USAARL Report No. 2011-02

The Effect of Sleep Deprivation on the Detection of Correlational and Causal Relationships, and Performance on an Engagement Skills Trainer Task in Soldiers

By Amanda Kelley
Jeremy Athy
Catherine Webb
Melody King

United States Army Aeromedical Research Laboratory
Warfighter Performance and Health Division

November 2010

Approved for public release, distribution unlimited.

Notice

Qualified requesters

Qualified requesters may obtain copies from the Defense Technical Information Center (DTIC), 8725 John J. Kingman Road, Suite 0944, Fort Belvoir, Virginia 22060-6218. Orders will be expedited if placed through the librarian or other person designated to request documents from DTIC.

Change of address

Organizations receiving reports from the U.S. Army Aeromedical Research Laboratory on automatic mailing lists should confirm correct address when corresponding about Laboratory reports.

Disposition

Destroy this document when it is no longer needed. Do not return it to the originator.

Disclaimer

The views, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other official documentation. Citation of trade names in this report does not constitute an official Department of the Army endorsement or approval of the use of such commercial items.

Human Use

Human subjects participated in this study after giving their free and informed voluntary consent. Investigators adhered to Army Regulation 70-25 and USAMRMC Regulation 70-25 on use of volunteers in research.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*
29-11-2010

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE
The Effect of Sleep Deprivation on Detection of Correlational and Causal Relationships, and Performance on an Engagement Skills Trainer Task in Soldiers

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Amanda Kelley
Jeremy Athy
Catherine Webb
Melody King

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Aeromedical Research Laboratory
P.O. Box 620577
Fort Rucker, AL 36362

8. PERFORMING ORGANIZATION REPORT NUMBER
USAARL 2011-02

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
504 Scott Street
Fort Detrick, MD 21702

10. SPONSOR/MONITOR'S ACRONYM(S)
USAMRMC

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

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Correlation detection, or rather the ability to determine whether two things are related, is an essential cognitive ability that underlies a number of other processes such as causal judgment. Under conditions of stress cognitive resources become taxed and the probability of a judgment error increases. To mititgate the risk of an error, people tend to use cognitive short-cuts. It was predicted that performance would reflect a “short-cut” technique and accuracy would be compromised during periods of sleep deprivation. Performance by 20 US Army Soldiers was evaluated on classic laboratory measures of correlation and causal judgment, measures sensitive to fatigue, and a correlation detection task employing the weapons simulator at baseline, after 24 and 36 hours of sleep deprivation, and after 8 hours of recovery sleep. The results of this study suggest that people use “cognitive short-cuts” under as few as 24 hours sleep deprivation thus decrementing performance. By understanding these cognitive short-cuts, cues presented in information displays can be manipulated to exploit these techniques and ultimately decrease the likelihood of an error.

15. SUBJECT TERMS
sleep deprivation, causal judgment, decision making

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Loraine St. Onge, PhD
UNCLAS	UNCLAS	UNCLAS	SAR	41	19b. TELEPHONE NUMBER *(Include area code)* 334-255-6906

Reset

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Acknowledgements

The authors would like to express their sincere gratitude to the following people for their contributions to this project.

- Mrs. Elizabeth Stokes for help with administrative matters
- Dr. Loraine St. Onge for her editorial assistance
- Ms. Melinda Vasbinder, Ms. Lana Milam, SGT Robert MacNeill, Mrs. Stephanie Moon, SGT Pedro Cruz, SGT Barbara Campbell, PFC Shannon Pippinger, SPC Navdeep Saini, SSG David Lopez, Mrs. Edna Rath, and Mr. Bradley Erickson for their assistance with the data collection
- CW3 Jay Bachman for his assistance in subject recruitment
- Dr. John Crowley and LTC Steven Gaydos for serving as study physicians
- Mrs. Jill Emerson for her assistance with human subjects protection matters
- Dr. Angus Rupert for serving as a medical monitor
- Mr. Andrew Lantz for generating figures

Table of contents

Page

Introduction 1

Background 2

Research objectives and hypotheses 4

Methods 5

General 5

Participants 5

Procedure 6

Task Battery 7

Engagement Skills Trainer 2000 7

Intervention task (causal judgment) 9

Observation task (correlation detection) 9

Sleep assessment 10

Subjective assessments 10

Psychomotor Vigilance Task 10

Results 11

Sleep assessment 11

Subjective assessments 12

Mood and alertness visual analogue scale 12

Profile of Mood States 14

Psychomotor Vigilance Task data analysis 16

Reaction time 16

Table of contents (continued)

Page

Lapses 17

Engagement Skills Trainer 2000 results 18

Correlation detection and judgment accuracy 19

Marksmanship accuracy 20

Intervention task (causal judgment) 23

Observation task (correlation detection) 25

Discussion 27

Deficits to positively and negatively correlated samples 27

Deficits in observation task performance 28

Positive bias 28

Deficits in intervention task performance 29

Tasks and abilities sensitive to sleep deprivation 29

Differences in sleep deprivation effects at 24 hours awake and 36 hours awake 29

Limitations and future studies 30

Conclusions 30

References 31

List of figures

1. A contingency table 4

2. Test schedule 7

3. EST 2000 set-up 8

4. Targets in EST 2000 task 9

List of figures (continued)

Page

5. Hand-held PVT device 11

6. Results of the Stanford Sleepiness Scale 12

7. Results of Mood and Altertness Visual Analogue Scale 13

8. Results of POMS 15

9. Mean PVT reaction time data by session 17

10. Mean PVT lapse data by session 18

11. EST 2000 task contingency table 19

12. Mean accuracy in EST 2000 friend/foe detection judgment task 20

13. Mean horizontal displacement by session 21

14. Mean reaction time by session 21

15. Mean vertical displacement by session 22

16. Mean proportion of "hits" by session 22

17. Intervention task relationship rating results 24

18. Intervention task confidence results 24

19. Observation task relationship rating results 26

20. Observation task confidence results 27

List of tables

1. Results of paired comparison *t*-tests for mean VAS adjective ratings at baseline and at 24 hours awake 13

2. Results of paired comparison *t*-tests for mean VAS adjective ratings at baseline and at 36 hours awake 13

List of tables (continued)

Page

3. Results of paired comparison *t*-tests for mean VAS adjective ratings at recovery and at 24 hours awake 14

4. Results of paired comparison *t*-tests for mean VAS adjective ratings at recovery and at 36 hours awake 14

5. Results of paired comparison *t*-tests for mean POMS sub-scale scores at baseline and at 24 hours awake 15

6. Results of paired comparison *t*-tests for mean POMS sub-scale scores at baseline and at 36 hours awake 15

7. Results of paired comparison *t*-tests for mean POMS sub-scale scores at baseline and at recovery 16

8. Results of paired comparison *t*-tests for mean POMS sub-scale scores at recovery and at 24 hours awake 16

9. Results of paired comparison *t*-tests for mean POMS sub-scale scores at recovery and at 36 hours awake 16

10. Significant results of paired samples *t*-tests for marksmanship data 23

11. Results of paired comparison *t*-tests for observation task 26

Introduction

The demands of modern warfare are rapidly changing and require Soldiers to think more flexibly and synthesize more information than ever before. The armed forces of the United States now possess unparalleled military capabilities in firepower, weapons targeting, and guidance systems, speed of weapons and troop delivery, maneuverability, electronic telecommunications, and nighttime operations technology. These advances mean that military operations proceed more rapidly and maintain a higher mission tempo than ever before (Richard & Huffman, 2002). With these advances in technology, lethality, and speed of delivery there is an ever-increasing need for commanders and front-line Soldiers to assimilate more information from diverse sources in less time and use that information to make critical battlefield decisions. Under stressful conditions, Soldiers and their leaders must make instantaneous decisions, often with incomplete data or insufficient time to process all available information. These split-second judgments often have life-and-death consequences and, if the wrong decision is made, destruction of friendly forces or loss of innocent civilian life may result. Recent media accounts from Operation Enduring Freedom and Operation Iraqi Freedom have repeatedly highlighted the grave loss of life and equipment that occurs when judgment fails at a critical moment and lethal weapons systems are engaged in error. Indeed, it has been suggested that human performance may be the most important limiting factor in the effectiveness of advanced weapons systems (Ramsey & McGlohn, 1997).

Recent neuropsychological studies have shown that the stresses associated with participation in combat may lead to changes in normal cognitive functioning (Vasterling, Duke, Tomlin, Lowery, & Kaplan, 2004; Vasterling, Proctor, Amoroso, Kane, Gackstetter et al., 2006; Vasterling, Proctor, Amoroso, Kane, Heeren et al., 2006). However, many of these studies have examined simple cognitive processes such as memory and attention. Researchers are beginning to evaluate higher order cognitive processes, such as judgment, decision making, planning, and risk-taking propensity under operational stress (e.g., Killgore & McBride, 2006).

One type of judgment that has not been examined under conditions of operational stress is the ability to judge causal and correlational relationships (how things vary together and relate to each other) in the environment. Accuracy in detection of causal and correlational relationships in the natural environment is essential to other cognitive functions such as learning, categorization, and decision making (see Shanks, 2004). More specifically, if these abilities are compromised in a combat environment, then Soldiers are more likely to make errors in predicting future events and taking precautionary actions. By further understanding the deficits to these abilities caused by combat stressors, it is possible that counteractions and techniques can be identified to reduce the probability of errors and mistakes. This study examined decisions made after 24 and 36 hours of sleep deprivation which required the decision maker to integrate prior beliefs and experiences with current situational information (e.g., causal reasoning and causal judgment scenarios). In other words, the present study was designed to evaluate whether and, if so, to what extent, performance on tasks related to judgment of correlational and causal relationships is influenced or compromised by conditions of sleep deprivation.

Background

The cognitive performance of a Soldier in combat operations is essential. There are a number of stressors, however, that compromise the cognitive capabilities of Soldiers such as fatigue and nutritional deprivation. It has been documented that sleep deprivation can lead to deficits in cognitive abilities specifically in shift-workers (Akerstedt, 2006; Himashree, Banerjee, & Selvamurthy, 2002; Rogers, Holmes, & Spencer, 2001). However, comparatively few studies have evaluated whether and to what extent sleep deprivation influences more complex behavior and cognition. Specifically, many sleep deprivation studies have relied on simple, basic, laboratory tasks such as reaction time tests and simple memory tasks that are sensitive to adverse effects of sleep deprivation given the monotony of the tasks. It is difficult to use these results to draw inferences about the impact of sleep deprivation in real-world settings and more complex tasks. In recent years, more studies have begun to emerge evaluating higher-order processes under conditions of sleep deprivation and fatigue; however, many of these studies are limited to performance in civilian medical professional shift-workers. Many sleep deprivation studies in military research have focused on evaluating performance of a fatigue management technique (e.g., pharmaceuticals, napping) on performance of a specific task (e.g., simulators, in-flight) in addition to sleep sensitive measures (e.g., psychomotor vigilance task).

Correlation detection, or rather the ability to accurately detect an association between two variables (i.e., determine whether two things are related to each other), is an essential cognitive ability that underlies a number of other processes such as categorization, learning, and causal judgment (Shanks, 2004). This ability allows us to explain the past, control the present, and predict the future (Crocker, 1981). People tend to use judgment heuristics which are "short-cuts" or judgment strategies that typically lead to a good response requiring only minimal effort expended. Heuristics can be very useful especially under conditions of stress or when cognitive resources are taxed but they can also lead to severe errors in judgment and biases (Tversky & Kahneman, 1982). One bias resulting from heuristics in correlation detection and causal judgment is illusory correlation which is defined as the perception of a relationship where in fact one does not exist (e.g., Chapman, 1967; Hamilton & Gifford, 1976). This bias can lead to false beliefs such as superstitions, inaccurate predictions about future events, and, consequently, inappropriate courses of action. Evidence of other errors in correlation detection has been demonstrated under conditions of cognitive overload such that people overestimate the importance of prior experiencs and tend to ignore new information in forming a judgment about the current situation or problem (i.e., related to causal and correlational relationships; e.g., Shanks, 2004). Typically, people use heuristics when resources are taxed such to increase the probability of an accurate response or judgment which can prove advantageous more often than not. However, the clear disadvantage to these short-cuts is that resulting errors can range from minor to severe. In isolation, such an error may not result in an accident as most accidents follow a series of mistakes/errors/malfunctions, however, it may be a contributing factor in the cause of an accident.

