INL/CON-07-12541
PREPRINT

Effects of Extreme Sleep Deprivation on Human Performance

Idaho Academy of Science Symposium and Meeting

Tuan Q. Tran
Kimberly R. Raddatz
Elizabeth T. Cady
Christopher L. Vowels
Bradford J. Amstutz
Peter D. Elgin

April 2007

Effects of Extreme Sleep Deprivation on Human Performance

Tuan Q. Tran
Idaho National Laboratory

Kimberly R. Raddatz
Kansas State University

Elizabeth T. Cady
Kansas State University

Christopher L. Vowels
Kansas State University

Bradford J. Amstutz
Kansas State University

Peter D. Elgin
Kansas State University

Manuscript submitted:

2007 IAS Conference Proceedings submitted to the Journal of Idaho Academy of Science

Keywords: Sleep Deprivation, Human Performance, Aviation

Author Note: This study was completed while the first author was a doctoral student at Kansas State University. Please send correspondence to: tuan.tran@inl.gov

Abstract

Sleep is a fundamental recuperative process for the nervous system. Disruption of this homeostatic drive can lead to severe impairments of the operator's ability to perceive, recognize, and respond to emergencies and/or unanticipated events, putting the operator at risk. Therefore, establishing a comprehensive understanding of how sleep deprivation influences human performance is essential in order to counter fatigue or to develop mitigation strategies. The goal of the present study was to examine the psychological effects of prolonged sleep deprivation (approx. 75 hrs) over a four-day span on a general aviation pilot flying a fixed-based flight simulator. During the study, a series of tasks were employed every four hours in order to examine the pilot's perceptual and higher level cognitive abilities. Overall, results suggest that the majority of cognitive and perceptual degradation occurs between 30-40 hours into the flight. Limitations and future research directions are also discussed.

Effects of Extreme Sleep Deprivation on Human Performance

Sleep allows the body to restore depleted energy, eliminate waste product from muscles, and repair cells (Wade & Tavris, 1993). Disruption of this homeostatic drive can lead to severe consequences especially in aviation. For example, 60% of Class A aircraft mishaps are attributed to sleep disruption each year (Wickens, Lee, Liu, & Becker, 2004). More specifically, the crashes of Korean Air Flight 801 (in 1997) and the China Airlines Flight 006 (in 1985) were both attributed to sleep deprivation (Caldwell & Gilreath, 2002). Unfortunately, sleep deprivation is still highly prevalent in the aviation industry. For example, in a survey of 401 Army aviators and aircrew members, 72% of pilots reported that they had flown under a state of high drowsiness (Caldwell & Gilreath, 2002). More importantly, 45% of the pilots acknowledged falling asleep while flying or sitting in the cockpit.

The goal of the present study was to provide a preliminary examination (and to generate testable hypotheses) of the psychological and emotional effects of prolonged sleep deprivation (approx. 75 hours) over a four-day span on a general aviation pilot flying a fixed-based flight simulator. While a series of behavioral measures were used in the present study, for the purpose of this paper, only five measures will be presented -- risk assessment, creativity, prospective memory, general affect, and sleepiness. Based upon past research showing performance degradations with prolonged lack of sleep, we expect to observed a continuous decline in cognitive abilities (as measured by cognitive tasks) and an increase in negative affect (as measured by an affect questionnaire) with as flight duration increased.

This study will add insight into the effects of extreme sleep deprivation and its potential impact on prolonged and demanding military flights (e.g., sorties) and flights associated with NASA's inter-planetary spaceflight program (e.g., mission to Mars). Understanding the degrading affects of prolonged sleep deprivation on human performance is essential in order to develop fatigue countermeasures and leverage mitigation strategies.

Method

Participant

A male general aviation instructor pilot at Kansas State University-Salina participated in the present study. The pilot had over 700 hours of total flight time as well as the following qualifications: Instrument, Commercial, Multi Engine, CFI (certified flight instructor), CFII (CFI-instrument), and MEI (multi-engine instructor).

Materials

Simulator

The simulator consisted of a 50 inch plasma TV connected to a PC-based Microsoft Flight Simulator 2006 simulating the virtual GlobalFlyer aircraft (by Aeroplane Heaven) and Saitek's X-52 flight controller.

