold prior to any time-on-task effects.

The distribution of values used in the model had a mean equal to either 1 or 2 depending on the true clock hand movement and a $\sigma = .252$ around each of the 'tick' values. This value for the parameter was found by calculating the d' obtained from the human data in the first block of 30 minutes and applying it in Equation 6 (see Fig. 7, left).

Production partial matching is leveraged in this model to account for the noise in the visual system in encoding the tick magnitudes. This allows the *respond* production to execute when the degree of movement is not exactly 2 ticks, which will generally be the case because

of noisy encoding of the stimulus. The *ppm* parameter specifies the degree to which the production's utility is discounted whenever the value is not exactly 2. As ACT-R does not provide a default value for *ppm*, the space of reasonable values was explored (in the range of 1 to 5, step size of .05). A *ppm* value of 2.5 produced the best results in anchoring the model to the first data point in the human data. The production partial matching mechanism requires a similarity calculation between the desired clock hand value (2) and the actual fuzzy signal in the visual buffer. This was defined as the negative of the absolute difference. Fig. 7 (right) depicts the utility value as a function of the encoded 'tick' value with the *ppm* parameter set to 2.5 (the noise-free value is indicated by the black line). The gray bands indicate possible range of utility values once ACT-R noise is added. The utility threshold of 3 (the best parameter setting for this data, as discussed below) is included for comparison.

8.3 Model Evaluation

The parameters were systematically varied in precisely the same manner as in the PVT, using slightly expanded parameter ranges. The results are reported in **Table 4**. Each parameter set was run 3000 times in order to increase confidence in the performance estimates. For each parameter set, the 23 accuracy scores and 23 average RT values were recorded and compared with the human data to obtain RMSE and R^2 values.

Table 4:

Parameter space explored and best fitting parameters for 62-minute Giambra & Quilter model.

Parameter	Range	Best Value
PU_{TOT}	-.01 to -1	-.21
UT_{TOT}	-.01 to -1	-.12
Initial Utility	1 to 5	3.6
Initial Utility Threshold	1 to 5	3
CDF RMSE (R^2)	0.03 (0.86)	
RT Distribution RMSE (R^2)	0.06 (0.75)	

Fig. 8 plots the relationship between Utility Threshold and Utility using the best parameter set (based on minimizing the RMSE between human and model data in both accuracy and RT and ensuring that performance reached asymptote in the last 20 minutes on task). Again, it is the relative difference between the PU_{TOT} and UT_{TOT} parameters that primarily affects model behavior. The best fitting model is plotted in Fig. 9 along with the human data for comparison. In general, the model fit was good in terms of both accuracy ($R^2=.86$) and response time ($R^2=.75$).

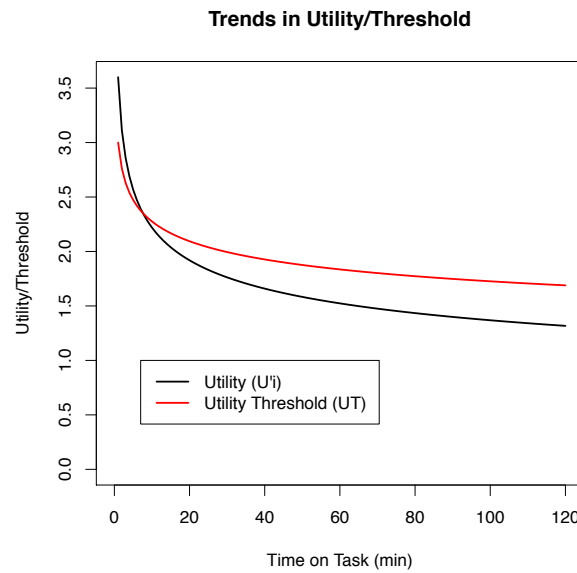
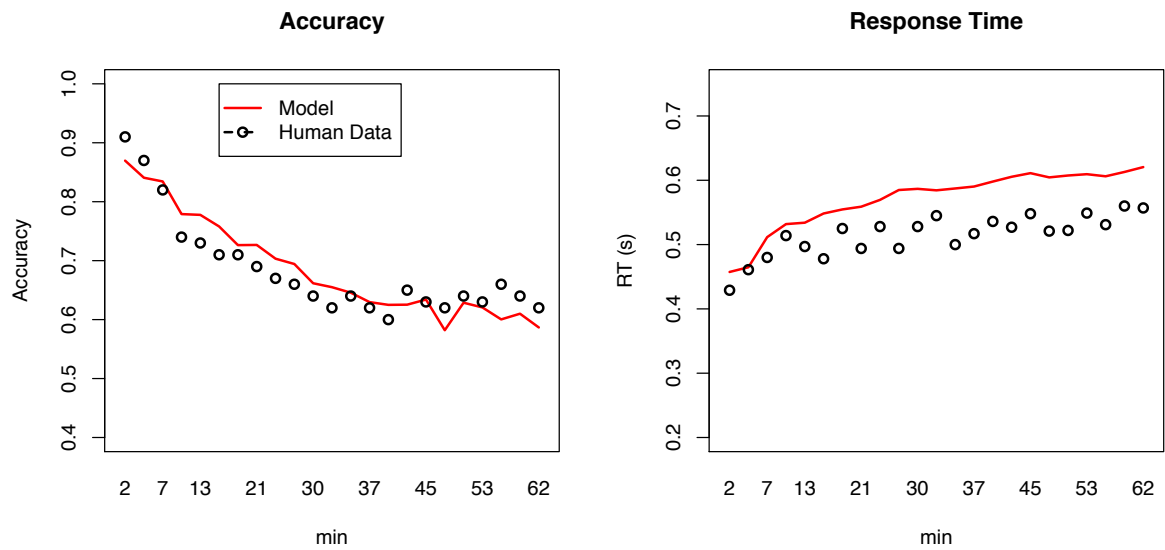


