An Integrative Approach to Understanding and Predicting the Consequences of Fatigue on Cognitive Performance

Glenn Gunzelmann and Kevin A. Gluck
Air Force Research Laboratory

The deleterious consequences of fatigue have motivated decades of research to understand the impact of inadequate sleep on cognitive performance. A key objective is to use insights from that research to develop predictive models that can serve as valid tools for managing work-rest schedules and making Go, No-Go mission decisions. Ultimately, this is about maximizing human performance and minimizing risk. In this paper, we describe a methodology that is moving us in the direction of achieving this goal, involving the integration of mathematical and computational process modeling approaches to understand how fatigue affects human cognitive processes. Mathematical models that capture the dynamics of the human arousal system are integrated with a cognitive architecture that instantiates a unified theory of the mechanisms of human cognition. The integration of these approaches leads to an enhanced ability to quantify the impact of fatigue on performance in particular tasks. We illustrate this by making principled, a priori predictions regarding how human performance in instrument flight with a Predator UAV synthetic task environment may change across 4 days without sleep.

KEYWORDS: Cognitive Architecture, Computational Model, Mathematical Model, Fatigue, Circadian Rhythm, Sleep

A major goal of research on fatigue is to understand how and why cognitive processing degrades in enough detail to prevent potentially catastrophic breakdowns in performance. This is true whether fatigue is a consequence of the interaction of time awake and circadian rhythms (Dinges, 2004; Klerman & St. Hilaire, 2007), or the result of extended time on task (Davies & Parasuraman, 1982; Hancock, Williams, Manning, & Miyake, 1995; Hockey, 1997; van der Linden, Frese, & Meijman, 2003). In this paper, we address degradations in performance associated with extended time awake and circadian rhythms, and thus preferentially use the term fatigue in the context of degradations to alertness stemming from these factors. However, we believe that our approach is readily generalizable to cognitive fatigue, as well as to other moderators of cognitive performance and decision making. As evidence for this, Ritter, Reifers, Klein, & Schoelles (2007) have used a similar methodology to account for the impact of stress on the cognitive system, and Gonzalez and colleagues (Fu, Gonzalez, Healy, Kole, & Bourne, 2006; Gonzalez, Fu, Healy, Kole, & Bourne, 2006) have taken an analogous approach to look explicitly at time-on-task effects.

An extensive history of empirical research has documented changes in performance as individuals are deprived of sleep for long periods of time (e.g., Doran, Van Dongen, & Dinges, 2001; Habeck et al., 2004; Horowitz, Cade, Wolfe, & Czeisler, 2003; Lisper & Kjellberg, 1972). In addition, neuropsychological and neurophysiological research has explored the neural bases of these deficits, adding to our understanding of fatigue as well as helping to identify basic cognitive mechanisms and their neural correlates (e.g., Drummond, Brown, Salamat, & Gillin, 2004; Saper, Scammell, & Lu, 2005; Thomas et al., 2000). Some of the insights from neurophysiological research have been captured by biomathematical models, which represent the dynamics of human alertness as a function of time awake and circadian rhythms (e.g., Hursh et al., 2004; Jewett & Kronauer, 1999; McCauley, Kalachev, Smith, Belenky, Dinges, & Van Dongen, 2009). These models are the centerpiece of the current generation of mathematical tools that can be used to prescribe work-rest schedules in order to maximize overall cognitive performance (e.g., Dean, Fletcher, Hursh, & Klerman, 2007; Hursh, Raslear, Kaye, & Fanzone, 2006).

The biomathematical models currently in use and being developed abstract away from the physiological processes themselves, while capturing the overall influence of various factors on cognitive functioning, or alertness. All of these models are grounded in the so-called two-process account of alertness, which places emphasis on sleep homeostasis and circadian rhythms (Borbély, 1982). However, they differ with regard to the way in which these processes are instantiated in the

models (i.e., what mathematical functions are used to represent the various influences). They also vary in other details. For instance, one model – the Circadian Neurobehavioral Performance and Alertness model, or CNPA – includes algorithms to represent the influence of light exposure on the circadian pacemaker (Jewett & Kronauer, 1999; Kronauer, Forger, & Jewett, 1999). This model also includes functions to represent sleep inertia, which is the general grogginess we experience shortly after awakening.

A formal description of the biomathematical models is beyond the scope of this paper. More information is available in a special issue of Aviation Space and Environmental Medicine (Neri, 2004), including an overview of some general features of the models (Mallis, Mejdal, Nguyen, & Dinges, 2004), a detailed evaluation of the strengths and weaknesses of seven specific models (Van Dongen, 2004), and an assessment of challenges being faced as these models evolve (Dinges, 2004). In his evaluation, Van Dongen utilized a least squares scaling method to align the generic output of the models (e.g., estimates of "cognitive throughput" in CNPA) to particular dependent measures. This requirement exposes an important limitation of these models for making performance predictions. In discussing research challenges, Dinges (2004) describes the use of biomathematical models as "more descriptive curve-fitting, than theoretically driven, hypothesis-generating, data-organizing mathematical approaches" (p. A182), and points to a need for increased attention to the association between alertness and cognitive processing. This is precisely where our research is situated.

The use of biomathematical models in technologies such as the Circadian Performance Simulation Software, or CPSS, which embodies the CNPA model (Jewett & Kronauer, 1999; Kronauer et al., 1999), and the Fatigue Avoidance Scheduling Tool, or FAST, which intantiates a different model (Hursh et al., 2004; 2006), represents an important step toward achieving the goal of applying research on fatigue to reducing the likelihood of fatigue-induced errors that can have costly consequences. Our research is intended to extend the explanatory power of such tools. For instance, biomathematical models currently do not address individual differences in the decrements in cognitive performance resulting from fatigue (Dinges, 2004; Van Dongen, Baynard, Maislin, & Dinges, 2004). This is a complex issue, as individual differences may arise from a variety of sources, only some of which are related to alertness or fatigue. In any task context, individuals will vary as a consequence of variability in knowledge, skill, or experience (e.g., Chase & Simon, 1973; Chi, 1978; Parasuraman, 1976).

Differences may relate to characteristics like age or gender, and often vary as a function of task (Benbow & Stanley, 1980; Salthouse, 1992). Individual differences in susceptibility to fatigue must take into account such preexisting differences in performance. Below we present an example of how individual differences in strategy – a difference in knowledge – can affect the severity of the performance decrements associated with fatigue.

