(2)

NAVAL AEROSPACE MEDICAL RESEARCH LABORATORY
NAVAL AIR STATION, PENSACOLA, FL 32508-5700

SIMULATED SUSTAINED FLIGHT OPERATIONS AND PERFORMANCE, PART 1: EFFECTS OF FATIGUE

D.F.Neri, S.A. Shappell, and C.A. DeJohn

Approved for public release; distribution unlimited.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)

2. REPORT DATE
July 1992

3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE
Simulated Sustained Flight Operations and Performance, Part 1: Effects of Fatigue

5. FUNDING NUMBERS

6. AUTHOR(S)
D.A. Neri, S.A. Shappell, and C.A. DeJohn

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Aerospace Medical Research Laboratory
Bldg. 1953, Naval Air Station
Pensacola, FL 32508-5700

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Naval Medical Research and Development Command
National Naval Medical Center
Building 1, Tower 12
8901 Wisconsin Avenue
Bethesda, MD 20889-5044

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

Published in *Military Psychology*, Vol. 4, No. 3, pp. 137-155, July 1992.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

Sustained flight operations are likely to produce fatigue and performance decrement in aviators. We assessed changes in cognitive performance using a work/rest schedule modeled on successive long-range attack missions. Twelve subjects performed several subtests of the Unified Tri-Service Cognitive Performance Assessment Battery and the Walter Reed Performance Assessment Battery 18 times during a simulated sustained operation. The scenario consisted of a 9-hr planning session followed by a 4-hr rest period and a 14-hr daytime mission. After 6 hr of rest, subjects repeated this schedule with a nighttime mission. For two spatial tests, subjects showed linear increases in response rate and one of its components, error rate. Subjects appeared to change strategy as the study progressed, possibly exchanging a higher failure rate for a savings in time. Any tendency to take chances when fatigued may have serious implications for aircrew in sustained operations.

14. SUBJECT TERMS

Sleep-loss, Sustained operations, Continuous operations

15. NUMBER OF PAGES
19

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std Z39-18
298-102

MILITARY PSYCHOLOGY, 4(3), 137–155

Simulated Sustained Flight Operations and Performance, Part 1: Effects of Fatigue

David F. Neri, Scott A. Shappell, and Charles A. DeJohn
Naval Aerospace Medical Research Laboratory
Pensacola, Florida

Sustained flight operations are likely to produce fatigue and performance decrement in aviators. We assessed changes in cognitive performance using a work/rest schedule modeled on successive long-range attack missions. Twelve subjects performed several subtests of the Unified Tri-Service Cognitive Performance Assessment Battery and the Walter Reed Performance Assessment Battery 18 times during a simulated sustained operation. The scenario consisted of a 9-hr planning session followed by a 4-hr rest period and a 14-hr daytime mission. After 6 hr of rest, subjects repeated this schedule with a nighttime mission. For two spatial tests, subjects showed linear increases in response rate and one of its components, error rate. Subjects appeared to change strategy as the study progressed, possibly exchanging a higher failure rate for a savings in time. Any tendency to take chances when fatigued may have serious implications for aircrew in sustained operations.

Sustained operations (SUSOPs) involve demanding, long work schedules that exceed a normal duty cycle and usually result in fatigue and sleep deprivation (Neri & Gadolin, 1990). They are often associated with training, ground-combat missions, or air-combat operations. In this article, aircrew SUSOPs refer to multiple, long-range, carrier-based, air-combat missions. These differ from other SUSOPs in their cyclical nature. For example, high-intensity aircrew SUSOPs can involve several demanding attack

Requests for reprints should be sent to David F. Neri, Aviation Performance Division, Naval Aerospace Medical Research Laboratory, Naval Air Station, Pensacola, FL 32508-5700.

missions in excess of 10 hr separated by only a few hours for rest/sleep and meals. This pattern results in a sliding work schedule during which aircrew must perform at peak levels at various times of day and night. Such SUSOPs include the early morning hours when individual circadian rhythms are in a trough (Nicholson & Stone, 1987). The SUSOPs may last several days before a significant break occurs, making it difficult to obtain adequate rest. Furthermore, the effects of cyclical operations and the sliding schedule will be magnified for flight leaders and mission commanders who have heavy tasking long before the first aircraft is launched. The planning demands placed on these men serve to "front load" them with fatigue. As a result, they could be struggling against stress and heavy fatigue well before the actual start of the mission.