As mentioned above, two sources of information used in correlation detection and causal judgment are prior beliefs and expectations (Alloy & Tabachnik, 1984). In some cases, people are very good at integrating their prior beliefs with current or new information in forming a judgment. Under some conditions, however, people show dependence on their prior beliefs and

inadequate weight is given to new information (Shanks, 2004). It has been suggested that overweighting of prior beliefs occurs under situations of cognitive overload as a means of simplifying the task (Arkes & Harkness, 1983). While this may seem to be rational, to re-use information that has proven successful in the past, the error comes from the lack of attention given to current and present information. For example, categorization can prove to be very useful tool for optimizing the accuracy of a judgment while conserving cognitive resources. However, stereotyping, a form of categorization, can prove to have severe negative social consequences. In fact, Friedland, Keinan, and Tytiun (1999) showed that Israeli Air Force cadets who were under high psychological stress demonstrated a greater tendency to provide stereotypic judgments than those were not. The consequences of this type of judgment behavior in military contexts or combat environments are potentially severe.

In experimental studies of correlation detection, a number of different tasks have been used. In one such task, participants are given a cover story which presents a hypothetical situation in which the relationship between two variables must be determined. For example, a commonly employed cover story describes a hypothetical situation in which the participant must imagine that he/she is an agricultural scientist and is investigating the effectiveness of a newly determined chemical compound as a fertilizer. In order to accomplish this task, participants are next presented with a set of data points (or a sample of data). Each data point indicates the occurrence/non-occurrence of both the causal candidate (e.g., chemical applied/not applied) and the effect variable (e.g., plant grew/plant did not grow). Specifically, in this example, a data point in the sample indicates whether a plant received the chemical compound and whether it had grown a significant amount. In the laboratory task, this set of data points is presented either sequentially or in a list format. It is common to summarize this data in a contingency table (figure 1) composed of four cells; Cell A is the frequency of observations where the causal candidate and effect variables are both present (e.g., plant received chemical compound and grew), Cell B is the frequency of observations where the causal candidate is present and the effect is absent (e.g., plant received chemical compound and did not grow), Cell C is the frequency of observations where the causal candidate is absent and the effect is present (e.g., plant did not receive chemical compound but did grow), and Cell D is the frequency of observations where both variables are absent (e.g., plant did not receive chemical compound and did not grow).

		EFFECT	
		Present	Absent
CAUSE	Present	A	B
	Absent	C	D

Figure 1. A contingency table. The cell entries are used to calculate the generally accepted measure of contingency between binary variables; ΔP. The cell entries A through D denote the number of observations in each cell of the contingency table.

Another commonly used task is the train/test correlation detection task which has been shown to be a reliable measure (e.g., Kareev, 1995) of correlation detection and was adapted for use with the weapons simulator (Engagement Skills Trainer [EST] 2000) in the present study. In this task, participants are told that they are to determine the relationship between two events in a test phase and then use one event to predict the occurrence/non-occurrence of another event. For example, a basic laboratory version of this task employs envelopes and coins of different colors. In the first stage of the task, the training stage, participants open a set number of envelopes (e.g., 20 envelopes), half of which are red and half of which are green. The envelopes contain either a gold coin or a silver coin. After the training stage is complete, participants then complete a test stage in which they are given another set of envelopes but this time they must predict the type of coin in the envelope based on its color. Each time that their prediction is correct, they receive a set amount of money (thus motivating them to increase their accuracy). In the present study, this task was adapted to the EST 2000 employing a friend/foe detection wherein participants predicted which target was foe. Performance on this task was correlated with performance on laboratory tasks; one measuring causal judgment, and one measuring correlation detection. Task accuracy (for each individual task) from each test session (baseline, after 24 hours sleep deprivation, after 36 hours sleep deprivation, and after 8 hours recovery sleep) was compared to determine if there were any deficits in this cognitive process associated with sleep deprivation.

Research objective and hypotheses

The main objective of this study was to assess the effects of sleep deprivation and recovery sleep on cognitive processes specific to correlation detection and causal judgment. Performance was evaluated on three tasks (EST 2000 adapted correlation detection task, laboratory correlation

detection task, and a causal judgment task) at baseline, after 24 and 36 hours of sleep deprivation, and after 8 hours of recovery sleep. It was predicted that when participants were sleep deprived they would show decreased accuracy on the tasks compared to their performance at baseline and after recovery sleep. A secondary hypothesis was that participants who were sleep deprived would be more susceptible to common errors and biases such as illusory correlation and would show strong dependence on prior beliefs and expectations. Finally, it was predicted that accuracy would increase after recovery sleep but not to the level of performance exhibited at rested baseline.

Methods

General

The protocol was reviewed and approved by the U.S. Army Medical Research and Materiel Command Human Subjects Research Review Board (USAMRMC HSRRB) prior to implementation. To test the above stated hypotheses, the study employed a within-subjects experimental design. The independent variable was session with four levels (baseline, at 24 hours awake, at 36 hours awake, after 8 hours recovery sleep).

Participants

Participants were 20 active-duty U.S. Army Soldiers. The mean age was 27.7 years and the mean education level was 13.8 years (e.g., 12 years = high school diploma). The mean number of years of military service was 7.35. Of the 20 participants, 18 were male, and 12 had combat experience. Ten participants were Caucasian, four were African-American, five were Hispanic-American, and one was American Indian. Three participants reported that English was not their first language. Volunteers on leave status received $300 for participation. Volunteers who held a graduate level degree (in any field) or a bachelor's degree in statistics were ineligible given that previous research has shown that advanced knowledge in statistics may skew performance on the tasks employed in this study (Doherty, Anderson, Kelley, & Albert, 2006). Volunteers who used nicotine regularly within the past year, whose daily caffeine intake exceeded 600 milligrams, whose alcohol consumption exceeded six beers/week or four mixed drinks/week, or whom the physician investigator deemed exclusionary based on these criteria were excluded from participation. Also, volunteers were excluded if they had a history or currently active condition of substance abuse or sleep disorders. One volunteer was excluded due to use of a medication that impacts wakefulness. Other exclusion criteria included a history of conditions (e.g., seizures, migraines) that may be aggravated by sleep loss. All volunteers were screened by a study physician and eligibility was ultimately determined at the discretion of the study physician. Although, there are no known effects of 36 hours of sleep deprivation on a developing fetus, to err on the side of safety, pregnant females were excluded from participation. Both female volunteers were administered a pregnancy test prior to the sleep deprivation phase of the study, the results of which were negative. Volunteers were recruited using word of mouth, email notifications, and posted flyers.

Procedure

Participants were asked to abstain from caffeine and alcohol consumption for 48 hours prior to arrival at the laboratory. Participants arrived at the laboratory on the night (approximately 2000 hours) of Day 1. Upon entering the laboratory, participants were briefed on the study's objective and procedures. After written consent was obtained, participants completed the medical screening form and met with the study physician to determine eligibility. Ineligible participants were thanked for their time and interest and dismissed from the laboratory. Eligible participants were given a tour of the laboratory and familiarized with the USAARL sleep suites. On Day 2, participants were awakened at 0700 hours. Participants had 2 hours for personal hygeine and breakfast before the first testing session at 0900 hours. At each test session, participants completed the brief demographics questionnaire (only completed at the initial test session), mood assessment, Stanford Sleepiness Scale, and Psychomotor Vigilance Task (PVT). Additionally, participants completed a set of observation and intervention tasks each measuring either causal judgment or correlation detection. For one of these tasks, participants completed a novel scenario using the EST 2000 system. In this task, participants learned friend/foe targets in a training phase (i.e., the training phase was structured to emulate well established causal learning tasks) after which they completed a test phase. The order of the tasks was counterbalanced per subject per test session to reduce any order effects. After initial (baseline) testing was complete, participants stayed awake until the next test session at 0700 hours (Day 3). At this point, participants had been awake for 24 hours. Participants were tested again at 1900 hours (at 36 hours awake). At 2300 hours on Day 3, participants were allowed to rest for 8 hours. Participants were awakened at 0700 hours (after 8 hours of recovery sleep) and given two hours for personal hygeine and breakfast before their 0900 test session. Participants were released from the study no later than 1200 hours on Day 4 after a medical assessment by the study phyician. In total, all participants were tested in 4 test sessions; baseline, 24 hours awake, 36 hours awake, and after 8 hours of recovery sleep. Total administration time was approximately 60 minutes. During the non-testing periods, participants were allowed to entertain themselves with the amenities of the USAARL facility, including a pool table, reading material, internet, movies and video games. Figure 2 illustrates the test schedule.

	Day 1	Day 2	Day 3	Day 4
0700		Wake up/Shower/ Breakfast	Test Session 2 (24 hours awake)	Wake up/ Shower/ Breakfast
0800				
0900		Test Session 1 (Baseline)	Shower/Breakfast	Test Session 4 (Recovery Sleep)
1000			Break	
1100		Break		Vitals Check and meeting with study physician
1200		Lunch	Lunch	Release
1300		Recreational Activities	Recreational Activities	
1400				
1500				
1600				
1700		P/T or Recreation	P/T or Recreation	
1800		Dinner	Dinner	
1900			Test Session 3 (36 hours awake)	
2000	Arrival at Lab/Consent			
2100	Medical Screening/Eligibility		Break	
2200				
2300	Sleep		Sleep	

Figure 2. Test schedule.

Task battery

Engagement Skills Trainer 2000

The EST 2000 is a U.S. Army's small arms training device. This device is used in the U.S. Army Infantry Schools Basic Rifle Marksmanship (BRM) strategy and allows for weapons training in a controlled (simulated) environment. As can be seen in figure 3, a participant fires from a lane (the USAARL laboratory has a five-lane configuration) at "targets" which appear on a projection screen at a distance of 26 feet and 3 inches from the firing line. The weapons have been modified to use with the EST 2000 but maintain their form, fit, feel, and function.

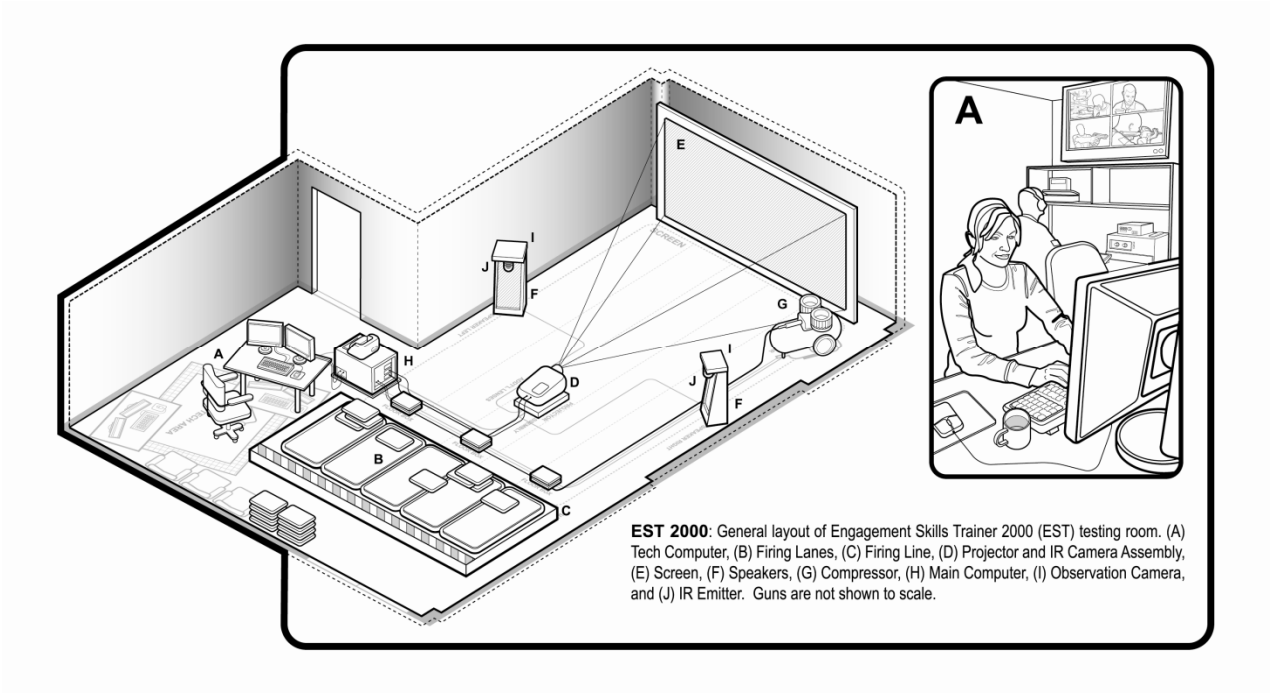

Figure 3. EST 2000 configuration.