In addition to individual differences, existing tools produce only a measure of overall cognitive functioning, such as "effectiveness" (Hursh et al., 2004) or "cognitive throughput" (Jewett & Kronauer, 1999). Though useful as descriptive models, such constructs do not allow for the generation of *in situ* performance predictions, as noted by Dinges (2004). To relate the output to some human performance measure, like reaction time or error probability, human performance data are required for the dependent measure of interest to drive a scaling process. Such data are often unavailable in applied contexts. Moreover, the scaling process provides no insight regarding the underlying cognitive mechanisms and how they are affected by fatigue, thereby limiting generalizability to other tasks and measures. All research is limited in one way or another, and in pointing out these limitations it is not at all our intention to denigrate or trivialize the important scientific contribution these models make to our understanding of fatigue. Once identified, however, weaknesses and limitations must be addressed. The models are, in fact, a critical component of our theoretical account, as we describe below.

Our research is based on the premise that an accurate understanding of fatigue necessitates an account of how changes in overall alertness impact particular human information processing capabilities. Without understanding these relationships, it is difficult or impossible to anticipate how performance will decline in novel contexts. Such knowledge has important implications for determining optimal work/rest schedules and sleep requirements to maximize performance in a task-sensitive manner. Thus, the research described here focuses on extending the potential of existing biomathematical models of alertness by integrating them with a theory of the human information processing system, implemented as a set of computational mechanisms that instantiates the theory - a *cognitive architecture* (Anderson, 2007; Newell, 1990). This integration is explicitly targeted at grounding estimates of alertness in cognitive processes, which allows for the generation of quantitative performance predictions on particular tasks.

Integrating mathematical models of alertness with computational mechanisms in a cognitive architecture requires the specification of a detailed theory regarding the manner in which cognitive processing changes as alertness declines. That is, changes in alertness that are predicted by the mathematical model must be translated into precise changes to specific mechanisms and parameters in the architecture. Once this is accomplished, quantitative performance predictions are possible, which we use to evaluate the validity of the model through direct comparison to human performance data on the same task. Our approach also establishes a framework for understanding individual differences, allowing for an investigation of how cognitive ability, task knowledge, and susceptibility to the negative effects of decreased alertness interact in particular task contexts to produce observed differences in performance. In what follows, we describe our research approach and examples of how these efforts have led to a more comprehensive and predictive explanation of fatigued performance than has been possible previously.

INTEGRATING METHODOLOGIES

At the core of our research approach is the goal of developing cognitive architectures, which are unifying theories of cognition (Newell, 1990). Cognitive architectures represent, within a single computational system, a theory of a general set of mechanisms that enable information processing abilities in humans. As Newell (1990) put it:

> "A single system (mind) produces all aspects of behavior. It is one mind that minds them all. Even if the mind has parts, modules, components, or whatever, they all mesh together to produce behavior… If a theory covers only one part or component, it flirts with trouble from the start. It goes without saying that there are dissociations, interdependencies, impenetrabilities, and modularities… But they don't remove the necessity of a theory that provides the total picture and explains the role of the parts and why they exist" (pp. 17-18).

We believe that this theoretical orientation is particularly relevant in the case of developing an understanding of fatigue. The effects of fatigue are widespread, with negative consequences observed in human performance across an extensive range of tasks and domains (Durmer & Dinges, 2005; Dinges, Baynard, & Rogers, 2005). With the 24/7 nature of modern society and the potential

for disastrous errors stemming from fatigue (e.g., Caldwell, 2003; Mitler et al., 1988; Pack et al., 1995), there is a pressing need for research that puts the pieces together to explain in detail how and why human cognitive performance changes as alertness varies. To that end, we are developing a detailed computational process theory that allows us to explain the effects of fatigue on the cognitive system and to predict the resulting impact on performance.

The cognitive architecture used in this research is ACT-R, which stands for Adaptive Control of Thought – Rational (Anderson, 2007; Anderson et al., 2004). ACT-R comprises a set of mechanisms representing a theory of the human information processing system. It is implemented in software and incorporates perceptual and motor modules, which allow ACT-R models to interact directly with numerous computer-based tasks by processing information available in the computer display and generating virtual mouse movements and keypresses. Because the perceptual, cognitive, and motor mechanisms are constrained by existing empirical data and quantitative theories regarding the capabilities and limitations of human cognition, the system produces behavior that corresponds closely with the performance of humans in a variety of contexts (for partial reviews see Anderson, 2007; Anderson & Lebiere, 1998; Anderson et al., 2004).

Importantly, while ACT-R represents a *unified* theory of cognition, it does not constitute a *complete* theory of cognition. That is, ACT-R contains representations and processing mechanisms that instantiate a theory of human cognition, but the theory has not been extended to all aspects of human cognitive performance. Of particular relevance here is that ACT-R does not include mechanisms to represent human alertness nor the negative consequences associated with remaining awake for extended periods of time. Our research helps to extend the psychological validity and explanatory breadth of the architecture by leveraging well-validated theories from the existing literature, that quantify the dynamics of alertness as a function of these factors.

An Emphasis on Mechanism

While mathematical models characterize the dynamics of alertness across time, other theories have been proposed to explain the changes in human performance that are observed. One common theory posits that fatigue produces two primary impacts on cognitive performance – a generalized slowdown in cognitive processing and an increase in the frequency and duration of lapses in cognitive performance (e.g., Bratzke, Rolke, Ulrich, &

Peters, 2007; Dinges & Kribbs, 1991). In our research, we have attempted to identify how mechanisms may be instantiated within the human information processing system to capture such effects, and what the performance implications are of those mechanisms. To do this, we have focused first on relatively simple laboratory tasks. This allows us to address specific aspects of cognitive functioning that correspond to particular components of the ACT-R cognitive architecture. As a result, we have been able to translate descriptive accounts of the impact of fatigue into computational mechanisms that lead to quantitative performance predictions in our computational cognitive models. While the mechanisms that have been implemented produce both *slowing* and *lapses* in the model's performance, they do so in unexpected ways. A description of our research in two task contexts follows to provide an introduction to the mechanisms that have been developed.