Risk Assessment.

The risk assessment task was designed to assess a pilot's sensitivity to environmental cues, especially those indicating an impending controlled flight into terrain (CFIT). Each trial consisted of a single static terrain-enhanced primary flight display (TE-PFD) that depicted a virtual two-dimensional wire mesh grid of a three-dimensional terrain scene displayed in conjunction with ecological representations of altitude and airspeed. The pilot was asked to judge whether an avoidance maneuver was required based on the airspeed, altitude, and relative heading cues provided in the TE-PFD. The pilot indicated whether a change in direction or altitude was required by selecting the corresponding key on a computer keyboard. This task has been used in previous studies to examine how expert and novice pilots differ in their ability to detect and avoid potential CFIT accidents by comparing and contrasting their ability to find the relevant information and their tolerance for risk, i.e., the extent to which they adopt a safety bias (Uhlarik, Peterson, & Herold, 1998). Both response times and accuracy rates were recorded.

Creativity Tasks.

Convergent creativity relates to the pilot's skill in finding common themes among several different concepts in order to assess which course of action would be best for achieving their goal. The tribond task was used because it was believed to test convergent creativity by requiring the pilot to find an underlying pattern in seemingly unrelated words (similar to Remote Associates Task in the literature; Mednick, 1962). It consisted of a series of three words or phrases, and the pilot's task was to identify a commonality among the tribonds. For example, given the words "cities," "songs," and "noses," the pilot should respond "all have bridges." The list of tribond words was previously rated for difficulty and each session consisted of five tribond sets of varying difficulty.

Divergent creativity, the ability to think "outside the box" in generating several possible courses of action toward completing a goal was tested using a word-naming task (similar to Guilford, 1967) in which the pilot was presented with a letter and was asked to write as many words beginning with that letter in the time allotted (1 min). The letters were chosen so that each had approximately the same number of possible words, as measured by the number of dictionary pages for each letter (Letter S, X, Y, and Z were omitted due to large or small number of possible words). Divergent thinkers will generate more words, and those words will tend to be less semantically, orthographically, or phonetically related than the words generated by less divergent thinkers.

Positive / Negative Affect.

The Positive Affect Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988) lists 10 positive affective states (interested, excited, strong, enthusiastic, proud, alert, inspired, determined, attentive, and active) and 10 negative affective states (distressed, upset, guilty, scared, hostile, irritable, ashamed, nervous, jittery, and afraid). The pilot indicated the degree to which he felt each of the 20 affective states on a 5-point scale with 1= "Very Slightly or Not at all" and 5 = "Extremely."

Stanford Sleepiness Scale.

The Stanford Sleepiness Scale (Hoddes, Dement, & Zarcone, 1972b) is a single item measure in a seven-point Likert-type format ranging from "very sleepy" to "very alert" and was used as a qualitative assessment of the effects of sleep deprivation. Scores also have been shown to be empirically related to degradation in mental performance such that lower scores correlate with high performance degradation (Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973).

Prospective Memory

Pilots must remember to perform many different functions at various times throughout flight. The test of prospective memory in this study was based on the pilot's ability to report the status of 7 different parameters at 20 minutes into every hour with no prompt or cue to give the report. An error of omission was assigned if the pilot failed to report within 15 minutes of the required time.

Design and Procedure

The pilot completed 11 testing sessions, including a baseline session that occurred the day before the flight simulation started, nine sessions that occurred during the approximately 75 hour flight, and a final session that occurred at the conclusion of the simulation. Testing session blocks lasted for four hours and were followed by a four-hour non-testing sessions. The pilot spent cumulatively two and one-half hours of the four hour testing block completing tasks, with 1-2 tasks given every half hour. The order of the tasks was counterbalanced across the sessions and the tasks themselves were randomly distributed across the duration of the four hour testing block (See Table 1).