A well-established reliable and valid measure of correlation detection was adapted for use with the EST 2000. In this task, there is a training phase and test phase for each test session. For this study, in the training phase, participants completed 8 trials. On each trial, participants were presented with two targets as depicted in figure 4. Participants were not told which target was friend and which was foe but had to learn this information. In order to do so, participants chose a target to shoot. They were then given auditory feedback as to whether their choice was correct (a correct choice would be shooting the foe target, and an incorrect choice would be shooting the friendly target). After the training session, participants took a short break (approximately 2 to 4 minutes) before beginning the test phase. In the test phase, participants completed a scenario similar to that in the training phase without auditory feedback. The key dependent variables in this task were reaction time, trial accuracy, and identification of foe target (participants were asked which target was foe and given the option of one circle target, two circle target, or could not tell). In addition, for a shot that hits a target, the horizontal and vertical distance from the center of target was also collected. The designation of friend and foe targets was random at each test session and participants were told that each time they completed the task that they must re-learn the identity of the targets.

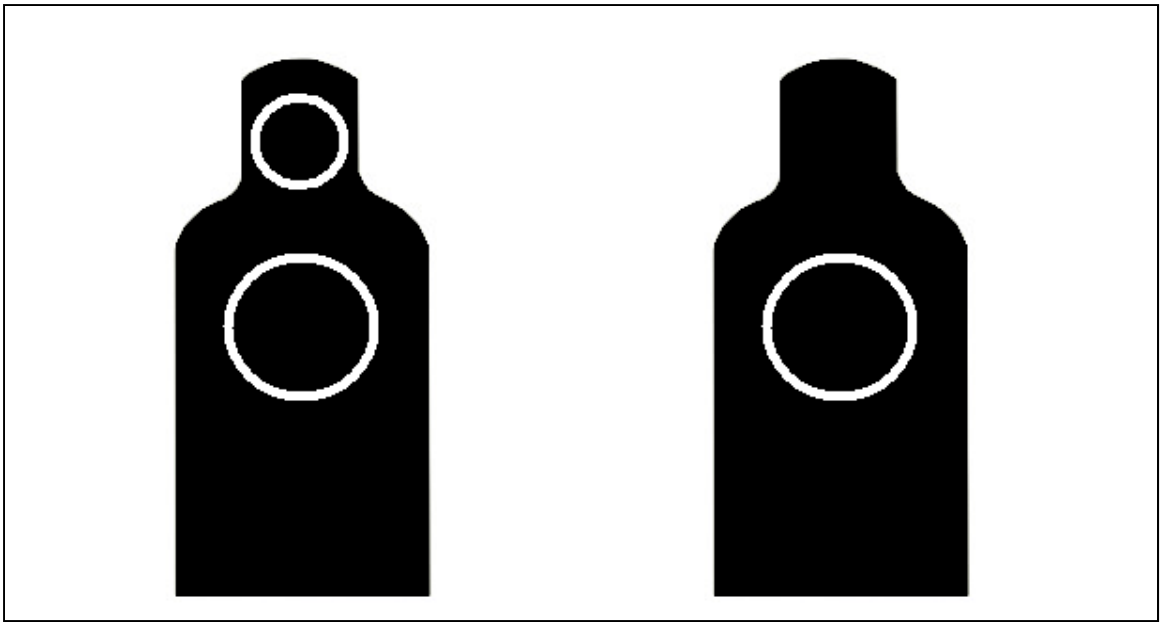

Figure 4. Targets in EST 2000 task.

Intervention task (causal judgment)

In the causal judgment task, participants were presented with a short cover story describing a causal candidate and an effect variable. To make a judgment about the relationship between the causal candidate and effect, participants evaluated some information about the occurrence and non-occurrence of the two variables. In this task, on a series of trials, participants allowed or prevented the causal candidate from occurring and observed the occurrence and non-occurrence of the effect variable. Participants then made a judgment about the relationship between the variables. Four concrete scenarios commonly employed in the causal judgment literature (e.g., fertilizers and plant growth, diets and weight loss, foods and allergic reactions, drugs and pain relief) were used such that a different scenario was presented at each test session. This intervention or "doing" on the part of the participant is thought to be essential to learning cause and effect relationships thus this task measures the ability to perceive causal relationships (Lagnado & Sloman, 2004). This task was completed on a computer using the Psychology Software Tools experiment generator software program E-prime version 2.0.

Observation task (correlation detection)

In the correlation detection task, participants were presented with a short cover story describing two variables and followed by a summary of observations of the occurrence and non-occurrence of the two variables. Participants were then asked to make a judgment about the relationship between the two variables on a scale from -10 (strong preventative relationship) to +10 (perfect generative relationship) and rate their confidence in that judgment on a scale from 0 (not confident at all) to 10 (very confident). Four concrete scenarios commonly employed in the causal judgment literature (e.g., fertilizers and plant growth, diets and weight loss, foods and allergic reactions, drugs and pain relief) were used such that a different scenario was presented at each test session. This task measures the ability to perceive correlational relationships (e.g.,

Hattori & Oaksford, 2008). This task was completed on a computer using the Psychology Software Tools experiment generator software program E-prime version 2.0.

Sleep assessment

The Stanford Sleepiness Scale was used to subjectively measure participants' sleepiness (e.g., Hoddes, et al., 1973; Killgore & McBride, 2006).

Subjective assessments

Administered at each test session were two assessments of subjective mood state and alertness; a visual analogue scale (VAS) response format (Penetar et al., 1993) and the Profile of Mood States (POMS), a 65 item adjective checklist with a Likert response scale (McNair, Lorr, & Droppleman, 1992). There are 8 state (versus trait) adjectives measured by the VAS: alert, anxious, energetic, confident, irritable, jittery, sleepy, and talkative. The POMS questionnaire yields six sub-scale scores: tension-anxiety, depression-dejection, anger-hostility, vigor-activity, fatigue-inertia, and confusion-bewilderment. A demographics questionnaire was also administered.

Psychomotor Vigilance Task

A 5-minute psychomotor vigilance task (PVT) was used to measure alertness by means of reaction time. The PVT was administered on a hand held personal digital assessment (PDA). This device was validated at Walter Reed Army Institute of Research and displayed in figure 5 (Thorne, Johnson, Redmond, Sing, & Belenky, 2005). Data collected from the PVT included mean response time, number of responses over 500 milliseconds (i.e., minor lapse), and number of responses over 3 seconds (i.e., major lapse).

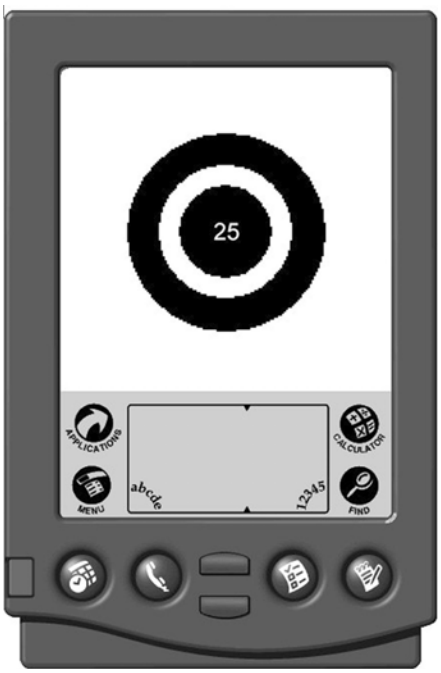

Figure 5. Hand-held PVT device (Thorne, et al., 2005).

Results

Sleep assessment

A repeated measures analysis of variance (ANOVA) of the Stanford Sleepiness Scale responses showed a significant main effect of session on self-reported sleepiness, $F(3, 57) = 44.29$, $p < 0.001$ (figure 6). Results of subsequent two-tailed paired samples *t*-tests showed that the mean score at baseline was significantly lower than that at 24 hours awake, $t(19) = -10.05$, $p < 0.001$, and that at 36 hours awake, $t(19) = -7.63$, $p < 0.001$, but not different from the mean score after recovery sleep. These tests also showed that the mean score after recovery sleep was significantly lower than that at 24 hours awake, $t(19) = 1.79$, $p < 0.001$, and that at 36 hours awake, $t(19) = 1.71$, $p < 0.001$. Mean scores at 24 hours awake was not significantly different from that at 36 hours awake.

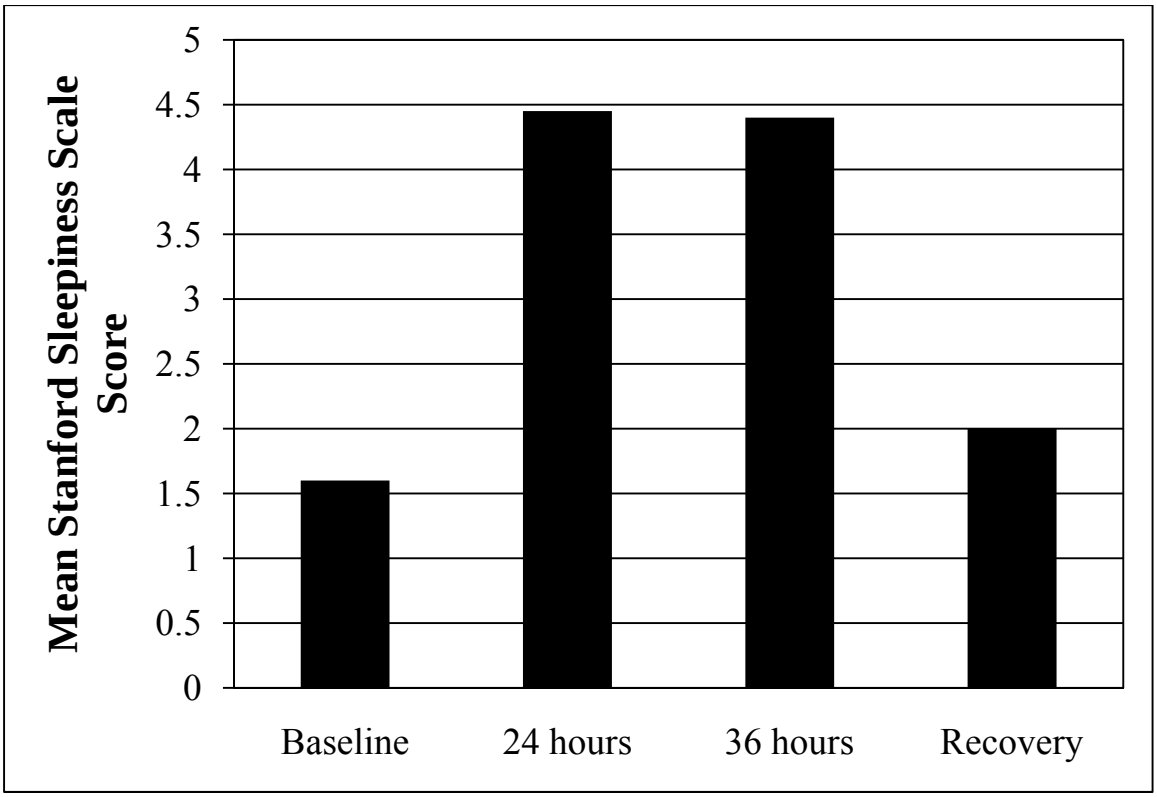

Figure 6. Results of the Stanford Sleepiness Scale.