The Psychomotor Vigilance Test (PVT)
The Psychomotor Vigilance Test, or PVT (Dinges & Powell, 1985), requires participants to attend to a known location on a computer monitor and respond when a stimulus appears (simple reaction time). Stimuli appear at intervals that vary from 2-10 seconds. The task is straightforward to perform, but the duration of a session (10 minutes) leads to performance decrements, particularly when individuals are deprived of sleep. These decrements show a consistent pattern of longer response times and increased probabilities of lapses (significantly delayed responses or non-responses), combined with higher proportions of false starts (responding before the stimulus appears). These changes are quite sensitive to changes in alertness stemming from time awake and circadian rhythms (e.g., Doran et al., 2001; Dorrian, Rogers, & Dinges, 2005; Van Dongen & Dinges, 2005).

The cognitive demands of the PVT are circumscribed enough to allow for a focused investigation of the impact of fatigue on a particular component of ACT-R – the production system. The production system in ACT-R represents a serial bottleneck in central cognition. Productions, in the form of if-then rules, are matched against the current state of the system. One of the productions is selected using an expected utility function (see Eq. 1), and then executed (fired) as long as it exceeds the utility threshold, T_u. The production serves to modify the system state in some way, and the cycle begins again. The state of the system is represented by the contents of buffers within ACT-R, which hold representations of internally maintained information including information encoded from the external environment. For the PVT, buffers hold information about the task context (i.e., doing the PVT), and also the task state (i.e., what is on the screen, where the item is located). The equation for expected utility is:

$$U_i = P_i G - C_i + \varepsilon \qquad \text{Eq. 1}$$

In the equation, P_i is the probability of achieving the goal with production i, C_i is the anticipated cost (in time) associated with using production i to achieve the goal, and G is a global variable that we associate with the concept of alertness from the sleep research literature. Finally, ε is a noise parameter that adds stochasticity to the utility computation.

To account for changes in human performance on the PVT under conditions of fatigue, we decrement the G parameter to reflect decreased levels of alertness. These reductions in G make it possible that on a given cognitive cycle, no productions will have a utility value that exceeds the utility threshold, T_u, resulting in no goal-directed cognitive processing for that cognitive cycle. We refer to these very brief gaps (c. 50 ms) in processing as microlapses. Because there is noise in the computation of expected utility (ε), it is possible for a microlapse to be followed by a cognitive cycle where a cognitive action is successfully executed.

Importantly, this mechanism can produce both slowdowns in cognitive processing and longer cognitive lapses. Intermittent microlapses will manifest as slowing, by causing response times to increase by a relatively small amount. With longer sequences, significantly delayed responses and failures to respond (i.e., lapses and non-responses) occur. In fact, with this mechanism, the ACT-R model exhibits behavior that closely resembles human performance under conditions of fatigue on these measures (see Gunzelmann, Gross, Gluck, & Dinges, 2009). The model provides a new perspective on how underlying information processing may change as a consequence of fatigue and how these changes lead to performance differences that are observed in humans deprived of sleep.

The Walter Reed Serial Addition/Subtraction Task (SAST)
We have also used ACT-R to investigate possible effects of fatigue on accessing and using declarative knowledge (Gunzelmann, Gluck, Kershner, Van Dongen, & Dinges, 2007). The task context for this research has been the Walter Reed Serial Addition/Subtraction Task (SAST; Thorne, Genser, Sing, & Hegge, 1985). In this task, two single-digit numbers and an operator (+ or -) are presented in succession for 200 ms each, with 200 ms intervals between items. The task is to perform the

operation (<N1> <Operator> <N2>), and respond with the ones digit if the answer is non-negative. If the answer is negative, participants are instructed to first add ten, and then to respond with this new result, which will be a positive one-digit number.

As with all ACT-R models, the central production system plays a critical role in the model's performance. In this model, however, ACT-R's declarative memory module is also crucial, since the model represents knowledge about numbers and math facts as chunks in declarative memory. These chunks have activation values, which represent the history (recency and frequency) of use as well as the influence of the current context, all of which impact the speed and probability of retrieval. Our model assumes adult-level experience with simple mathematical calculations (derived from Lebiere, 1999), with the context defined by the problem elements presented on each trial. The equation for activation in ACT-R is:

$$A_i = B_i + \sum_j W_j S_{ji} - D_{ip} + \sigma \qquad \text{Eq. 2}$$

B_i represents base-level activation, which reflects the recency and frequency with which chunk i has been accessed. The summation represents the influence of context, with spreading activation, W, divided equally across the j elements of the goal. Activation spreads to chunk i based upon the strength of association (S_{ji}) between element j in the goal and chunk i. Errors may be generated in this process when the wrong chunk is retrieved from memory (error of commission), which can occur because of the noise in activation values (σ). Items that are similar to each other are more likely to be confused – a mechanism referred to as partial matching (see Anderson & Lebiere, 1998 for a thorough description). This is reflected in the D_{ip} term, which decreases activation as a function of dissimilarity between the chunk i and the request p. In our model, similarity between numbers declines exponentially as the difference between them increases.

When a retrieval request is made by the central production system, the chunk in declarative memory with the highest activation (including partial matching and noise) is selected. As in procedural knowledge, a retrieval request in declarative memory is only successful if the activation of the chunk exceeds a threshold, referred to as the retrieval threshold, or T_r. If not, nothing is retrieved (error of omission). When a chunk is successfully retrieved, its activation influences the speed of retrieval through the following equation:

$$T_i = Fe^{-A_i} \qquad \text{Eq. 3}$$

Here, A_i is the activation of the chunk i, and F is a scaling factor (the default value of 1 is used here). Retrieval times play an important role in the SAST model at various points in the solution process. The model uses stored knowledge of numbers, mathematical operators, and addition and subtraction facts to determine the response on each trial. To encode a number or operator from the screen, the model attends to it by providing requests to ACT-R's vision module, and then uses the visual representation to retrieve the symbolic representation of the number/operator from memory. Once the entire problem has been presented, ACT-R uses the symbolic information to guide the retrieval of a chunk representing an addition or subtraction fact in memory. The result encoded in the math fact that is retrieved is then used to respond, or as the basis for performing the subsequent addition if the result is negative.

In extending our fatigue theory to ACT-R's declarative knowledge, we implemented it in a way that is analogous to the procedural mechanism described above. Specifically, activation values in declarative knowledge are decreased to represent greater fatigue, corresponding to lower utility values for procedural knowledge. This is achieved by decreasing the base-level activation (B_i) through manipulation of a scaling parameter associated with the calculation of this value.