Fig. 8.

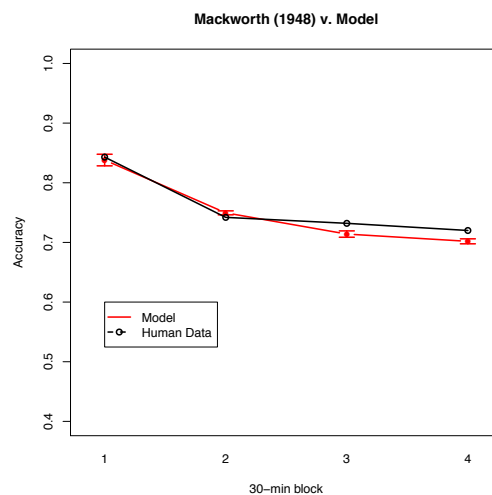
Trends of production utilities and utility thresholds of best fitting model parameters for Giambra & Quilter (1987) model.

To test whether this model would generalize to a longer duration Mackworth Clock Task, the same parameter set was used to generate predictions for a 2-hour Mackworth Clock Task aggregated over 30 minute blocks for 2 hours (N. H. Mackworth, 1948). The participants in Mackworth's original study actually had better accuracy at baseline than Giambra and Quilter's participants. Both the IU and UT (0) parameters were adjusted to reflect this better performance, shifting both values down by .4. This shifted all accuracy values up, but did not alter the slope of the decline found in the Giambra & Quilter Model. Notionally, this may reflect initially higher motivation or effort on the part of the participants in the original Mackworth (1948) study to perform well on the task.

**Fig. 9.**

Model fit to Giambra & Quilter (1987) data.

The model was rerun 3000 times using Mackworth's presentation schedule over a 2 hour period. The results of the model runs are presented in Fig. 10. Again, the fit was reasonably good (RMSE=0.01) compared to the human accuracy results.

**Fig. 10.**

Model fit to Mackworth (1948) data.

General Discussion

Computational cognitive models have made great strides in explaining and predicting performance on a variety of tasks (Anderson et al., 2004; Laird, 2012; Sun, 2006). However, these models typically disregard performance changes that may occur as a result of fluctuations in arousal and alertness due to sleep loss, time on task, and other cognitive moderators (i.e. caffeine, drugs, emotions). To make valid predictions about task performance, computational models need to have a mechanism that can capture how performance changes with respect to these factors (Gluck, Gunzelmann, Gratch, Hudlicka, & Ritter, 2006; Gluck & Gunzelmann, 2013). The introduction of mechanisms for fatigue to the ACT-R cognitive architecture is a good starting point for exploring the effects that degradations in central cognition have on overall task performance (Gunzelmann et al., 2009). Without such a mechanism in place, performance will be largely unchanging regardless of the duration of the task.

The current modeling work extends and validates the fatigue mechanisms in the context of vigilance, bridging the gap between sleep loss and vigilance fields of study and providing theoretical unification. The original fatigue mechanism in ACT-R modeled the effects of sleep loss on performance over the course of 88 hours awake on a 10-minute Psychomotor Vigilance Task (Gunzelmann et al., 2009). We ascertained the mechanism's generalizability by conducting a longer duration PVT task in alert participants. Remarkably, we found that performance degraded in a manner similar to that found in participants from a sleep deprivation study after sleep loss of one night (Tucker et al., 2010). To truly test the robustness of the fatigue mechanism, we turned to the most traditional of vigilance tasks, the Mackworth Clock Task. Both accuracy and response time data in this task were captured by

the model, providing support for the conclusion that degradations in performance associated with the vigilance decrement are functionally equivalent to the declines associated with sleep loss.