Subjective assessments

Mood and alertness visual analogue scale

The assumption of equal variances was violated for the following dependent measures: irritable, jittery, and sleepy. Thus, a Greenhouse-Geisser correction was applied. A repeated measures multivariate analysis of variance (MANOVA) showed a significant main effect of session on the following state adjectives: alertness, $F(3, 54) = 34.82$, $p < 0.001$; energy, $F(3, 54) = 23.82$, $p < 0.001$; confidence, $F(3, 54) = 15.68$, $p < 0.001$; sleepiness, $F(2.35, 42.26) = 16.25$, $p < 0.001$; and talkativeness, $F(3, 54) = 5.12$, $p = 0.003$ (figure 7). Specifically, participants reported feeling more alert, energetic, confident, and less sleepy at baseline than at 24 hours (table 1) and at 36 hours awake (table 2). Likewise, participants felt more alert, energetic, confident, and less sleepy after recovery sleep than at 24 hours awake, (table 3), and at 36 hours awake (table 4). Additionally, participants reported feeling less talkative at 24 hours awake than at baseline, $t(19) = 2.87$, $p = 0.01$, and after recovery sleep, $t(19) = -3.25$, $p = 0.004$. These results suggest that participants did, in fact, feel sleep deprived.

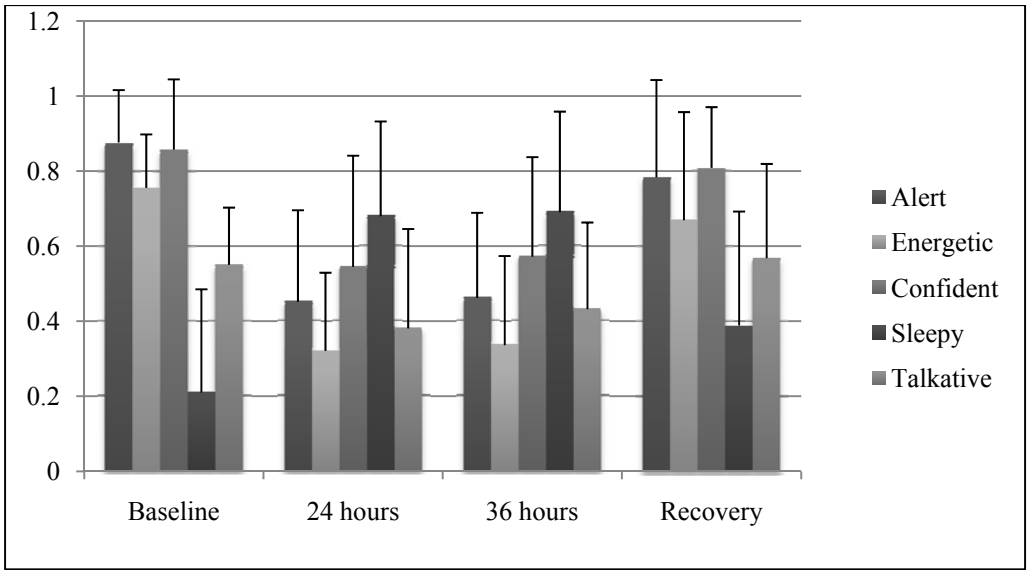

Figure 7. Results of mood and alertness VAS. Error bars represent standard error of the mean.

Table 1.
Results of paired comparison *t*-tests for mean VAS adjective ratings at baseline and at 24 hours awake.

Adjective	Baseline Mean	24 hours Mean	Statistic
Alert	0.85	0.45	$t(19) = 8.04$, $p < 0.001$
Energetic	0.74	0.32	$t(19) = 9.59$, $p < 0.001$
Confident	0.85	0.54	$t(19) = 4.56$, $p < 0.001$
Sleepy	0.20	0.70	$t(19) = -6.30$, $p < 0.001$

Table 2.
Results of paired comparison *t*-tests for mean VAS adjective ratings at baseline and at 36 hours awake.

Adjective	Baseline Mean	36 hours Mean	Statistic
Alert	0.85	0.44	$t(19) = 9.09$, $p < 0.001$
Energetic	0.74	0.33	$t(19) = 7.20$, $p < 0.001$
Confident	0.85	0.56	$t(19) = 5.05$, $p < 0.001$
Sleepy	0.20	0.71	$t(19) = -5.21$, $p < 0.001$

Table 3.
Results of paired comparison *t*-tests for mean VAS adjective ratings at recovery and at 24 hours awake.

Adjective	Recovery Mean	24 hours Mean	Statistic
Alert	0.78	0.45	$t(19) = -5.78$, $p < 0.001$
Energetic	0.67	0.32	$t(19) = -4.97$, $p < 0.001$
Confident	0.81	0.54	$t(19) = -5.41$, $p < 0.001$
Sleepy	0.38	0.70	$t(19) = 3.82$, $p < 0.001$

Table 4.
Results of paired comparison *t*-tests for mean VAS adjective ratings at recovery and at 36 hours awake.

Adjective	Recovery Mean	36 hours Mean	Statistic
Alert	0.78	0.44	$t(19) = -6.82$, $p < 0.001$
Energetic	0.67	0.33	$t(19) = -4.56$, $p < 0.001$
Confident	0.81	0.56	$t(19) = -5.01$, $p < 0.001$
Sleepy	0.38	0.71	$t(19) = 4.03$, $p < 0.001$

Profile of Mood States

The assumption of equal variances was violated as determined by Mauchly's test for equal variances for the following dependent measures: fatigue-inertia sub-scale. Thus, a Greenhouse-Geisser correction was applied. A repeated measures MANOVA showed a significant main effect of session for the vigor-activity sub-scale, $F(3, 57) = 39.61$, $p < 0.001$, the fatigue-inertia sub-scale, $F(2.28, 43.28) = 29.41$, $p < 0.001$, and a marginally significant effect for the confusion-bewilderment sub-scale, $F(3, 57) = 2.31$, $p = 0.086$ (figure 8). Subsequent paired comparisons *t*-tests showed that participants reported being the least fatigued and most vigorous at baseline (see tables 5-9). Paired comparisons *t*-tests also showed that participants reported being more confused/bewildered at 36 hours awake than at 24 hours awake, $t(19) = 2.27$, $p = 0.035$.

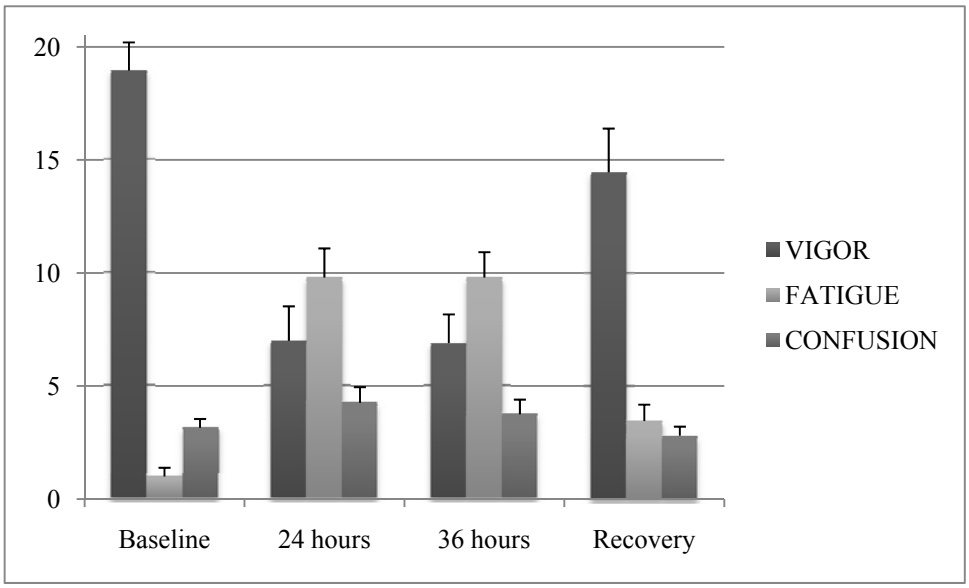

Figure 8. Results of POMS. Error bars represent standard error of the mean.

Table 5.
Results of paired comparison *t*-tests for mean POMS sub-scale scores at baseline and at 24 hours awake.

Sub-scale	Baseline Mean	24 hours Mean	Statistic
Vigor-Activity	18.95	7.00	$t(19) = 11.33$, $p < 0.001$
Fatigue-Inertia	1.00	9.80	$t(19) = -7.68$, $p < 0.001$

Table 6.
Results of paired comparison *t*-tests for mean POMS sub-scale scores at baseline and at 36 hours awake.

Sub-scale	Baseline Mean	36 hours Mean	Statistic
Vigor-Activity	18.95	6.90	$t(19) = 9.49$, $p < 0.001$
Fatigue-Inertia	1.00	9.80	$t(19) = -8.42$, $p < 0.001$

Table 7.
Results of paired comparison *t*-tests for mean POMS sub-scale scores at baseline and at recovery.

Sub-scale	Baseline Mean	Recovery Mean	Statistic
Vigor-Activity	18.95	14.45	$t(19) = 2.73$, $p = 0.013$
Fatigue-Inertia	1.00	3.45	$t(19) = -3.49$, $p = 0.002$

Table 8.
Results of paired comparison *t*-tests for mean POMS sub-scale scores at recovery and at 24 hours awake.

Sub-scale	Recovery Mean	24 hours Mean	Statistic
Vigor-Activity	14.45	7.00	$t(19) = -5.07$, $p < 0.001$
Fatigue-Inertia	3.45	9.80	$t(19) = 4.40$, $p < 0.001$

Table 9.
Results of paired comparison *t*-tests for mean POMS sub-scale scores at recovery and at 36 hours awake.

Sub-scale	Recovery Mean	36 hours Mean	Statistic
Vigor-Activity	14.45	6.90	$t(19) = -5.76$, $p < 0.001$
Fatigue-Inertia	3.45	9.80	$t(19) = 5.01$, $p < 0.001$

PVT data analysis

Reaction time

Mean PVT reaction time data are presented in figure 9. A repeated measures ANOVA revealed a significant effect of session with regard to mean reaction time data, $F(1.537, 29.201) = 6.418$, $p = 0.008$. Pairwise comparisons revealed that participants had a slower mean reaction time after 24 hours of sleep deprivation compared to the baseline ($p = 0.003$) and recovery sessions ($p < 0.001$). In addition, participants' mean reaction time was significantly faster in the recovery session than the baseline session ($p = 0.044$).

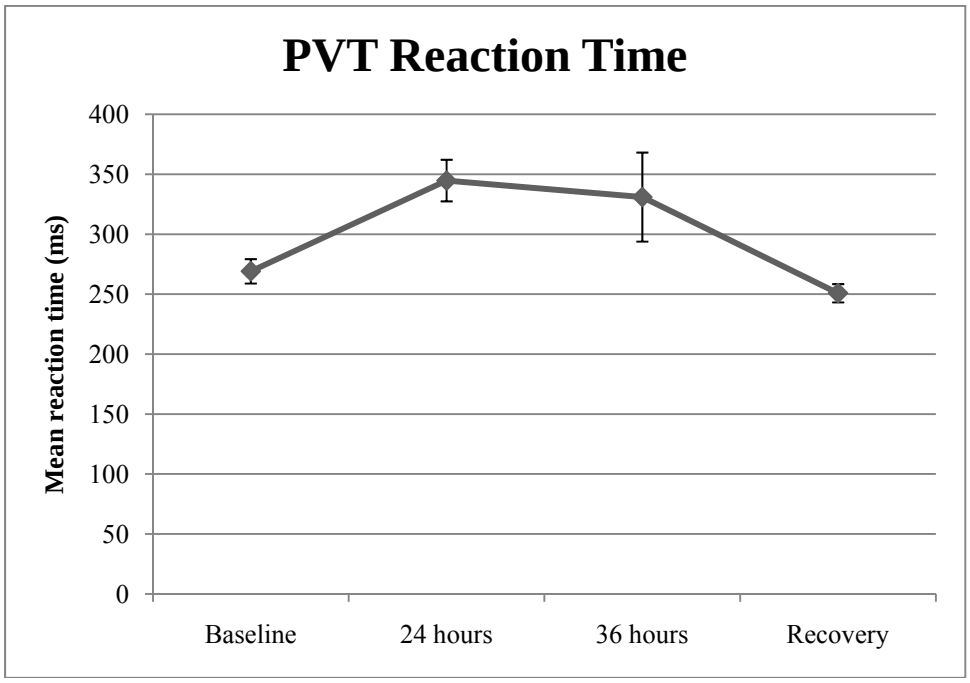

Figure 9. Mean PVT reaction time data by session. Error bars represent standard error of the mean.