As activation values decrease, retrievals take longer and are less likely to succeed. Longer retrieval times result in longer response times, simulating cognitive slowing. Because the knowledge in the SAST is well-learned, failures to retrieve the necessary information are relatively rare. However, during the encoding phase, longer retrieval times increase the likelihood of errors by making it more likely that the model will fail to encode portions of the stimulus. Because presentation time is short (200 ms), more time taken to retrieve the symbolic representation of one of the numbers can lead to a failure to encode a subsequent item.

Lastly, we note that the procedural microlapse mechanism described for the PVT model was included in the SAST model, reflecting an aspiration to achieve a more comprehensive theory. Of course, microlapses can also delay the production of a response and increase the likelihood of an encoding lapse, but the shorter duration of the SAST (typically less than 3 minutes for a 50-problem session) and the faster pacing of the task (200 ms between problem elements) serve to limit opportunities for procedural activity to decline, minimizing the impact of that mechanism in this model. Thus, it is changes in the accessibility of declarative

knowledge that drives the changes in the model's performance. As people become increasingly fatigued, both response times and errors increase in the SAST. The set of mechanisms described here accurately captures the performance decrements exhibited by human participants across an extended period of sleep deprivation (Gunzelmann et al., 2007).

Integrating Mathematical Modeling

So far, the discussion has focused on the computational mechanisms we have developed to account for the effects of fatigue. However, a critical piece of this research involves linking changes in information processing mechanisms to predictions about overall cognitive functioning derived from biomathematical models of the human arousal system that were mentioned above. Models that can predict the dynamics of human alertness are essential for making *a priori* predictions about performance in particular task contexts given particular histories of wakefulness and sleep.

To integrate the predictions of these models into ACT-R, linear functions were estimated to map predictions of overall alertness to each of the relevant parameters in the architecture. Thus, alertness maps directly to values of *G* in the procedural system and to the scaling parameter for computing base-level activation (B_i) in the declarative system. Though we have not emphasized them in this paper, alertness also influences the threshold parameters in each of these systems, representing effortful compensation in the production system (T_u) and a secondary consequence of fatigue in declarative memory (T_r). Table 1 illustrates this integration. It presents the predicted level of alertness from the CNPA model for 0800 after a full night's sleep, and after 1, 2, and 3 days of total sleep deprivation (TSD). It also presents the slope and intercept for the linear function mapping that value onto the four parameters that vary in the SAST,

including the resulting parameter values for the model at 0800 on each day of the study. These parameter values are the ones used in generating performance predictions in the aircraft instrument flight basic maneuvering task described below. Incorporating the mathematical models of alertness serves to enhance the theoretical constraint on the account, which reduces degrees of freedom in fitting the data and extends the predictive capacity of the model. Further details can be found in Gunzelmann et al. (2007; 2009).

As the slopes presented in Table 1 suggest, changes in activation across days of sleep deprivation are primarily responsible for the performance changes in the model. In Gunzelmann et al. (2007) we included the other parameters to emphasize the goal of developing a comprehensive, general, and integrated theory of the impact of alertness on the cognitive system. Significant research is needed to understand the interactions of the parameters with each other, with internal factors like motivation and interest, and with external factors like task dynamics and time on task.

The mapping functions in Table 1 that link alertness to ACT-R parameters were optimized on the basis of existing empirical data for the SAST. However, the focus of this paper is on how the methodology enables making *in situ*, *a priori* performance predictions in novel tasks. In the next section we provide a use case for this capability by integrating mechanisms for fatigue into an existing model that executes basic maneuvers in a Predator Unmanned Aerial Vehicle (UAV) Synthetic Task Environment (STE). Such a generalization is not possible using biomathematical models of alertness in isolation, because data are not available to support a scaling process to map alertness onto measures like the probability of failure or deviation from optimal performance in this task context.

Table 1. *Integration of biomathematical predictions of alertness with ACT-R.*

			Parameter Values Used in Model			
Linear Function			**Baseline**	**Day 1 TSD**	**Day 2 TSD**	**Day 3 TSD**
Parameter	*Intercept*	*Slope*	*(0.889)*	*(0.521)*	*(0.159)*	*(0.038)*
G	4.21	0.03	4.24	4.23	4.22	4.21
T_u	2.39	-0.08	2.31	2.34	2.37	2.38
Scaler (B_i)	2.60	23.40	117.04 (4.45)	74.00 (3.58)	31.64 (2.65)	17.41 (2.05)
T_r	-0.06	-0.62	-0.61	-0.39	-0.16	-0.09

Notes: *G* is the global parameter in ACT-R's utility equation (Eq. 1) and T_u is the utility threshold. B_i is the base-level activation (Eq. 2) for declarative knowledge and T_r is the retrieval threshold. For each day of total sleep deprivation (TSD), predictions of the Jewett & Kronauer (1999) model for cognitive throughput for 0800 on that day are shown in parentheses under the column label, which are the inputs to the linear functions producing the parameter values. In conjunction with the scaler parameter for B_i, the corresponding B_i value is presented in parentheses..

PERFORMANCE PREDICTION
IN THE CONTEXT OF SIMULATED PREDATOR
BASIC MANEUVERS

To demonstrate the potential of our approach for predicting cognitive performance under conditions of fatigue, this section describes a set of predictions that have been made regarding degradations in human instrument-flight piloting resulting from an extended period of sleep deprivation. The particular task context used here is a Predator unmanned aerial vehicle (UAV) synthetic task environment (STE). This is a complex task simulation requiring control of the UAV from a remote location using a narrow field-of-view video. In the basic maneuvering task, no out-of-cockpit view is provided, necessitating reliance on instrument readings to successfully perform the task. The STE incorporates a high-fidelity model of the aerodynamics of the Predator RQ-1A System 4 UAV (Martin, Lyon, & Schreiber, 1998). Prior research using this STE resulted in three ACT-R model variants that fly a series of "basic maneuvers" requiring carefully controlled, constant-rate changes to the Predator's altitude, airspeed, and/or heading (Gluck, Ball, & Krusmark, 2007; Gluck, Ball, Krusmark, Rodgers, & Purtee, 2003). The heads-up display and task screen are shown in Figure 1. Ideal performance is obtained when control settings (i.e., power, pitch, and bank) are set appropriately to achieve particular flight performance characteristics (i.e., airspeed, altitude, and rate of heading change).