Planning and performance demands combined with stress and fatigue probably result in fragmented sleep and, if continued, circadian rhythm desynchronization. These disruptions increase the probability of performance degradation. Exactly how performance degrades during SUSOPs is difficult to predict. The relation between stress, fatigue, and cognitive performance is complex (Holding, 1983). Skill level, environmental conditions, and task variables—such as type, complexity, duration, pacing (self-paced vs. work paced), and location within duty periods—all play a role in this relation (Heslegrave & Angus, 1985; Hockey, 1986; Johnson & Naitoh, 1974; Naitoh, 1981). Task repetitiveness and the extent to which it requires sustained effort, attention, activity, and stamina are important (Krueger, 1991). Individual characteristics of age, temperament, personality, and intelligence also can affect performance (Hockey, 1986).

When subjects are totally deprived of sleep, there are clear decrements in cognitive performance (Mullaney, Kripke, Fleck, & Johnson, 1983; Mullaney, Kripke, Fleck, & Okudaira, 1983) that increase with the continuousness of the tasks (Angus & Heslegrave, 1985). Sleep loss can even cause a phase delay of several hours in the natural circadian rhythm of performance (Babkoff, Mikulincer, Caspy, Carasso, & Sing, 1989). In SUSOPs and continuous operations, workload, physical activity, sleep deprivation, and time of day can interact to affect performance (Babkoff et al., 1985; Englund, Ryman, Naitoh, & Hodgdon, 1985; Krueger, 1989). Lack of sleep is generally considered to cause a monotonic decrease in performance and diurnal factors, leading to changes in the performance rhythm (Babkoff, Caspy, Mikulincer, & Sing, 1991). Although all aspects of performance might not degrade in the same way or at the same rate with fatigue, higher level cognitive abilities (e.g., short-term memory and decision making) are likely to be affected. Bartlett's (1943) Cambridge cockpit studies demonstrated that with prolonged work attention narrows and the variability of response latencies often increases. This effect may be a reflection of increased lapses or blocks (unusually long response times). The likelihood of errors, particularly errors of omission, also increases (Hockey, 1986). This finding is especially relevant to long-range attack missions, when target detection is critical and one error of omission can be catastrophic.

An interesting result of prolonged work and its accompanying fatigue can be the selection of easy but risky alternatives. The Choice of Probability and Effort model (COPE; Holding, 1983) was developed from studies in which subjects controlled the effort they apply to the solution of a problem. Probability of success corresponded to the level of effort. After a day of performance testing, subjects were likely to choose a strategy involving less effort even when they knew it had less probability of success. This strategy has important ramifications for aircrew SUSOPs in which low success rates are unacceptable.

These facts are not always adequately considered in planning and executing missions, even when flexibility exists. For example, aircrew are frequently asked to work on continually sliding work/rest schedules, even though shift rotations on successive days (particularly in the counterclockwise direction) (a) can dissociate the naturally occurring circadian rhythm (Coleman, 1986), (b) are less preferred (Czeisler, Moore-Ede, & Coleman, 1982), and (c) negatively affect performance (Nicholson & Stone, 1987). Doctrine for managing the problems that can be associated with aircrew SUSOPs must be developed and implemented. First, the type and magnitude of fatigue-induced performance degradation in realistic SUSOP scenarios must be categorized and quantified. This can be done in the laboratory by reproducing a cyclical aircrew SUSOP schedule and measuring changes in cognitive performance. The goal of this study is to provide a first step in documenting the nature and severity of problems associated with a specific aircrew SUSOP before pursuing appropriate countermeasures.

METHOD

Subjects

Twelve male U.S. Marines, ranging in age from 23 to 28 years, volunteered for the experiment. All were college graduates awaiting initial flight training. The subjects had current flight physicals and underwent medical screening by a flight surgeon. Heavy nicotine or caffeine users and those on medication were excluded.

Apparatus

All tests were presented on six microcomputers equipped with color monitors. Each work station was outfitted with a Mini-Modulus III™ input device configured with numeric keypad, resistive joystick, and tapping key. The Mini-Modulus III™ was interfaced to the computer with a Systems Research Laboratories Labpak™ multifunction data-acquisition board with a 1-MHz clock. This board provided timing resolution of 1 ms. The six work stations were linked by a local area network, ensuring simultaneous presentation of performance tests.