Lapses

Mean PVT lapse data are presented in figure 10. A lapse is defined as a response greater than 500 milliseconds. A repeated measures ANOVA revealed a significant effect of session with regard to mean lapses, $F(2.009, 38.178) = 10.849$, $p < 0.001$. Pairwise comparisons revealed that participants had a greater mean number of lapses after 24 hours of sleep deprivation than at the baseline ($p = 0.001$) and recovery sessions ($p < 0.001$).

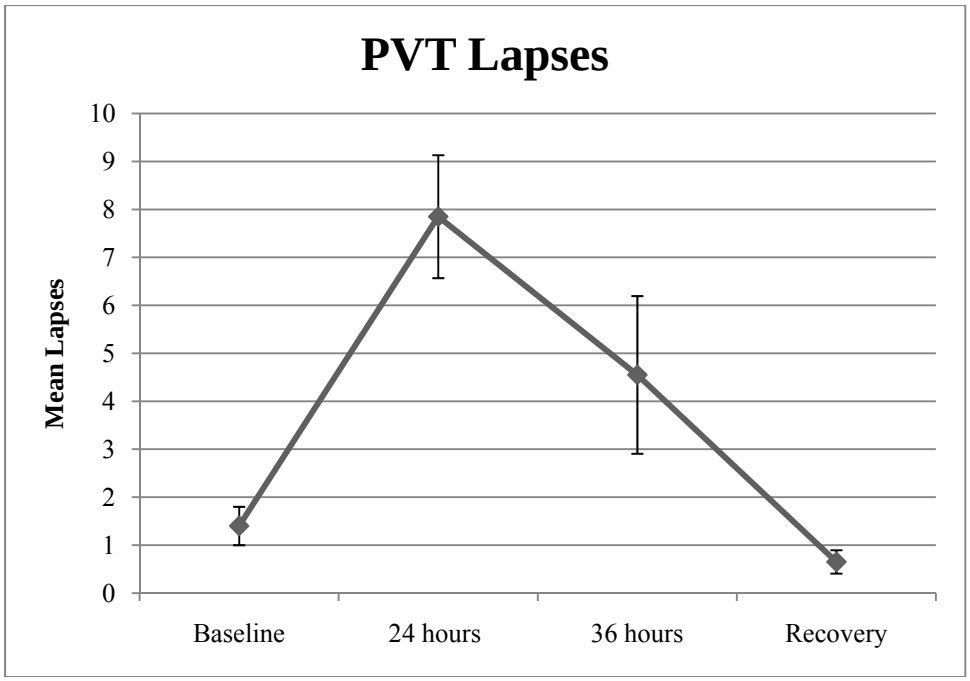

Figure 10. Mean PVT lapse data by session. Error bars represent standard error of the mean.

Engagement Skills Trainer 2000 results

Two sets of analyses were conducted on the EST data; a set to assess correlation detection and judgment accuracy and a set to assess marksmanship accuracy. Mean reaction time, mean horizontal displacement (shot accuracy), mean vertical displacement (shot accuracy), and the proportion of hits was calculated for each session. Each session consisted of two phases: a training phase and a test phase. In order to assess accuracy in the judgment aspect of the task, two other variables were calculated first; *generated sample type* and *test decision*. In the training phase, participants saw eight pairs of targets, fired at one target, and received auditory feedback indicating if they shot the foe target or friend target. Thus, there are four potential observations as illustrated by the contingency table in figure 10. To calculate the *generated sample type* in the training phase, equation 1 was used if the participant varied his/her behavior (fired at least once at each target type).

$$\Delta P = A/(A+B) - C/(C+D) \quad (1)$$

If equation 1 yielded a positive number, then the *generated sample type* was labeled *two-circles foe*, if a negative number, then labeled *one-circle foe*, and if zero then labeled *random* (three possible categories for this nominal variable). If the participant did not vary his/her behavior, then the ratio of Cell A to B observations or Cell C to D observations (dependent upon at which target he/she shot) was calculated. If Cell A > B then *positive*, if A < B then *negative*, and if A = B then *zero*. Likewise, if Cell C > D then *negative*, if C < D then *positive*, and if C = D then *zero*. In the test phase, participants saw eight pairs of targets and were instructed to shoot the target he/she believed to be the foe target based on the training phase feedback. Thus, *test decision* was determined by the frequency of shots fired at each target. If more shots were fired at the one-

circle target, the *test decision* was labeled as *one-circle foe.* If more shots were fired at the two-circles target, then the *test decision* was labeled *two-circles foe*. If equal shots were fired at both targets, then the *test decision* was labeled *random* (three possible categories for this nominal variable).

		Feedback	
		Incorrect	Correct
Target	One-Circle	A	B
	Two-Circles	C	D

Figure 11. EST 2000 task contingency table. The cell entries were used to calculate the *generated sample type*. The cell entries A through D denote the number of observations in each cell of the contingency table.

Correlation detection and judgment accuracy

To evaluate the accuracy of judgments in the friend/foe detection task, *generated sample type* and *test decision* were compared. If these two were congruent, for example, if both *generated sample type* and *test decision* equaled *two circles foe*, then the judgment was scored as accurate (1). If *generated sample type* and *test decision* were incongruent, then the judgment was scored as inaccurate (0). A repeated measures ANOVA was conducted with accuracy as the dependent measure and session as the independent measure and was not significant, $F(3, 57) = 1.647$, $p = 0.189$ (figure 12).

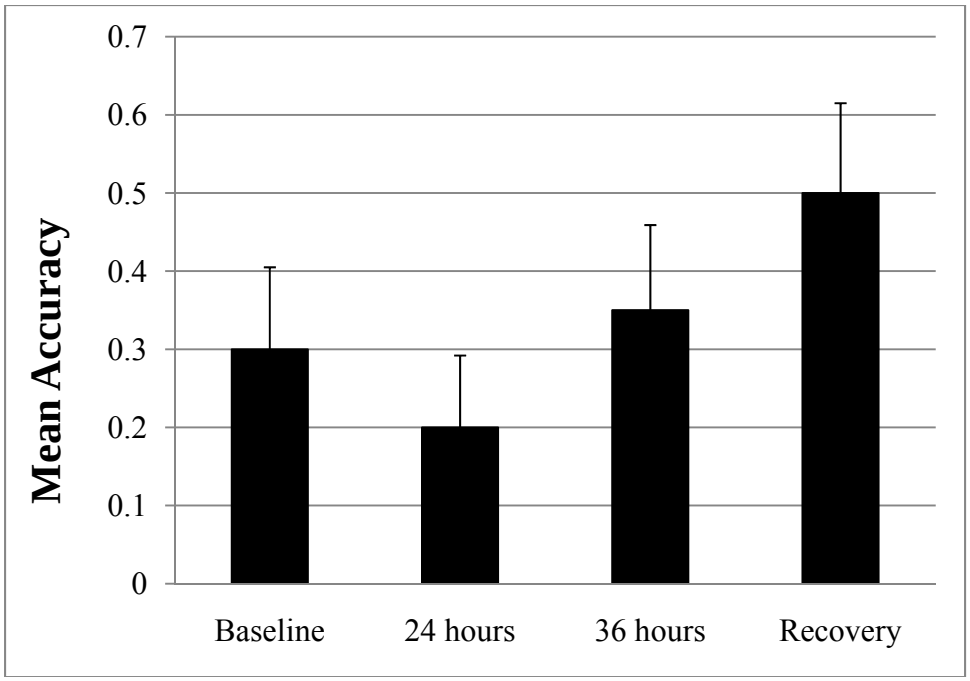

Figure 12. Mean accuracy in EST 2000 friend/foe detection judgment task. Error bars represent standard error of the mean.

As stated previously, it was hypothesized that current judgments would be related to prior judgments, more so than by current observations under conditions of sleep deprivation. Given the small sample size and very large standard errors of parameter estimates, the proposed multinomial logistic regression was deemed an inappropriate method of analysis given that the large standard errors suggest an unreliable model. Subsequently, chi-square tests were conducted which did not yield any significant results.

Marksmanship accuracy

To assess the marksmanship accuracy of the participants, repeated measures ANOVAs and paired samples *t*-tests were conducted. The assumption of equal variance was violated for the analyses of horizontal displacement, vertical displacement, and reaction time as shown by Mauchly's Test of Sphericity. Therefore, a Greenhouse-Geisser correction was applied. There was not a significant main effect of session on horizontal displacement, $F(1.963, 37.291) = 1.248$, $p = 0.298$, or on reaction time, $F(2.127, 40.415) = 2.97$, $p = 0.06$ (figures 13 and 14, respectively). However, there was a significant main effect of session on vertical displacement, $F(1.963, 38.968) = 6.052$, $p = 0.005$, and the proportion of "hits," $F(3, 57) = 9.736$, $p < 0.001$ (figures 15 and 16, respectively). Paired samples *t*-tests indicated that participants performance tended to strengthen over the sessions rather than worsen under conditions of sleep deprivation as predicted (summarized in table 10).

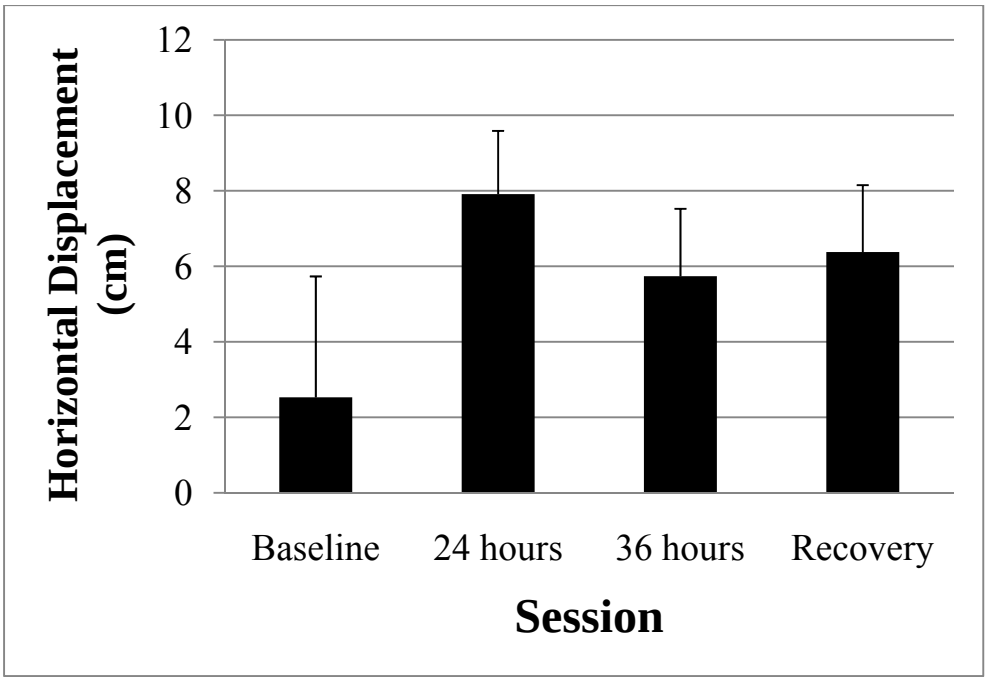

Figure 13. Mean horizontal displacement by session. Error bars represent standard error of the mean.