The three model variants represent three different strategies for performing the task, varying in terms of whether they know about appropriate control settings for the maneuvers and how much they attend to the control instruments at the beginning of the maneuvers. The Performance variant (Model P) does not know about the use of control settings in instrument flight and attends exclusively to the Performance instruments throughout the maneuver. The Control and Performance variant (Model CP) does know about control instrument settings and divides its attention between the control and performance instruments as it performs a *crosscheck* of all of them over the course of each maneuver. The Control Focus and Performance variant (Model CFP) also knows about control instrument settings, but at the beginning of each maneuver this model focuses explicitly on getting the control instruments set to desired values. After establishing desired control settings, Model CFP then divides its attention across the control and performance instruments as it engages in a crosscheck of the instruments throughout the remainder of the maneuver.

The implementation of the CFP strategy is based on the process Air Force pilots are trained to use for instrument flight and has been validated through comparison with expert pilot performance (Gluck et al., 2003). The CP and P model variants represent less sophisticated strategies, but are not necessarily intended to represent strategies used by particular individuals for the task (Gluck et al., 2007). More detailed descriptions of the validation of the CFP model, rationale for these model variants, and additional implementation details are available in Gluck et al. (2003, 2007) and in Ball, Gluck, Krusmark, and Rodgers (2003).

We implemented the fatigue mechanisms described earlier directly into these three models. Importantly, the ACT-R models that pilot the simulated UAV are tightly integrated with the STE. The models move visual attention around the screen and generate virtual stick and throttle movements, which are relayed directly to the STE. Thus, the performance of the models can be measured in exactly the same manner as human performance, since the model flies the maneuvers and the STE generates output data, just as it would with a human participant. As an initial illustration of the capacity to make principled predictions regarding the impact of fatigue in a complex naturalistic task such as this, we used the prescribed parameter values taken directly from the SAST for Baseline (0 hours without sleep), 24, 48, and 72 hrs without sleep (see Table 1; Gunzelmann et al., 2007). We use the parameter values from the SAST because the duration of each SAST session in the available empirical data are close to the same as the basic maneuvering tasks, which last 70 or 100 seconds, depending on the particular maneuver. Also, the relatively rapid pace of the SAST is a closer match to the dynamic piloting task than the PVT, which has long time periods with no environmental activity

We report two model performance measures. First is the frequency of "failing" the maneuver, which means the model did not remain within a pre-determined deviation threshold for each of the flight performance dimensions (altitude, heading, and airspeed). These thresholds were selected by the original developers of the STE to be challenging enough to stress the capabilities of even

experienced pilots, but still be reasonable performance targets for that population (Schreiber, Lyon, Martin, & Confer, 2002). The second performance measure is a normalized, aggregate measure of deviation from optimal performance, across those same three performance dimensions. Optimal performance requires precisely maintaining specific rates of change for altitude, heading, and airspeed throughout the duration of the maneuver. It may be theoretically possible to perform optimally in this task, but in practice the quality of a person's or model's performance on each trial is measured as a deviation from that optimal level.
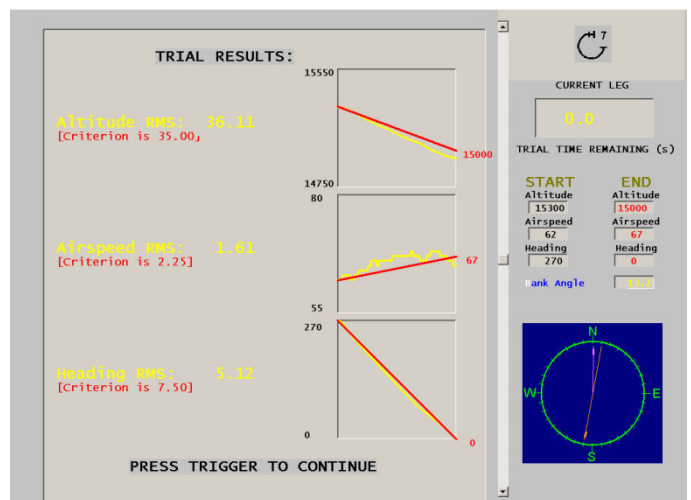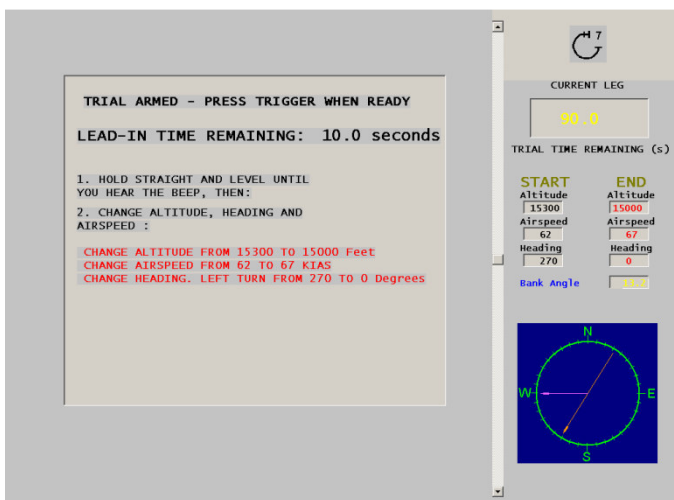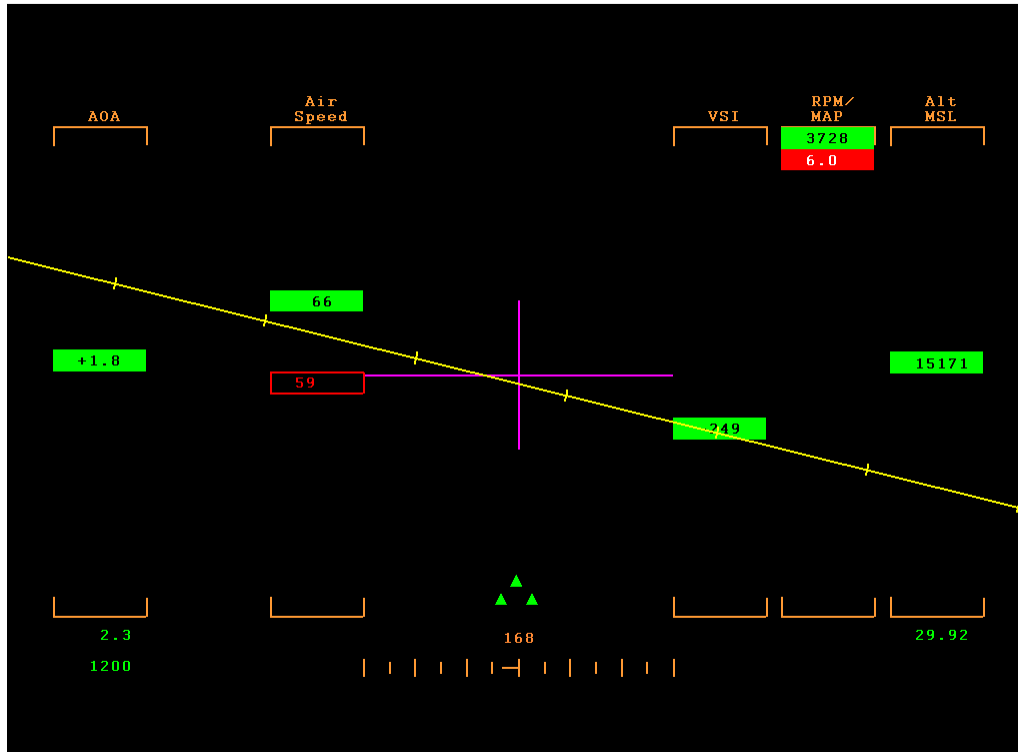
Figure 1. **Display components of the Predator Unmanned Aerial Vehicle (UAV) Synthetic Task Environment. The top panel illustrates the heads-up display (HUD), which replicates the display used on the actual Predator UAV. The bottom-left panel illustrates the task window before the trial starts, indicating the task goals, and providing information about bank angle and time remaining. Finally, the bottom-right panel shows the feedback screen, showing deviation from optimal performance for each of the three measures of interest, (1) altitude, (2) airspeed, and (3) heading. The pass/fail determination is made based upon whether the average deviations on each of these measures remain within a prespecified limit, shown in red on the feedback screen.**