Table 1. *Experimental Testing Schedule*

Pre-Test	PANAS (PANAS) Sleepiness Scale (SSS) CFIT Convergent Creativity (CC) Psychotic Symptoms (PS) Divergent Creativity (DC) Word Problems (WP)					
Day 1	11 am – 3pm Test 2	3 pm – 7 pm No Test	7 pm – 11 pm Test 3	11 pm – 3 am No Test	3 am – 7 am Test 4	7 am – 11 am No Test
	CFIT DC/WP PANAS/SSS CC		WP/DC CC PANAS/SSS CFIT		DC/WP CFIT CC PANAS/SSS	
Day 2	11 am – 3pm Test 5	3 pm – 7 pm No Test	7 pm – 11 pm Test 6	11 pm – 3 am No Test	3 am – 7 am Test 7	7 am – 11 am No Test
	CC PANAS/SSS WP/DC CFIT		WP/DC CFIT PANAS/SSS CC		PANAS/SSS DC/WP CC CFIT	
Day 3	11 am – 3pm Test 8	3 pm – 7 pm No Test	7 pm – 11 pm Test 9	11 pm – 3 am No Test	3 am – 7 am Test 10	7 am – 11 am No Test
	CC DC/WP PANAS/SSS CFIT		CFIT PANAS/SSS CC WP/DC		CC CFIT WP PANAS/SSS	
Post-Test	PANAS/SSS CFIT CC WP/DC					

The length of the flight was approximately 75 hours and the pilot stayed in the "cockpit" for the duration of the flight with the exception of bathroom breaks. An experimenter was continuously present and the simulator was placed near the door of the student center on campus, so students, faculty, administration, and other individuals were walking by or observing his progress. During each testing session, the experimenter administered the tasks and the pilot recorded his answers on either a computer or on paper. When not administering tests, the experimenters strove to limit interactions with the pilot in order to maintain the solitary aspect of the flight. Based on the pilot's preference, he drank water and weight-loss shakes but did not eat solid food for the duration of the flight.

Results

It was expected that a continual degradation of the pilot's performance would be observed on various cognitive tasks as well as a gradual increase in negative affect for the duration of the study. Unfortunately, scarce data points prevented the use of inferential statistics. Nevertheless the aim of the study was to provide preliminary observation of meaningful patterns in the effects of prolonged sleep deprivation and to explore possible hypotheses that can be examined statistically in a larger sample in the near future.

Risk Assessment

The results of the risk assessment revealed some interesting findings (see Figure 1). First, instead of a linear degradation of performance as a function of time, performance fluctuated across the experimental session, with a drop in both accuracy and response time in sessions 6 and 10, approximately 36 and 68 hrs into the study, respectively. Second, accuracy was fairly high (above 90%) throughout the experiment but came at the expense of increased response time (i.e., a speed-accuracy trade-off). This is consistent with the literature suggesting that sleep deprivation induces a speed-accuracy trade-off (Angus & Heslegrave, 1985; Webb & Levy, 1982) where accuracy trumps speed. However the speed-accuracy tradeoff was not constant across sessions. Reasons for this inconsistent speed-accuracy tradeoff will be elaborated on later in the discussion section.