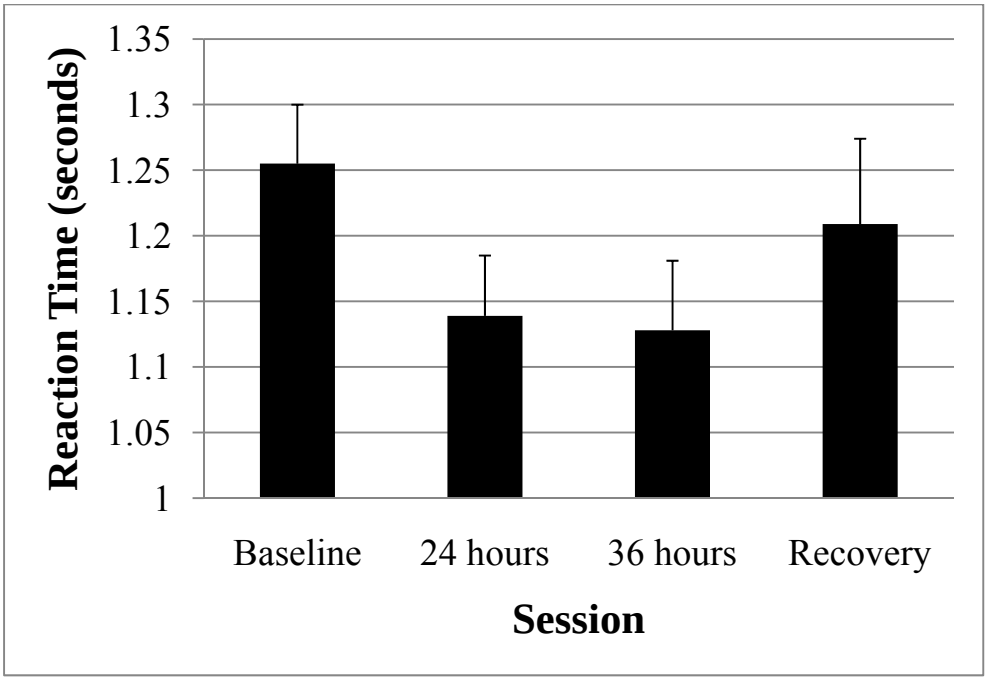

Figure 14. Mean reaction time by session. Error bars represent standard error of the mean.

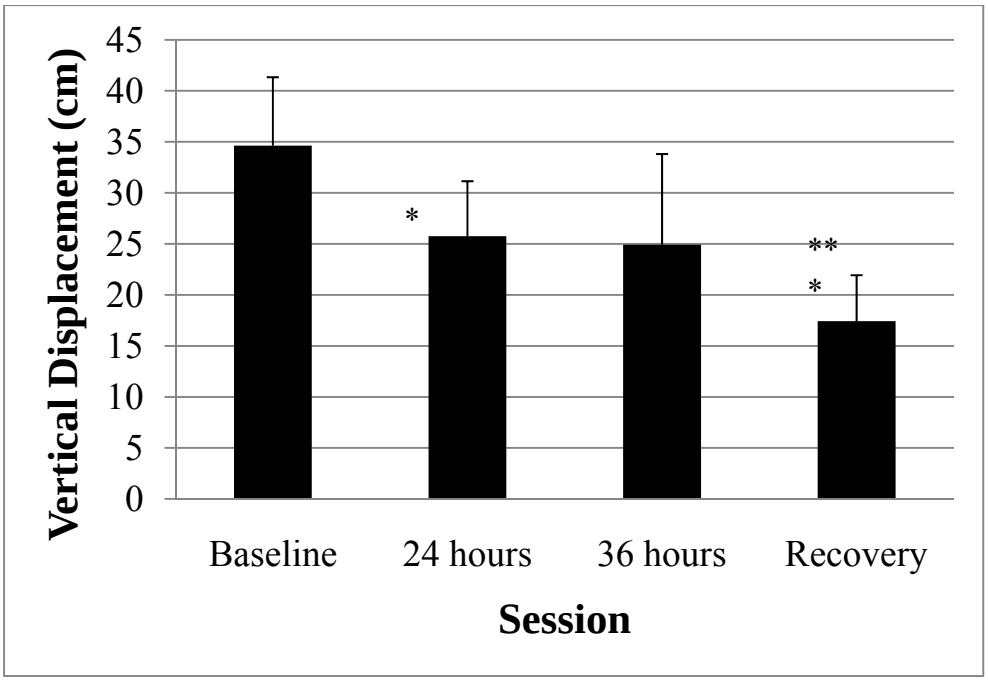

Figure 15. Mean vertical displacement by session. Note that * indicates significant difference from baseline and ** indicates significant difference from 24 hours awake. Error bars represent standard error of the mean.

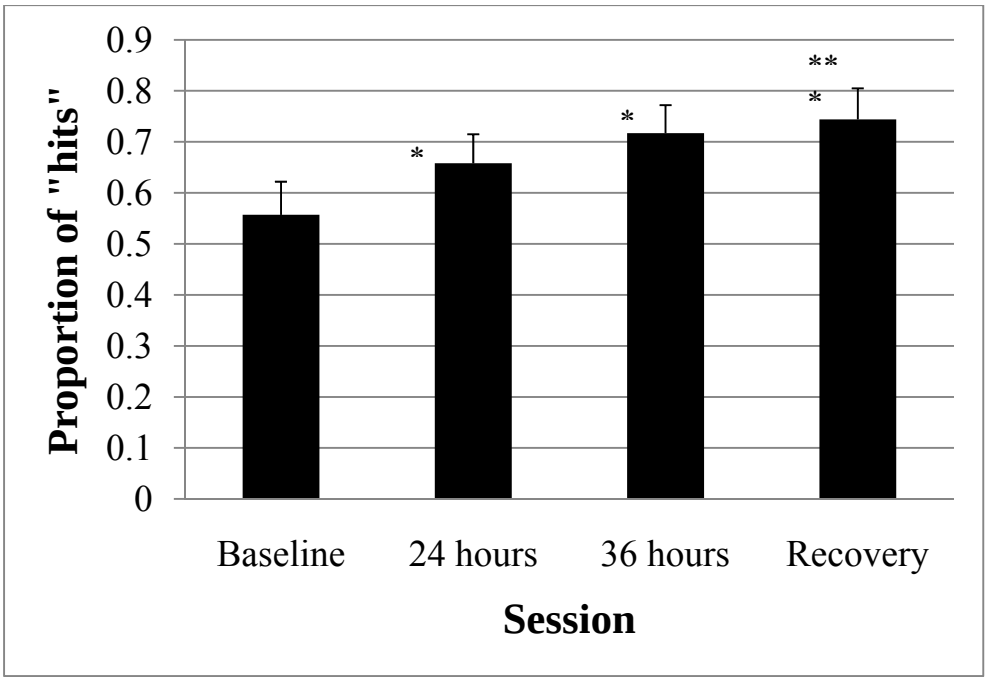

Figure 16. Mean proportion of “hits” by session. Note that * indicates significant difference from baseline and ** indicates significant difference from 24 hours awake. Error bars represent standard error of the mean.

Table 10.
Significant results of paired samples *t*-tests for marksmanship data.

Dependent Measure	Comparison	df	*t* value	*p* value
Proportion of "hits"	baseline – 24 hours	19	-2.711	0.014
Proportion of "hits"	baseline – 36 hours	19	-3.869	0.001
Proportion of "hits"	baseline – recovery	19	-4.486	< 0.001
Proportion of "hits"	24 hours – recovery	19	-2.299	0.033
Vertical displacement	baseline – 24 hours	19	2.879	0.010
Vertical displacement	baseline – recovery	19	4.628	< 0.001
Vertical displacement	24 hours – recovery	19	3.083	0.006

Intervention task (causal judgment)

To analyze the intervention task causal judgments, first, the samples generated by the participants in the task were assessed. Across all participants and all sessions, 1200 samples were generated of which 599 were correlationally indeterminate. Specifically, participants did not vary in their behavior choice thus the level of the causal candidate did not vary, rendering the correlational relationship indeterminate. However, for purposes of analysis, samples were categorized with respect to the ratio of Cell A observations to Cell B observations (see figure 1), when the participant chose to apply the causal candidate on each observation, such that if the ratio was greater than one then the sample was categorized as positive, equal to one then the sample was no relationship, and less than one then the sample was negative. When the participant chose not to apply the causal candidate on each observation the generated samples were categorized with respect to the ratio of Cell C to Cell D observations, such that if the ratio was less than one then the sample was categorized as positive, equal to one then no relationship, and greater than one as negative. Further, this independent variable will be referred to as the *generated sample relationship type* which has three levels; positive, zero, and negative.

The relationship ratings data were analyzed using a 4 (session) X 3 (*generated sample relationship type*) repeated measures ANOVA which revealed a significant main effect of *generated sample relationship type*, $F(2, 16) = 22.57$, $p < 0.001$ (figure 17). Subsequent pairwise comparisons revealed that participants rated *positive generated samples*, $M = 3.9$, 95% CI (1.94, 4.84), significantly higher than *negative generated sample,* $M = 0.96$, 95% CI (-0.75, 2.67), and *zero generated samples*, $M = 0.12$, 95% CI (-1.97, 2.21). Comparisons between the other two groups were not statistically significant at $p > 0.05$. Participants rated how confident they felt regarding these judgments and these data were also analyzed using a 4 (session) X 3 (*generated sample relationship type*) repeated measures ANOVA which yielded no significant effects (figure 18). There were no significant effects of session for either dependent variable. It should be noted that eleven participants were excluded from the ANOVA for incomplete data (i.e., given that the task requires participants to generate their own samples, experimental control is

sacrificed, thus some participants did not generate each possible trial type in the intervention task resulting in incomplete data). The observed power equaled 0.47 which may be insufficient to detect an effect of session.

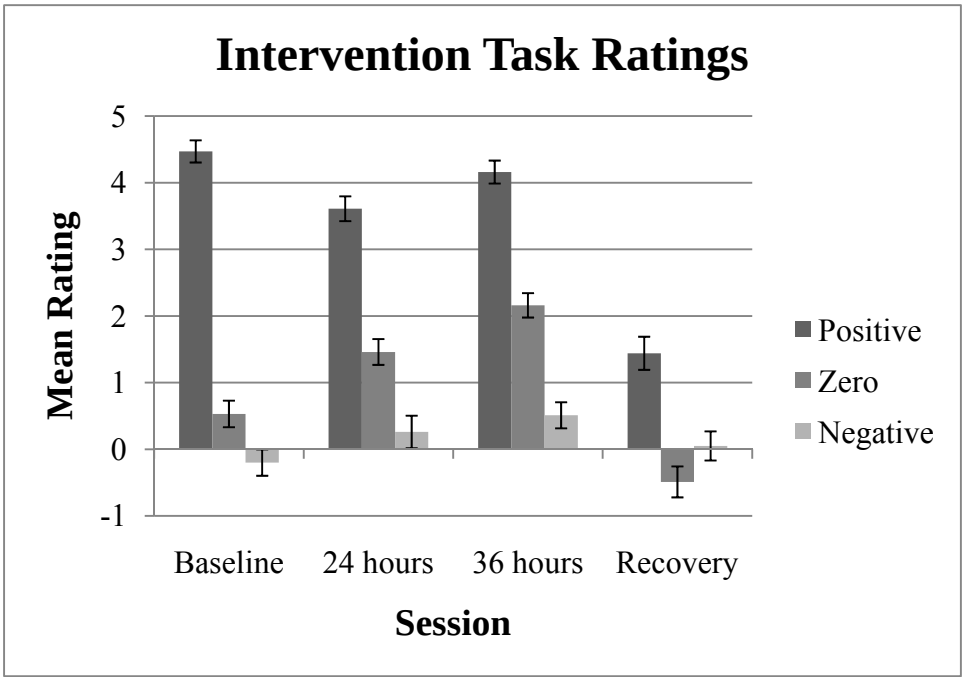

Figure 17. Intervention task relationship rating results. Error bars represent standard error of the mean.