Table 2. *Performance for Each of the Model Variants at Baseline, and after 1, 2, and 3 Nights without Sleep.*

Model	Dependent Measure	Hours Without Sleep			
		0	**24**	**48**	**72**
Model P	*Failures*	42	48	47	49
	Aggregate Z-Score (SD)	3.40 (1.92)	3.85 (1.77)	5.11 (1.69)	6.95 (1.51)
Model CP	*Failures*	42	43	48	50
	Aggregate Z-Score (SD)	3.06 (2.91)	4.30 (3.41)	6.81 (4.95)	10.79 (5.08)
Model CFP	*Failures*	11	13	25	38
	Aggregate Z-Score (SD)	1.22 (1.29)	1.05 (1.37)	1.48 (2.06)	4.09 (3.86)

Notes: P = Performance Model. CP = Control and Performance Model. CFP = Control Focus and Performance Model. Baseline simulates normal, rested performance. Days 1, 2, and 3 simulate performance after that many days of total sleep deprivation. Failures are the number of trials out of 50 in which performance exceeded a pre-defined threshold for deviation from optimal maneuvering. Z-Agg are normalized model performance scores aggregated across altitude, heading, and airspeed. Z-scores were computed using means and standard deviations from the sample of expert-level human pilots previously reported in Ball et al. (2003), and in Gluck et al. (2007).

Each of the three model variants was run at each of the four levels of sleep deprivation on the most difficult maneuver built into the Predator STE, which requires constant-rate changes to altitude, heading, and airspeed simultaneously over a 90-second period (see Figure 1). Due to the stochastic architectural characteristics in ACT-R, each model variant was run through 50 iterations at each level of sleep deprivation to ensure an adequate estimate of its central tendency and variability at the different fatigue levels. Table 2 shows the performance predictions.

At Baseline, the models' rank order performance results are as we would expect given prior results with these models (Gluck et al., 2007). CFP is the best performing model and P is the worst (see Table 2). The fatigue simulations produce a main effect of fatigue level $F(3, 441) = 79.77$, $p < .001$ ($MSE = 10.036$), and also reveal an interaction between model variant and fatigue level $F(6, 441) = 8.24$, $p < .001$ ($MSE = 10.036$), with the largest impact of fatigue on Model CP. In contrast, Model CFP is noteworthy in that performance remains quite good even after two days without sleep, then spikes up dramatically after a third day of sleep deprivation. Even after three days of sleep deprivation, however, predicted performance using the CFP strategy only degrades to approximately the level of performance observed at Baseline for the other two model variants. This reflects the superiority of the CFP strategy for carefully controlled instrument flight maneuvers and suggests that it is more robust in the face of sleep deprivation. In addition, this result illustrates the point from the introduction that individual differences in

baseline knowledge or strategies can have important implications for how decreased alertness impacts performance. We address this, and other points, in the conclusion.

CONCLUSIONS

The models that fly the basic maneuvers in the UAV STE vary in terms of whether they have knowledge about the relationship between control settings and performance characteristics, as well as in the sophistication with which this knowledge is used. The most naïve model, Model P, has no knowledge of the relationship between control settings and performance characteristics of the plane. Model CP does have this knowledge, but does not use knowledge of appropriate control settings to deliberately establish them when the trial begins. Model CFP has knowledge of the correct control settings, and also focuses on establishing appropriate control settings early in the trial, which tends to set the plane on an appropriate course, leading to better performance in general.

It is interesting that our fatigue mechanisms have varying impacts for these different models. As in the model for the SAST, the main impact of decreased alertness in the model is to reduce the activation of declarative knowledge, leading to longer retrieval times and occasional retrieval failures (when activation fails to exceed the threshold). This has the general effect across all three models of slowing down the crosscheck process, leading to slower compensations and adjustments to keep the plane's flight characteristics in line. For Model CFP,

this effect is minimized due to the initial focus on establishing appropriate control settings. Since these initial adjustments tend to set the UAV on a path that is close to the desired one, the adjustments made during the remainder of the trial are typically minor changes that serve to optimize the UAV's performance given the goals of the maneuver. Of course, with greater levels of fatigue, even the initial process of establishing control settings is delayed significantly, as are the subsequent minor adjustments, leading to larger decrements in performance if not catastrophic failure.

For the other two models, performance is more dependent on attending to the instruments and making continual adjustments to bring the UAV's performance in line with the desired characteristics. This dependence is more pronounced in Model CP, which is a result of the knowledge it has regarding control instrument settings. Accessing this knowledge is costly in terms of time, and becomes even more costly as alertness declines. Because Model P is less dependent on declarative knowledge for making adjustments to control the UAV, its performance is impaired less as a result of the reduced activation of declarative knowledge stemming from fatigue than is Model CP. However, it is also the case that Model P instantiates the least effective strategy initially, and so its performance is never particularly good.