In terms of the resource theory of vigilance, our work demonstrates how the progressive declines in utility in ACT-R's central production system can lead to performance changes consistent with observed effects in human behavior. These degradations play out at various time scales (hours to days in the case of sleep deprivation and minutes to hours for vigilance tasks). In addition to supporting current theories of the vigilance decrement, the model is also consistent with research showing that the responsiveness of neural populations declines as neurons "tire out" (Krueger et al., 2008).

One objection that may arise is that the parameters used for the PVT and Mackworth models are different. One reason for this is that the task contexts are different, which may impact task difficulty, motivation, and other factors. Additionally even in the context of the Mackworth task, baseline performance of participants in Mackworth's (1948) study was noticeably better than Giambra and Quilter's (1987), although Giambra and Quilter did not comment on this and despite similar methodologies. Further validation of the fatigue mechanism in other vigilance task contexts is needed, as is a better understanding of the interaction of the task context and the environment with alertness and cognitive performance to calibrate the parameter settings.

In conclusion, by validating the fatigue mechanism originally proposed in the context of sleep loss within the context of time on task, we have demonstrated the functional equivalence of the declines in performance observed in both domains, at least for sustained attention tasks.

The degradations to central cognition that have been proposed and validated in this context aid unification of these previously separate domains. In the future we hope results from the domains of sleep deprivation and vigilance will inform each other and that computational cognitive models will make more extensive use of cognitive moderators that influence performance.

Acknowledgments

The views in this paper reflect those of the author, and are not the official views of the United States Air Force. We would like to thank Monica Nguyen for supporting data collection, Jack Harris for experiment setup and MindModeling@Home support, Kevin Gluck, Matthew Walsh, and Rachel Boyd for providing insightful comments on earlier drafts of this manuscript.

References

- Adams, J. A., & Boulter, L. R. (1960). *Monitoring of complex visual displays: I. Effects of response complexity and Intersignal Interval on Vigilant Behavior when Visual Load is Moderate*. Urbana, IL: University of Illinois.
- Adams, J. A., Stenson, H. H., & Humes, J. M. (1961). Monitoring of Complex Visual Displays— II. Effects of Visual Load and Response Complexity on Human Vigilance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 3(4), 213–221.
- 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), 1036–1060.
- Bakan, P. (1955). Discrimination decrement as a function of time in a prolonged vigil. *Journal of Experimental Psychology*, 50(6), 387.
- Benitez, P. L., Kamimori, G. H., Balkin, T. J., Greene, A., & Johnson, M. L. (2009). Modeling fatigue over sleep deprivation, circadian rhythm, and caffeine with a minimal performance inhibitor model. *Methods in Enzymology*, 454(C), 405–421.
- Davies, D., & Parasuraman, R. (1982). *The psychology of vigilance*. London: Academic Press.
- Dinges, D. F., & Powell, J. W. (1985). Microcomputer analyses of performance on a portable simple visual RT task during sustained operations. *Behavior Research Methods, Instruments & Computers*, 17(6), 652–655.
- Doran, S. M., Van Dongen, H. P. A., & Dinges, D. F. (2001). Sustained attention performance during sleep deprivation: Evidence of state instability. *Archives of Italian Biology: Neuroscience*, 139(3), 253–267.
- Fisk, A. D., & Schneider, W. (1981). Control and automatic processing during tasks requiring sustained attention: A new approach to vigilance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 23(6), 737–750.
- Giambra, L. M., & Quilter, R. E. (1987). A two-term exponential functional description of the time course of sustained attention. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 29(6), 635–643.
- Gluck, K. A., & Gunzelmann, G. (2013). Computational Process Modeling and Cognitive Stressors: Background and Prospects for Application in Cognitive Engineering. In J. D. Lee & A. Kirlik (Eds.), *The Oxford Handbook of Cognitive Engineering* (pp. 424–432). New York, NY: Oxford University Press.