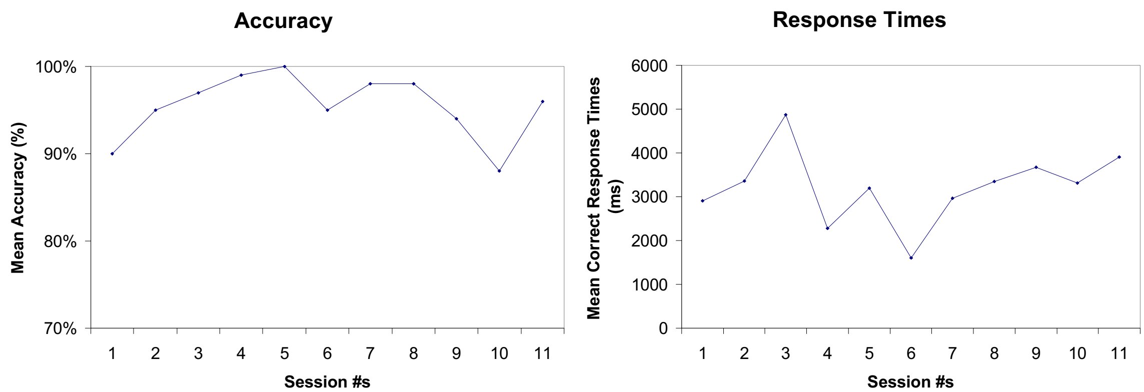

Figure 1. Risk Assessment Results

Creativity Task

Similar to the risk assessment results, both divergent and convergent tasks results fluctuated. Severe performance degradation for both tasks occurred in session 3 and 6 (approx. 12 and 36 hours into the study, respectively). Results also revealed different performance trends for the divergent and convergent tasks (See Table 2). Specifically, in session 7 (approx 44 hours into the study), the pilot scored high on the convergent task but poorly on the divergent task, whereas, in session 10 (approx 68 hours into the study), the pilot performed well in the divergent task but poorly on the convergent task. This different performance trends between divergent and convergent task may be of interest to examine in future studies.

Prospective Memory

Results again revealed some fluctuating performance degradation regarding the pilot's prospective memory. The pilot successfully remembered to report the flight parameters at exactly 20 minutes past the hour 40% of the time, but he was late in reporting the flight parameters 29% of the time. Most troubling was that the pilot failed to report the status of his equipment 29% of the time (i.e., error of omission), with six consecutive errors of omission between 31 and 36 hours into the study.

Table 2. *Creativity and Self-Ratings Results*

Sessions	**Sleep Deprivations (# hours into the study)**	**Divergent (# words)**	**Convergent (% correct)**	**Sleep Scale**	**PANAS Positive Affect Scores**	**PANAS Negative Affect Scores**
1	Pre-test	18	20%	2	4.7	1.2
2	4	19	60%	2	4.1	1.2
3	12	11	40%	4	2.9	1.3
4	20	21	80%	5	1.9	2.1
5	28	21	80%	2	2.9	1.5
6	36	22	40%	3	1.7	2.3
7	44	14	80%	5	1	1.7
8	52	31	100%	3	3.3	1.2
9	60	28	100%	1	3.7	1.1
10	68	31	40%	3	2.9	2.4
11	Post-Test	24	100%	2	3.5	1.2

Sleep Ratings/PANAS

The SSS ratings revealed that the pilot was sleepiest in sessions 3, 4, 6, and 7 (approx. 12 to 44 hrs into the study). Also in those sessions, the PANAS ratings revealed the pilot had a decrease in positive affect and a slight increase in negative affect. Thus, when the pilot reported a high level of sleepiness and low affect ratings, performance on both the prospective memory and creativity task suffered. This suggests a relationship that may be worthwhile to examine in future studies.

Discussion/Conclusion

The aim of the current study was to provide preliminary observations of the effects of prolonged sleep deprivation to leverage hypotheses for future studies. Consistent with the literature, the present study found that sleep deprivation degraded performance. This study found large performance degradation around 12 hours into the study as evidenced by: 1) the largest spike in risk assessment response time (while still maintaining a high accuracy rates), 2) a decrease in both convergent and divergent performances, and 3) the first observation of an error of omission in the prospective memory task. The sleep and affect scales indicated that the pilot was experiencing a high degree of sleepiness and low levels of positive and negative affect in this session. Post-study interview with the pilot revealed that he was able to obtain only 3-4 hours of sleep the night prior to the study due to enthusiasm. Thus, these results suggest that minimal sleep can hinder performance within a 12-hour period after waking, a finding relevant in many industries where workers tend to get less than 5 hours of sleep a night. Performance continued to degrade approximately 45 hours into the study as response times increased and risk assessment accuracy suffered. Also, drops in both the convergent and divergent task performance and a string of errors of omissions in the prospective memory task were observed. The sleep and affect scales again revealed that, at these times, the pilot was experiencing a high degree of sleepiness and low level of positive and negative affect. These relationships may be worthwhile to examine in future studies.