Finally, note that the model predictions reported here are real predictions. Unfortunately, there are no empirical data available regarding the effects of sleep deprivation on this basic maneuvering task. Thus, we are unable to validate the predictions the models make. However, what is most important about this research is that it demonstrates real progress in the direction of precise, *a priori* quantitative performance predictions. Our approach to accomplishing this combines mathematical and computational architectural modeling approaches. The use of a cognitive architecture to represent the mechanisms of human information processing maps overall alertness onto cognitive functions that translate into task performance. Biomathematical models of alertness are fundamentally limited in this regard, since they do not incorporate a theory of human cognition. They cannot be used to make performance predictions in a novel context without being scaled to existing data. Similarly, cognitive architectures, including ACT-R but also all of the others, are currently limited in the sense that they do not incorporate a theory of how alertness

varies with fatigue level or how those variations influence cognitive processing. By bringing these modeling approaches together, these limitations can be overcome, and quantitative predictions of task performance under fatigued conditions are possible.

REFERENCES

Anderson, J. (2007). *How Can the Mind Exist in a Physical Universe?* Oxford University 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*, 1036-1060.

Anderson, J. R., & Lebiere, C. L. (1998). *The atomic components of thought*. Hillsdale, NJ: Lawrence Erlbaum.

Ball, J.T., Gluck, K.A., Krusmark, M.A, & Rodgers, S.M. (2003). Comparing three variants of a computational process model of basic aircraft maneuvering. In *Proceedings of the 12th Conference on Behavior Representation in Modeling and Simulation* (pp. 87-89). Orlando, FL: Institute for Simulation & Training.

Benbow, C.P., & Stanley, J. C. (1980). Sex differences in mathematical ability: Fact or artifact? *Science, 210*, 1262-1264.

Borbély, A.A. (1982). A two process model of sleep regulation. *Human Neurobiology, 1*, 195–204.

Bratzke, D., Rolke, B., Ulrich, R., & Peters, M. (2007). Central slowing during the night. *Psychological Science, 18(5)*, 456-461.

Caldwell, J.A. (2003, Fall). Wake up to the importance of sleep for air safety! *Flightline*, 30-33.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*, 55-81.

Chi, M. T. H. (1978). Knowledge structures and memory development. In R. S. Siegler (Ed.), *Children's Thinking: What Develops?* (pp. 73-96). Hillsdale, NJ: Lawrence Erlbaum Associates.

Davies, D.R., & Parasuraman, R. (1982). *The psychology of vigilance*. New York, NY: Academic Press.

Dean, D.A. III, Fletcher, A., Hursh, S.R., & Klerman, E.B. (2007). Developing mathematical models of neurobehavioral performance for the "real world." Jo*urnal of Biological Rhythms, 22(3)*, 246-258.

Dinges, D.F. (2004). Critical research issues in development of biomathematical models of fatigue and performance. *Aviation, Space, and Environmental Medicine, 75*(3), A181-A191.

Dinges, D.F., Baynard, M., & Rogers, N.L. (2005). Chronic sleep deprivation. In M.H. Kryger, T. Roth, and W.C. Dement (Eds.), *Principles and Practice of Sleep Medicine (4th edition)* (pp. 67-76). Philadelphia, PA: W. B. Saunders.

Dinges, D.F., & Kribbs, N.B. (1991). Performing while sleepy: effects of experimentally-induced sleepiness. In Monk, T.H. (Ed.) *Sleep, sleepiness and performance* (pp. 97-128). Chichester: John Wiley & Sons.

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., & Dinges, D.F (2001). Sustained attention performance during sleep deprivation: evidence of state instability. *Archives of Italian Biology: Neuroscience 139(3)*, 253-267.

Dorrian, J., Rogers, N.L., & Dinges, D.F. (2005). Psychomotor vigilance performance: Neurocognitive assay sensitive to sleep loss. In Kushida, C.A. (Ed.), *Sleep Deprivation: Clinical Issues, Pharmacology and Sleep Loss Effects*. Marcel Dekker, Inc., New York, NY, pp 39-70.

Drummond, S.P., Brown, G.G., Salamat, J. S., & Gillin, J. C. (2004). Increasing task difficulty facilitates the cerebral compensatory response to total sleep deprivation. *Sleep, 27*, 445– 451.

Durmer, J. S., & Dinges, D. F. (2005). Neurocognitive consequences of sleep deprivation. *Seminars in Neurology, 25(1)*, 117-129.

Fu., W., Gonzalez, C., Healy, A.F., Kole, J.A., & Bourne, L.E., Jr. (2006). Building predictive models of skill acquisition in a data entry task. *Human Factors and Ergonomics Society Annual Meeting (HFES 50th Annual Meeting)* (pp. 1122-1126). Santa Monica, CA: Human Factors and Ergonomics Society.

Gluck, K.A., Ball, J.T., & Krusmark, M.A. (2007). Cognitive control in a computational model of the Predator pilot. In W. D. Gray (Ed.) *Integrated Models of Cognitive Systems* (pp. 13-28). New York, NY: Oxford University Press.

Gluck, K.A., Ball, J. T., Krusmark, M.A., Rodgers, S.M., & Purtee, M. D. (2003). A computational process model of basic aircraft maneuvering. In F. Detje, D. Doerner, & H. Schaub (Eds.), In *Proceedings of the Fifth International Conference on Cognitive Modeling* (pp. 117-122). Bamberg, Germany: Universitats-Verlag Bamberg.

Gonzalez, C., Fu, W., Healy, A.F., Kole, J.A., & Bourne, L.E., Jr. (2006). ACT-R models of training data entry skills. In *Proceedings of the Fifteenth Conference on Behavior Representation in Modeling and Simulation* (pp. 101-109). Orlando, FL: Simulation Interoperability Standards Organization.

Gunzelmann, G., Gluck, K.A., Kershner, J., Van Dongen, H.P.A., & Dinges, D.F. (2007). Understanding decrements in knowledge access resulting from increased fatigue. In D.S. McNamara and J.G. Trafton (Eds.), *Proceedings of the Twenty-Ninth Annual Meeting of the Cognitive Science Society* (pp. 329-334). Austin, TX: Cognitive Science Society.

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.