- Gluck, K. A., Gunzelmann, G., Gratch, J., Hudlicka, E., & Ritter, F. (2006). Modeling the Impact of Cognitive Moderators on Human Cognition and Performance. In R. Sun & N. Miyake (Eds.), *Twenty-Eighth Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Lawrence Earlbaum Associates.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Society (Vol. 1). Wiley.
- 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.
- Gunzelmann, G., Moore, L. R., Gluck, K. A., Van Dongen, H. P. A., & Dinges, D. F. (2010). Fatigue in sustained attention: Generalizing mechanisms for time awake to time on task. In P. L. Ackerman (Ed.), *Cognitive Fatigue: Multidisciplinary Perspectives on Current Research and Future Applications* (pp. 83–96). Washington, DC: American Psychological Association.
- Harris, J. (2008). MindModeling@Home: a large-scale computational cognitive modeling infrastructure. In *Proceedings of the Sixth Annual Conference on Systems Engineering Research 2008* (pp. 246–252). Los Angeles, CA, USA.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index). In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139–183). Elsevier Inc.
- Hitchcock, E. M., Dember, W. N., Warm, J. S., Moroney, B. W., See, J. E., & Services, L. T. (1999). Effects of cueing and knowledge of results on workload and boredom in sustained attention. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 41(3), 365–372.
- Kaida, K., Akerstedt, T., Kecklund, G., Nilsson, J. P., & Axelsson, J. (2007). Use of subjective and physiological indicators of sleepiness to predict performance during a vigilance task. *Industrial Health*, 45(4), 520–526.
- Kass, S. J., Vodanovich, S. J., Stanny, C. J., & Taylor, T. M. (2001). Watching the clock: Boredom and vigilance performance. *Perceptual and Motor Skills*, 92(3), 969–976.
- Kilpeläinen, A. A., Huttunen, K. H., Lohi, J. J., & Lyytinen, H. (2010). Effect of Caffeine on Vigilance and Cognitive Performance During Extended Wakefulness. *The International Journal of Aviation Psychology*, 20(2), 144–159.
- Krueger, J. M., Rector, D. M., Roy, S., Van Dongen, H. P. A., Belenky, G., & Panksepp, J. (2008). Sleep as a fundamental property of neuronal assemblies. *Nature Reviews Neuroscience*, 9(12), 910–919.
- Laird, J. E. (2012). *The Soar Cognitive Architecture*. Cambridge, MA: MIT Press.

- Lanzetta, T. M., Dember, W. N., Warm, J. S., & Berch, D. B. (1987). Effects of task type and stimulus heterogeneity on the event rate function in sustained attention. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 29(6), 625–633.
- Lichstein, K. L., Riedel, B. W., & Richman, S. L. (2000). The Mackworth Clock Test: A computerized version. *The Journal of Psychology*, 134(2), 153–161.
- Lim, J., & Dinges, D. F. (2008). Sleep deprivation and vigilant attention. *Annals Of The New York Academy Of Sciences*, 1129(1), 305–322.
- Mackworth, J. F. (1964). Performance Decrement in Vigilance, Threshold, and High-Speed Perceptual Motor Tasks. *Canadian journal of psychology*, 18(3), 209–23.
- Mackworth, N. H. (1948). The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology*, 1, 6–21.
- Parasuraman, R. (1979). Memory load and event rate control sensitivity decrements in sustained attention. *Science*, 205(4409), 924–927.
- Ramakrishnan, S., Rajaraman, S., Laxminarayan, S., Wesensten, N. J., Kamimori, G. H., Balkin, T. J., & Reifman, J. (2013). A biomathematical model of the restoring effects of caffeine on cognitive performance during sleep deprivation. *Journal of Theoretical Biology*, 319, 23–33.
- Sun, R. (2006). The CLARION cognitive architecture: Extending cognitive modeling to social simulation. In R. Sun (Ed.), *Cognition and MultiAgent Interaction* (pp. 79–99). Cambridge University Press.
- Teichner, W. H. (1974). The detection of a simple visual signal as a function of time of watch. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 16(4), 339–352.
- Temple, J. G., Warm, J. S., Dember, W. N., Jones, K. S., LaGrange, C. M., & Matthews, G. (2000). The Effects of Signal Salience and Caffeine on Performance, Workload, and Stress in an Abbreviated Vigilance Task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 42(2), 183–194.
- Tucker, A. M., Whitney, P., Belenky, G., Hinson, J. M., & Van Dongen, H. P. A. (2010). Effects of sleep deprivation on dissociated components of executive functioning. *Sleep*, 33(1), 47–57.
- Van Dongen, H. P. A., Belenky, G., & Krueger, J. M. (2011). A Local , Bottom-Up Perspective on Sleep Deprivation and Neuro- behavioral Performance. *Sleep Rochester*, 11(19), 2414–2422.

- Van Dongen, H. P. A., & Dinges, D. F. (2005). Circadian rhythms in sleepiness, alertness, and performance. In M. H. Kryger, T. Roth, & W. C. Dement (Eds.), *Principles and practice of sleep medicine* (pp. 435–443). Philadelphia, PA: Elsevier Saunders.
- Van Dongen, H. P. A., Price, N. J., Mullington, J. M., Szuba, M. P., Kapoor, S. C., & Dinges, D. F. (2001). Caffeine eliminates psychomotor vigilance deficits from sleep inertia. *Sleep*, 24(7), 813–9.
- Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic bulletin & review*, 7(3), 424–65.
- Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance Requires Hard Mental Work and Is Stressful. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(3), 433–441.
- Wiener, E. L. (1977). Stimulus presentation rate in vigilance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 19(3), 301–303.