Tests

The generic performance assessment battery (G-PAB) employed here was composed of tests from the Unified Tri-Service Cognitive Performance Assessment Battery (UTC-PAB; Englund et al., 1987; Perez, Masline, Ramsey, & Urban, 1987) and the Walter Reed Performance Assessment Battery (WR-PAB; Thorne, Genser, Sing, & Hegge, 1985). The tests were chosen to sample several stages of information processing: (a) input, detection, and identification (Wilkinson's Four-Choice Reaction Time Test); (b) linguistic/symbolic manipulation (Grammatical Reasoning and Serial Add/Subtract tests); and (c) spatial-processing manipulation (Manikin, Pattern Recognition I, and Time Estimation tests). These stages are arguably very important for aircrew during piloting, navigation, and target acquisition. The performance tests, with their parent performance battery in parentheses, are described in detail in the following sections.

Four-choice serial reaction time test (UTC-PAB). This test is a modification of the Four-Choice Reaction Time Test developed by Wilkinson and Houghton (1975) and adapted for the personal computer by Ryman, Naitoh, and Englund (1984). It evaluates information-processing resources dedicated to stimulus encoding, categorization, and response selection. The subject was presented with a plus sign in one of four quadrants on the CRT. His task was to indicate which quadrant it occupied by pressing one of four keys on the Mini-Modulus III™ keypad. The plus sign remained visible until a response was made, then reappeared randomly in one of the quadrants. Each test administration consisted of 50 trials.

Grammatical reasoning test (UTC-PAB and WR-PAB). This test is an adaptation of one designed by Baddeley (1968) to test logical-reasoning ability. Subjects were presented a pair of letters (AB or BA) with a statement describing their sequential arrangement. They were instructed to determine whether the statement accurately described the letter pair by pressing one of two keys on the keypad corresponding to true and false responses. Thirty-two trials were presented during each test administration.

Stimulus items were presented one at a time and were centered on the screen.

Serial add/subtract test (WR-PAB). This test is a version of the Serial Add/Subtract Test used by Wever (1981) and requires rapid arithmetic manipulations. It places demands on working memory and sustained attention. The subject was presented two digits in succession followed by either a plus or minus sign. His task was to perform the indicated operation and enter the last digit of the solution on the keypad. If a negative solution was obtained, the subject was instructed to add 10 and enter the last digit of the new solution. Fifty trials were presented for each test administration.

Manikin test (UTC-PAB). This test is a modification of one developed by Benson and Gedye (1963) to assess the ability to perform rotations and related transformations of a mental image. During each trial, a human figure (the manikin) was displayed inside either a green circle or red square. The manikin held a red square in one hand and a green circle in the other. The manikin was randomly presented in one of four orientations: (a) upright and facing the subject, (b) upright and facing away, (c) upside down and facing the subject, or (d) upside down and facing away. The task was to indicate which hand held the same symbol as that surrounding the manikin by pressing a key on the keypad. Sixty-four trials were presented during each test administration.

Pattern recognition I test (WR-PAB). This test places demands on spatial memory. A random pattern of asterisks was displayed for 1.5 sec, followed by a blank screen for 3.5 sec, followed by a second pattern of asterisks. The task was to indicate with a key press whether the two asterisk patterns were the same or different. Each administration included 20 trials. For one half of the trials, three randomly selected asterisks exchanged horizontal position while retaining vertical position.

Time estimation test (UTC-PAB and WR-PAB). This test is a variation of the Time Wall Test developed by Jerison and Argintenu (1958). It examines the ability to estimate when a target, moving at a constant rate, has traveled a predetermined distance. During each trial, a small brick emerges from the top of the display and descends at a constant velocity toward a solid color barrier occupying the lower third of the screen. When the brick reaches the top of the barrier, it is no longer visible. The task was to estimate the moment at which the brick reached the bottom of the barrier by pressing the tapper key. Actual time was 10 sec. Six trials were presented for each test administration.