Habeck, C., Rakitin, B.C., Moeller, J., Scarmeas, N., Zarahn, E., Brown, T., & Stern, Y. (2004). An event-related fMRI study of the neurobehavioral impact of sleep deprivation on performance of a delayed-match-to-sample task. *Cognitive Brain Research, 18*, 306-321.

Hancock, P.A., Williams, G., Manning, C.M., & Miyake, S. (1995). Influence of task demand characteristics on workload and performance. *International Journal of Aviation Psychology, 5(1)*, 63-86.

Hockey, G.R.J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology, 45*, 73-93.

Horowitz, T. S., Cade, B. E., Wolfe, J. M., & Czeisler, C. A. (2003). Searching night and day: A dissociation of effects of circadian phase and time awake on visual selective attention and vigilance. *Psychological Science, 14(6)*, 549-557.

Hursh, S.R., Raslear, T.G., Kaye, A.S., & Fanzone, J.F. Jr. (2006). Validation and calibration of a fatigue assessment tool for railroad work schedules, summary report (United States Department of Transportation Report DOT/FRA/ORD-06/21). Washington, DC: Office of Research and Development.

Hursh, S.R., Redmond, D.P., Johnson, M.L., Thorne, D. R., Belenky, G., Balkin, T. J., Storm, W.F., Miller, J. C., & Eddy, D. R. (2004). Fatigue models for applied research in warfighting. *Aviation, Space, and Environmental Medicine, 75(3)*, A44-60.

Jewett, M.E., & Kronauer, R.E. (1999). Interactive mathematical models of subjective alertness and alertness in humans. *Journal of Biologicla Rhythms, 14*, 588-597.

Klerman, E. B., & St. Hilaire, M. (2007). On Mathematical Modeling of Circadian Rhythms, Performance, and Alertness. *Journal of Biological Rhythms, 22(2),* 91-102.

Kronauer, R.E., Forger, D.B., & Jewett, M.E. (1999). Quantifying human circadian pacemaker response to brief, extended, and repeated light stimuli over the photopic range. *Journal of Biological Rhythms, 14*, 500-515.

Lebiere, C. (1999). The dynamics of cognition: An ACT-R model of cognitive arithmetic. *Kognitionswissenschaft., 8(1)*, 5-19.

Lisper, H., & Kjellberg, A. (1972). Effects of 24-hour sleep deprivation on rate of decrement in a 10-minute auditory reaction time task. *Journal of Experimental Psychology, 96*, 287-290.

Mallis, M., Mejdal, S., Nguyen, T., & Dinges, D. (2004). Summary of the key features of seven biomathematical models of human fatigue and performance. *Aviation, Space, and Environmental Medicine, 75(3)*, A4-A14.

Martin, E., Lyon, D. R., & Schreiber, B. T. (1998). Designing synthetic tasks for human factors research: An application to uninhabited air vehicles. In *Proceedings of the Human Factors and Ergonomics Society*.

McCauley, P., Kalachev, L.V., Smith, A.D., Belenky, G., Dinges, D.F., & Van Dongen, H. P. A. (2009). A new mathematical model for the homeostatic effects of sleep loss on neurobehavioral performance. *Journal of Theoretical Biology, 256,* 227-239.

Mitler, M.M., Carskadon, M.A., Czeisler, C.A., Dement, W.C., Dinges, D.F., & Graeber, R.C. (1988). Catastrophes, sleep, and public policy: consensus report. *Sleep 11(1)*, 100-109.

Neri, D.F. (Ed.). (2004). Fatigue and performance modeling workshop, June 13-14, 2002 [Special Issue]. *Aviation, Space, and Environmental Medicine, 75(3, Supplement)*.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Pack, A.I., Pack, A.M., Rodgman, E., Cucchiara, A., Dinges, D.F., & Schwab, C.W. (1995). Characteristics of crashes attributed to the driver having fallen asleep. *Accident Analysis & Prevention, 27(6)*, 769-775.

Parasuraman, R. (1976). Consistency of individual differences in human vigilance performance: An abilities classification analysis. *Journal of Applied Psychology, 61(4)*, 486-492.

Ritter, F.E., Reifers, A.L., Klein, L.C., & Schoelles, M. . (2007). Lessons from defining theories of stress for cognitive architectures. In W.D. Gray (Ed.), *Integrated Models of Cognitive Systems* (pp. 254-262). New York, NY: Oxford University Press.

Salthouse, T.A. (1992). *Theoretical Perspectives on Cognitive Aging*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Saper, C.B., Scammell, T.E., & Lu, J. (2005). Hypothalamic regulation of sleep and circadian rhythms. *Nature. 437(7063)*, 1257-1263.

Schreiber, B. T., Lyon, D. R., Martin, E. L., & Confer, H. A. (2002). Impact of prior flight experience on learning Predator UAV operator skills (AFRL-HE-AZ-TR-2002-0026). Mesa, AZ: Air Force Research Laboratory, Warfighter Training Research Division.

Thomas, M., Sing, H., Belenky, G., Holcomb, H., Mayberg, H., Dannals, R., Wagner, H. Jr., Thorne, D., Popp, K., Rowland, L., Welsh, A., Balwinski, S., & Redmond, D. (2000). Neural basis of alertness and cognitive performance impairments during sleepiness. I. Effects of 24 h of sleep deprivation on waking human regional brain activity. *Journal of Sleep Research, 9*, 335-352.

Thorne, D.R., Genser, S.G., Sing, G.C., & Hegge, F.W. (1985). The Walter Reed performance assessment battery. *Neurobehaviora. Toxicology and Teratology, 7*, 415-418.

Van der Linden, D., Frese, M., & Meijman, T.F. (2003). Mental fatigue and the control of cognitive processes: Effects on perseveration and planning. *Acta Psychologica, 113*, 45-65.

Van Dongen, H.P.A. (2004). Comparison of mathematical model predictions to experimental data of fatigue and performance. *Aviation, Space, and Environmental Medicine, 75(3)*, A15-A36.

Van Dongen, H.P.A., Baynard, M.D., Maislin, G., & Dinges, D. F. (2004). Systematic interindividual differences in neurobehavioral impairment from sleep loss: Evidence of trait-like differential vulnerability. *Sleep, 27*, 423-433.

Van Dongen, H.P.A., & Dinges, D.F. (2005). Sleep, circadian rhythms, and psychomotor vigilance. *Clinical Sports Medicine, 24*, 237-249.

AUTHOR NOTES