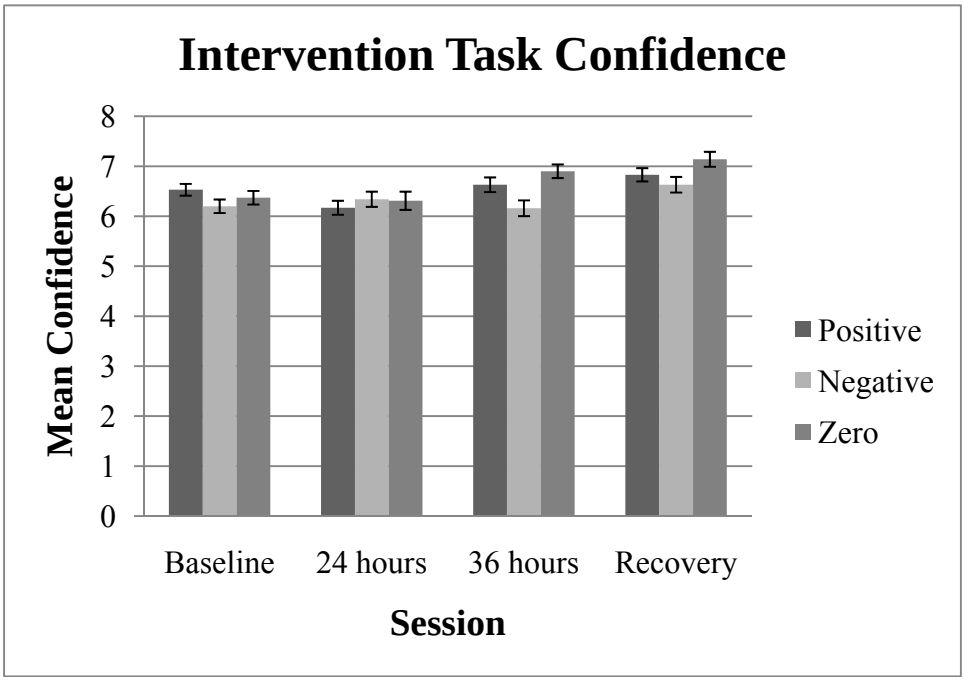

Figure 18. Intervention task confidence results. Error bars represent standard error of the mean.

Observation task (correlation detection)

This task presented participants with five types of samples (nine trials of each *sample correlational relationship type* thus a total of 45 trials): positively correlated, negatively correlated, uncorrelated/zero correlation, correlationally indeterminate indicative of a positive relationship, and correlationally indeterminate indicative of a negative relationship (following the same criteria as described for determining *generated sample type* in the intervention task).

The mean ratings data were analyzed using a 4 (session) X 5 (*sample correlational relationship type*) repeated measures ANOVA which revealed a significant main effect of *sample correlational relationship type*, $F(4, 76) = 76.31$, $p < 0.001$, and a significant interaction between *sample correlational relationship type* and session, $F(2, 228) = 6.23$, $p < 0.001$ (figure 19). Bonferroni corrected post-hoc pairwise comparisons indicated that participants rated positively correlated samples, $M = 6.88$, 95% CI (5.97, 7.80), significantly higher than all other *sample correlational relationship types.* The comparisons also showed that participants rated negatively correlated samples, $M = -3.98$, 95% CI (-5.70, -2.27), significantly lower than all other *sample correlational relationship types.* Indeterminate positive samples, $M = 3.17$, 95% CI (2.37, 3.97), were rated significantly lower than positively correlated samples and significantly greater than all other *sample correlational relationship types*. Alternatively, indeterminate negative samples, $M = -0.95$, 95% CI (-1.97, 0.08), were rated significantly greater than negatively correlated samples and significantly lower than all other *sample correlational relationship types*. Finally, zero correlation samples, $M = 1.56$, 95% CI (0.81, 2.31), were rated significantly greater than negatively correlated samples and indeterminate negative samples and rated significantly lower than positively correlated samples and indeterminate positive samples. In regard to the significant interaction, paired-samples *t*-tests revealed two significant differences at the bonferroni corrected alpha level ($\alpha = 0.002$). Specifically, mean ratings of negatively correlated samples given at 24 hours awake were significantly greater than mean ratings of negatively correlated samples given at baseline, 36 hours awake (marginally significant), and recovery sessions (table 11). There was also a marginally significant difference between mean ratings of positively correlated samples such that ratings given at 24 hours were less than ratings given at 36 hours. These results suggest that participants' accuracy in rating negatively correlated samples was decreased at 24 hours awake compared to baseline, 36 hours awake, and recovery sessions and for positively correlated samples greater accuracy was exhibited at 36 hours awake than 24 hours awake.

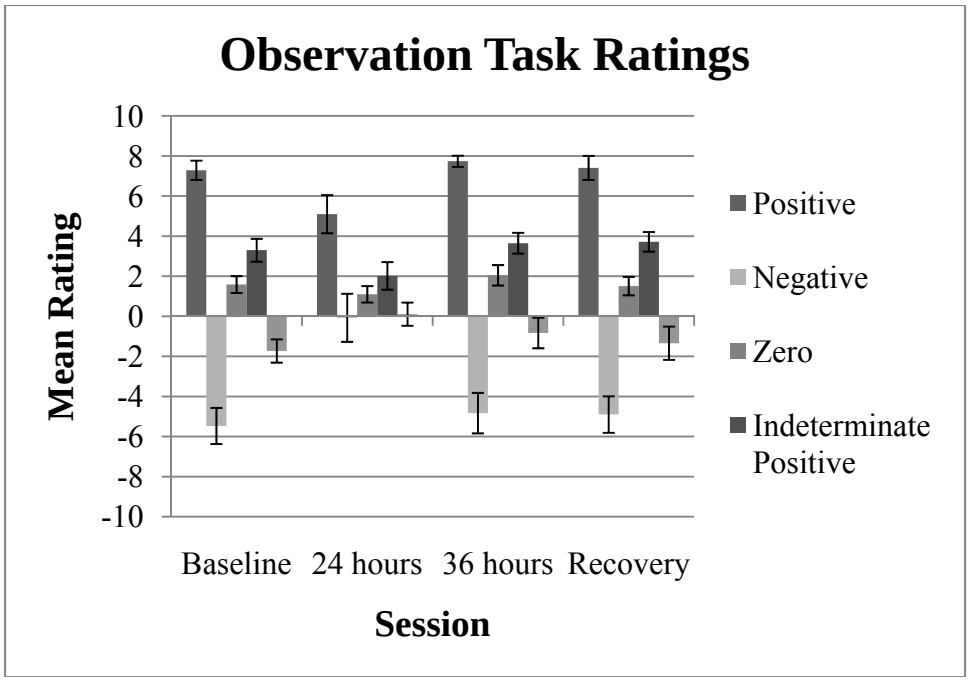

Figure 19. Observation task relationship rating results. Error bars represent standard error of the mean.

Table 11.
Results of paired samples *t*-tests for observation task.

Sample Type	Comparison	df	*t* value	*p* value
Positive	24 hours < 36 hours	19	-2.973	0.008*
Negative	Baseline < 24 hours	19	-3.785	0.001
Negative	24 hours > 36 hours	19	3.430	0.003*
Negative	24 hours > Recovery	19	4.082	0.001

* denotes marginal significance.

The mean confidence ratings data were also analyzed using a 4 (session) X 5 (*sample correlational relationship type*) repeated measures ANOVA which revealed a significant main effect of *sample correlational relationship type*, $F(4, 76) = 2.63$, $p = 0.041$ (figure 20). However, none of the post-hoc pairwise comparisons were significant at the Bonferroni corrected level.

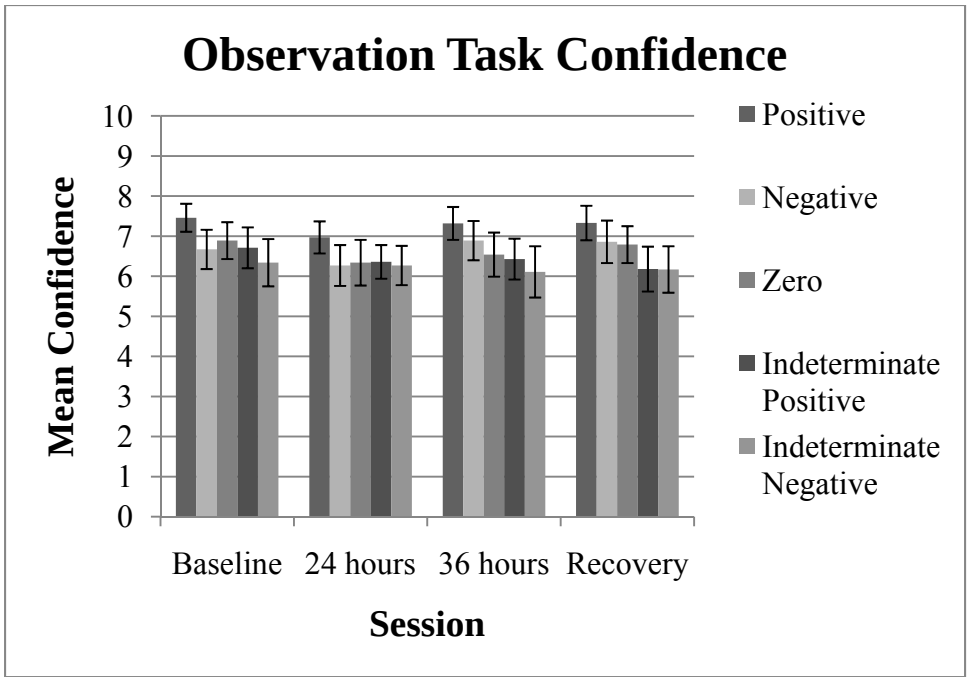

Figure 20. Observation task confidence results. Error bars represent standard error of the mean.

Discussion

The primary objective of this study was to evaluate the impact of sleep deprivation on the detection of correlational relationships and judgment of causality. Tasks measuring these cognitive abilities included two classic laboratory tasks administered on a computer and a version of the correlation detection task adapted for administration employing the weapons simulator. The main findings suggest that accuracy in detection and judgment of correlational and causal relationships declined for some types of relationships but not others during a period of sleep deprivation. Specifically, participants' ratings of uncorrelated samples and correlationally indeterminate samples were consistent across all four test sessions. In contrast, decreases in perception and judgment accuracy were limited to the observation and intervention tasks and the positively and negatively correlated relationships such that participants' ratings diminished with respect to accuracy when 24 hours sleep deprived. There are three main points of interest to discuss in regard to these findings; limitation of deficits to determinate positively and negatively correlated samples, judgment and perception in the marksmanship task were largely unaffected despite shown deficits in tasks sensitive to sleep deprivation (PVT), and discrepancy between performance at 24 hours awake and at 36 hours awake.

Deficits to positively and negatively correlated samples

The results indicate that judgment accuracy was compromised for some types of samples in the observation and intervention tasks. Consider that accuracy is reflected by the mean ratings such that the higher the mean rating of positive relationships (both determinate and indeterminate) the greater the accuracy. Alternatively, the closer the mean rating of negative relationships (both determinate and indeterminate) is to -10, the greater the accuracy. To apply

this definition of accuracy to the zero correlation samples, it is necessary to take into account the standard errors of the means. The mean ratings and standard errors for zero correlation samples reported suggest that the ratings were "close" to zero, slightly above, and not contaminated by a large number of extreme values in both directions (both positive and negative).