Procedure

Training. Subjects trained on the cognitive tests over 6 weekdays beginning on a Monday. The final training session was conducted on a Monday to counteract any training loss over the weekend. On the first day of training, subjects had a single session in the morning. Subsequent days consisted of both morning and afternoon sessions, each lasting about 20 min. The order of the six tests remained the same throughout training and testing. Before the experiment, subjects completed 11 administrations of the test battery and received feedback. The number of test administrations is considered enough to ensure asymptotic performance (Englund et al., 1987).

Experimental design. The experimental design was based on a realistic SUSOP schedule obtained in 1988 from the Commander, Medium Attack Tactical Electronic Warfare Wing of the U.S. Pacific Fleet. One cycle of this scenario consisted of 9 hr of preflight planning, 4 hr of rest, a 14-hr mission, and 6 hr of rest. The planning session, although not applicable to all aircrew in the Navy, is particularly relevant for flight leaders and mission commanders. The 4-hr and 6-hr rest periods represented best-case situations because key planners may not have the luxury of significant rest between mission planning, execution, and subsequent planning sessions. The 14-hr mission included two 2-hr blocks at the beginning and end for brief/preflight and postflight/debrief. To simulate cyclical operations, two iterations of the preceding schedule were incorporated in the experiment, separated by the 6-hr rest period. The total 60-hr schedule is diagrammed in Figure 1.

Subjects were tested in two groups of six. They were brought into the laboratory the evening before the experiment to allow them to become accustomed to the sleeping accommodations and to allow us to control bedtime. After the final training session the following morning, the experiment began at 1800 and proceeded continuously until 0500 2½ days later. The first simulated planning session occurred from 1800 the first day to 0300 the following morning. During this period, subjects were seated in front of computer work stations separated by partitions and performed cognitive tests almost continuously. They were free to interact with other subjects and to leave their work stations during brief breaks. During the 4-hr (and 6-hr) rest period, subjects were allowed to lay on their beds to rest or sleep. During the simulated 14-hr missions, subjects were restricted to their seats with feet on the floor to maintain a greater fidelity to the actual cockpit. Subjects performed slightly different batteries of cognitive tests

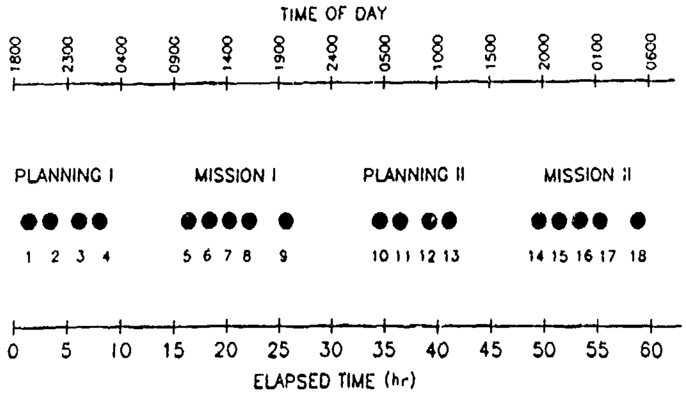

FIGURE 1 The schedule used in this experiment. Time of day is shown along the top axis, and time from the beginning of the experiment is shown along the bottom axis. The dots represent the 18 administrations of the G-PAB: four in each of two simulated planning and five in each of two simulated mission segments. The spaces between the planning and mission segments represent the rest periods. The spaces between Times 2 and 3 and 11 and 12 are for meals. The spaces between Times 8 and 9 and 17 and 18 indicate administration of a computerized flight simulator.

during the planning and mission segments. The G-PAB was administered during both planning and mission periods.

The test room was physically isolated, and the windows were covered to eliminate distractions. Subjects were denied access to clocks or watches to prevent their use in the Time Estimation Test. All subjects were on a balanced, controlled diet of about 3,100 calories daily. Nicotine, caffeine, and all medications were prohibited to prevent their uncontrolled impact on the cognitive measures.

The G-PAB was but one of several visual, auditory, and cognitive tests (including a computer-administered flight simulator) presented throughout planning and mission segments. The tests were selected to measure cognitive processes most representative of planning and flying. Tests were separated by short breaks. The G-PAB was administered four times during planning segments and five times during mission segments. Measures collected from the flight simulator (not included in this analysis) accounted for larger gaps in time between the 8th and 9th administrations and the 17th and 18th administrations of the G-PAB (Figure 1). A 20-min block of time was allotted for each G-PAB session. Subjects were given explicit instructions to proceed as quickly as possible while maintaining accuracy. These instructions were emphasized repeatedly during the experiment. Response accuracy and duration were collected by computer for later analysis.