A consistent finding observed in all behavioral measures was the inconsistent performance patterns. This finding contrasts with our expectation that linear performance degradation would emerge as the length of sleep deprivation increased (such results should be statistically tested in future study with a larger sample size). Tasks were counterbalanced across each of the testing sessions. Furthermore, within each task, item/trial difficulty was controlled across time. Thus, the results cannot be attributed to any experimental order effects. The study constraints permitted the pilot to freely use the restroom and to not be disturbed during his incidental naps that tended to be 5-20 minutes in length. It is possible that these incidental naps at least partially explain our study results. Performance degradation may have occurred prior to these naps while performance enhancement may have occurred after these incidental naps. Steve Fossett, owner of several record-breaking flight endurance adventures including the Global Flyer flight, says his ability to take 5 minute naps every hour is a primary reason for his record-breaking success. Thus, it seems that very short naps may be an effective sleep mitigation strategy. Another possible explanation is personal contact. Due to constraints beyond the control of the experimenters, the flight simulator was located near the door of the student center on campus, so students, faculty, administration, and other individuals who were walking by could observe the pilot's behaviors. Furthermore, an experimenter was constantly present for the duration of the flight, either to administer the experimental task or to reduce bystander disruption on the pilot's performance in the simulator. Thus, these personal contacts between the pilot and his environment (e.g., bystanders, experimenters, or ambient noises) may have provided sufficient stimulation that enabled the pilot to maintain a high level of performance on the experimental tasks. Periods of low performance were found to correlate with periods where one would expect low levels of personal contact to occur (e.g., late evening).

In conclusion, we did not observe a linear performance degradation of prolonged sleep deprivation. Instead, we found performance fluctuation across the experimental session. This behavioral pattern may be attributed to either naps or the amount of personal contact (an assumption that was unable to be examined in the current study due to design/methodology limitation). Thus, this study provides some groundwork for using naps and/or personal contact as mitigation strategies to counter the effect of fatigue. Future studies are needed to empirically examine not only the effects of prolonged sleep deprivation under more socially deprived situations but also the use of short concentrating napping as a mitigation strategy for fatigue.

References

Angus, R. G. and Heslegrave, R. J. (1985). Effects of sleep and sustained cognitive performance during a command and control simulation. *Behav. Res. Methods, Instruments Computers*, 17, 55–67.

Caldwell, J. A., and Gilreath, S. R. (2002). A survey of aircrew fatigue in a sample of U.S. Army aviation personnel. *Aviation, Space, and Environmental Medicine, 73*(5), 472-479.

Graeber, R.C., Dement, W.C., Nicholson, A.N., Sasaki, M., & Wegman, H.M. (1986). International cooperative study of aircrew layover sleep: Operational summary. *Aviation, Space, and Environmental Medicine, 57,* B10-B13.

Guilford, J.P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.

Hoddes, E., Dement, W.C., & Zarcone, V. (1972b). Cross-validation of the Stanford sleepiness scale (SSS). *Sleep Research, 1,* 91.

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W.C. (1973). Quantification of sleepiness: A new approach. *Psychophysiology, 10,* 431-436.

Knapp, C. J., and Johnson, R. (1996). F-16 Class A Mishaps in the U.S. Air Force. *Aviation,Space, and Environmental Medicine, 67*(8), 777-783.

Mednick, S. (1962). The associative basis of creative problem solving process. *Psychological Review*, 69, 200-232.

Sternberg, R. J. (1999). Cognitive psychology (2^{nd} ed.). Fort Worth, TX: Harcourt Brace College Publishers.

Uhlarik, J., Peterson, D., & Herold, J.M. (1998). Terrain awareness using a terrain enhanced primary flight display. *Proceedings of the Human Factors and Ergonomics Society, 42^{nd} Annual meeting,* Chicago, IL (pp. 102-106).

Wade, C., & Tavris, C. (1993). *Psychology*. New York: HarperCollins College Publishers.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54,* 1063-1070.

Webb, W. and Levy, C. (1982). Age, sleep deprivation and performance. *Psychophysiology*, 19 (3): 272–276.

Wickens, C.D., Lee, J.D., Liu, Y., & Gordon Becker, S.E. (2004). *An Introduction to Human Factors Engineering 3rd Edition*. New Jersey: Pearson Education, Inc.