Deficits in observation task performance

Given the above definition of accuracy and the observation task results, one could argue that overall participants displayed greater accuracy in the assessment of determinate positive and negative relationships than in the assessment of indeterminate and zero correlation samples. Also, the results of the observation task indicate that judgments of zero and indeterminate correlation samples were seemingly unaffected by sleep deprivation whereas the judgment accuracy of positively and negatively correlated samples was compromised at 24 hours awake. A possible explanation for this may be that participants used a successful strategy in judgment of positive and negative relationships at baseline but then after 24 hours awake, participants abandoned this strategy and adopted an alternative approach as has been shown in other studies of sleep deprivation (Wimmer, Hoffman, Bonato, & Moffit, 1992; Horne, 1988). In judgment of correlationally indeterminate and zero correlation samples, however, the strategies implemented did not yield a level of accuracy high enough to show deficits under conditions of sleep deprivation. It is possible that participants changed their strategies given the depletion of cognitive resources under the condition of sleep deprivation. Arkes and Harkness (1983) demonstrated that under conditions of cognitive overload, people shift strategies (from a complex strategy to a simple strategy) in correlation judgment in an attempt to retain as much accuracy as possible. It is reasonable to suggest that updating strategies given the degree of one's cognitive resources may be applicable to conditions of sleep deprivation.

Positive bias

Another possible explanation for the observation task pattern of results is the demonstration of a bias for positively correlated relationships. In many aspects of cognition, a bias for affirmational information (positive relationships) over negational information (negative relationships) has been exhibited (e.g., Mandel & Lehman, 1998). Specific to the perception of correlational relationships and causal judgment, Kareev (1995) investigated a positive bias such that people are inclined to perceive positive relationships relative to negative or null relationships. Kareev argued that participants' goal is not to actually assess the relationship but to maximize the number of correct predictions about future events based on the observed relationship. By this argument, a positive bias may actually prove beneficial given that the use of this heuristic yields a hit rate of 0.5 or better for a two-valued criterion. The results of this study do in fact support this assertion such that while mean ratings of positive relationships decreased at 24 hours awake, they were still above zero. In other words, the mean ratings decreased from arguably high to moderate. Alternatively, the mean ratings in judgment of negative relationships regressed to zero. It should be noted also that a positive bias can be seen in the assessment of zero correlation samples across all four sessions which is a robust finding in the literature (see Shanks, 2004 for a review).

Deficits in intervention task performance

In the intervention task, judgments of positively correlated samples were more accurate than judgments of negative or zero correlation samples. In comparison to the observation task results, it appears from the data patterns that participants' were overall less successful in the intervention task. Given that the intervention task results were not significant with respect to session which is speculated to be a result of the inadequate power to detect an effect if one does truly exist, it is inappropriate to speculate further on the interpretation of the data pattern.

Tasks and abilities sensitive to sleep deprivation

The vigilance and mood measures (Stanford Sleepiness Scale, VAS, POMS, and PVT) were included to demonstrate that participants did, in fact, feel the effects of sleep deprivation given that these measures have repeatedly been shown to be sensitive to sleep deprivation. The results on these measures suggest that participants were sleep deprived through subjective reporting of fatigue and sleepiness and also showing decremented objective, vigilance performance. Accuracy was maintained in the EST task and marksmanship performance improved across sessions. One explanation for this discrepancy is that the more monotonous, tedious laboratory computer tasks did not engage the participants sufficiently to put forth the effort to overcome sleepiness while the more arousing task of the weapons simulator did engage the participants. Previous research on effects of sleep deprivation has shown that more basic tasks including reaction time and vigilance tasks show deficits under conditions of sleep deprivation. However, these deficits do not necessarily transfer to more complex tasks. It has been argued that high level complex tasks are relatively unaffected due to the arousal they generate and the need for energy to be expended to overcome any fatigue or sleepiness. A review of sleep deprivation and decision making (Harrison & Horne, 2000) critically examined this assumption and indicated that while this may be true for some tasks, there are tasks involving complex skills that rely heavily on prefrontal cortex function. Research has shown that this region of the brain is particularly impacted after as little as one night of sleep deprivation (e.g., Horne, 2000). According to this review, tasks that seem to be unaffected are complex, logical, interesting, and rule-based. Tasks that are uninteresting, monotonous, too simple, or too long in duration are affected. Additionally, tasks that require divergent skills (such as higher-order decision making which can involve unfamiliar scenarios) are affected by sleep deprivation.

Differences in sleep deprivation effects at 24 hours awake and 36 hours awake

A number of sleep deprivation studies have shown a drop in performance at 24 hours sustained wakefulness followed by an increase in performance abilities the second day (e.g., Caldwell, Caldwell, Brown, & Smith, 2004). In accordance with previous research on sleep deprivation, there was a discrepancy in judgment accuracy such that participants' accuracy decreased at 24 hours but was comparatively recovered at 36 hours. One possible reason for this resurgence in performance is that the circadian cycle rises during the day thus boosting performance (Eddy & Hursh, 2001). Another possible explanation is that participants were aware that they would be able to end the period of sustained wakefulness soon and subsequently were highly motivated to complete the tasks (Previc, et al., 2009).

Limitations and future studies

One limitation of the present study was the inexperience with the adapted correlation detection task administered via the weapons simulator. Specifically, it seems that participants struggled to understand this task and may have benefitted from multiple trials per session as well as a preliminary practice session. The marksmanship results strongly indicate the necessity of practice sessions. The investigators did not incorporate a practice session as previous experience with the laboratory computerized version of the task did not indicate that one was necessary. However, this previous experience was largely gained in an academic environment. The differences between a military and an academic environment with regard to the contextual settings, social cultures, and subject populations may have contributed to this discrepancy. Also, it is possible that the adaptation itself did not translate to an applied scenario with as much validity as assumed. Additional reliability and validation studies are necessary prior to future implementation.

Conclusions

Results of this study are inconclusive in regards to whether participants overweigh prior information under conditions of sleep deprivation. However, the results do suggest that accuracy in judgments were compromised for some relationship types (e.g., determinate positive and negative correlational relationships) under conditions where cognitive resources were taxed.

References

Akerstedt, T. 2006. Altered sleep/wake patterns and mental performance. Physiology & Behavior. 90: 209-218.

Alloy, L. B. and Tabachnik, N. 1984. Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. Psychological Review. 91: 112-149.

Arkes, H. R., and Harkness, A. R. 1983. Estimates of contingency between two dichotomous variables. Journal of Experimental Psychology: General. 112: 117-135.

Caldwell, J. A., Caldwell, J. L., Brown, D. L., and Smith, J. K. 2004. The effects of 37 hours of continuous wakefulness on the physiological arousal, cognitive performance, self-reported mood, and simulator flight performance of F-117A pilots. Military Psychology. 16: 163-181.

Chapman, L. J. 1967. Illusory correlation in observational report. Journal of Verbal Learning and Verbal Behavior. 6: 151-155.

Crocker, J. 1981. Judgment of covariation by social perceivers. Psychological Bulletin. 90: 272-292.

Doherty, M. E., Anderson, R. B., Kelley, A. M., and Albert, J. 2006. Probabilistically valid inference of covariation from a single *xy* observation when univariate characteristics are known. Poster presented at the annual meeting of the Society for Judgment and Decision Making, November, at Houston, TX.

Eddy, D. R., and Hursh, S. R. 2001. Fatigue avoidance scheduling tool: Phase 1 SBIR final report. Brooks Air Force Base, TX: Air Force Research Laboratory. AFRL-HE-BR-TR-2001-0140.

Friedland, N., Keinan, G., and Tytiun, T. 1999. The effect of psychological stress and tolerance of ambiguity on stereotypic attributions. Anxiety, Stress, & Coping. 12: 397-410.

Hamilton, D. L., and Gifford, R. K. 1976. Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. Journal of Experimental Social Psychology, 12, 392-407.

Harrison, Y., and Horne, J. A. 2000. The impact of sleep deprivation on decision making: A review. Journal of Experimental Psychology: Applied. 6: 236-249.

Hattori, M., and Oaksford, M. 2007. Adaptive non-interventional heuristics for covariation detection in causal induction: Model comparison and rational analysis. Cognitive Science. 31: 765-814.

Himashree, G., Banerjee, P. K., and Selvamurthy, W. 2002. Sleep and performance--recent trends. Indian Journal of Physiology & Pharmacology. 46: 6-24.

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., and Dement, W. C. 1973. Quantification of sleepiness: a new approach. Psychophysiology. 10: 431-436.

Horne, J. A. 1988. Sleep deprivation and divergent thinking ability. Sleep. 11: 528-536.

Horne, J. A. 2000, February 10. Images of lost sleep. Nature. 403: 605-606.

Kareev, Y. 1995. Positive bias in the perception of covariation. Psychological Review. 102: 490-502.

Killgore, W. D. S., and McBride, S. A. 2006. Odor identification accuracy declines following 24 h of sleep deprivation. Journal of Sleep Research. 15: 111-116.

Lagnado, D. A., and Sloman, S. A. 2004. The advantage of timely intervention. Journal of Experimental Psychology: Learning, Memory, and Cognition. 30: 856-876.

Mandel, D. R., and Lehman, D. R. 1998. Integration of contingency information in judgments of cause, covariation, and probability. Journal of Experimental Psychology: General. 127: 269-285.

McNair, D. M., Lorr, M., and Droppleman, L. F. 1992. Manual: Profile of Mood States--Revised. San Diego, CA: Education and and Industrial Testing Service.

Penetar, D., McCann, U., Thorne, D., Kamimori, G., Galinski, C., Sing, H., Thomas, M., and Belenky, G. 1993. Caffeine reversal of sleep deprivation effects on alertness and mood. Psychopharmacology. 112: 359-365.

Previc, F. H., Lopez, N., Ercoline, W. R., Daluz, C. M., Workman, A. J., Evans, R. H., and Dillon, N. A. 2009. The effects of sleep deprivation on flight performance, instrument scanning, and physiological arousal in pilots. The International Journal of Aviation Psychology. 19: 326-346.

Ramsey, C. S., and McGlohn, S. E. 1997. Zolpidem as a fatigue countermeasure. Aviation Space and Environmental Medicine. 68: 926-931.

Richard, L. S., and Huffman, A. H. 2002. The impact of commuter war on military personnel. Military Medicine. 167: 602-605.

Rogers, A., Holmes, S., and Spencer, M. 2001. The effect of shiftwork on driving to and from work. Journal of Human Ergology (Tokyo). 30: 131-136.

Shanks, D. R. 2004. Judging covariation and causation. In D. J. Koehler, and N. Harvey (Eds.), Blackwell Handbook of Judgment and Decision Making, 220-239. Oxford, England: Blackwell Publishing.

Thorne, D. R., Johnson, D. E., Redmond, D. P., Sing, H. C., and Belenky, G. 2005. The Walter Reed palm-held psychomotor vigilance test. Behavior Research Methods. 37: 111-118.

Tversky, A., and Kahneman, D. 1982. Judgments of and by representativeness. In D. Kahnmen, P. Slovic, and A. Tversky (Eds.), Judgment under Uncertainty: Heuristics and Biases. Cambridge, UK: Cambridge University Press.

Vasterling, J. J., Duke, L. M., Tomlin, H., Lowery, N., and Kaplan, E. 2004. Global-local visual processing in posttraumatic stress disorder. Journal of the International Neuropsychological Society. 10: 709-718.

Vasterling, J. J., Proctor, S. P., Amoroso, P., Kane, R., Gackstetter, G., Ryan, M. A., et al. 2006. The Neurocognition Deployment Health Study: a prospective cohort study of Army Soldiers. Military Medicine. 171: 253-260.

Vasterling, J. J., Proctor, S. P., Amoroso, P., Kane, R., Heeren, T., and White, R. F. 2006. Neuropsychological outcomes of army personnel following deployment to the Iraq war. Journal of the American Medical Association. 296: 519-529.

Wimmer, F., Hoffman, R. F., Bonato, R. A., and Moffitt, A. R. 1992. The effects of sleep deprivation on divergent thinking and attention processes. Journal of Sleep Research. 1: 223-230.

Department of the Army
U.S. Army Aeromedical Research Laboratory
Fort Rucker, Alabama, 36362-0577
www.usaarl.army.mil

U.S. Army Medical Research and Materiel Command