RESULTS

Multiple measures were collected on all cognitive tests except for the Time Estimation Test. The basic measures of reaction time (RT) for correct responses, accuracy (percentage correct), and response rate were collected by computer. Two components of response rate—hit and error rates—were derived from the accuracy and response rate measures. *Hit rate* is the number of correct responses per minute (also called *throughput*), and *error rate* is the number of errors per minute. The five measures were computed for the Pattern Recognition, Manikin, Grammatical Reasoning, Serial Add/Subtract, and Four-Choice Serial Reaction Time tests. For the Time Estimation Test, the only relevant measure was the subject's estimate of target arrival time.

Because one of the main objectives of this study was to examine cumulative effects of fatigue over the course of a realistic SUSOP schedule, analyses focused on an examination of performance across the entire 60-hr experiment. A one-way repeated measures analysis of variance (ANOVA) was first performed on each response measure to ascertain the presence of statistical significance in the data set. We used a liberal significance level of .10 because our main concern in this initial study was to increase power and reduce Type II error. In this type of military scenario, the importance of detecting the presence of performance changes as a result of the work/rest schedule is considered greater than the risk of falsely rejecting the null hypothesis of no change. Thus avoiding Type II error was considered more important than committing Type I error (Keppel, 1973, pp. 153–154). Most important, the experimental paradigm was repeated with a very similar pattern of results for the spatial-processing and linguistic/symbolic tests (Shappell, Neri, & DeJohn, in press). For those measures showing a significant omnibus *F*, trend analyses were performed on each response measure to examine the change in performance over time. Only significant linear and quadratic trends are reported for two reasons. First, most psychological theories limit themselves to predictions involving linear or quadratic trends (Keppel, 1973, p. 116). Second, higher order trends are appropriate for an analysis of time-of-day effects rather than detection of a monotonic degradation in performance presumably due to fatigue. Results for the spatial processing tests; linguistic/symbolic manipulation tests; and the input, detection, identification test are described in the following sections.

Spatial-Processing Tests

The two spatial-processing tests with multiple measures showed a consistent pattern of results. The Pattern Recognition and Manikin tests both showed significant linear decreases in RT for correct responses and increases in response rate as the experiment progressed, according to the trend analyses (Table 1). All response measures for the spatial-processing tests, except RT, are also plotted in Figures 2 and 3. For the Pattern Recognition Test, increase in response rate was due to a linear increase in error rate. The increase in error rate, coupled with no significant change in hit rate, resulted in a linear decline in accuracy. For the Manikin Test, increase in response rate was composed of linear increases in both hit and error rates. Neither hit rate nor error rate significantly outweighed the other, resulting in no change

TABLE 1
Trend Analysis Results for the Cognitive Tests

Task	*Trend*	*F*	*df*[a]
Spatial Manipulation tests			
Pattern Recognition			
RT	Linear	9.34**	1,11
Percentage correct	Linear	5.44**	1,11
Response rate	Linear	9.73**	1,11
Hit rate	ns		
Error rate	Linear	9.57**	1,11
Manikin			
RT	Linear	4.86*	1,10
Percentage correct	ns		
Response rate	Linear	6.38**	1,10
Hit rate	Linear	4.19*	1,10
Error rate	Linear	3.28*	1,10
Time Estimation	ns		
Linguistic/Symbolic Manipulation tests			
Grammatical Reasoning			
RT	Linear	5.14**	1,10
Percentage correct	ns		
Response rate	Linear	4.69*	1,10
Hit rate	Linear	4.44*	1,10
Error rate	ns		
Serial Add/Subtract			
RT	Quadratic	10.14**	1,9
Percentage correct	ns		
Response rate	Quadratic	6.45**	1,9
Hit rate	Quadratic	6.28**	1,9
Error rate	ns		
Input, Detection, and Identification test			
Four-Choice Serial Reaction Time			
RT	ns		
Percentage correct	ns		
Response rate	ns		
Hit rate	ns		
Error rate	ns		

[a]Different degrees of freedom are due to several instances of missing data.
*$p \leq .10$. **$p \leq .05$.

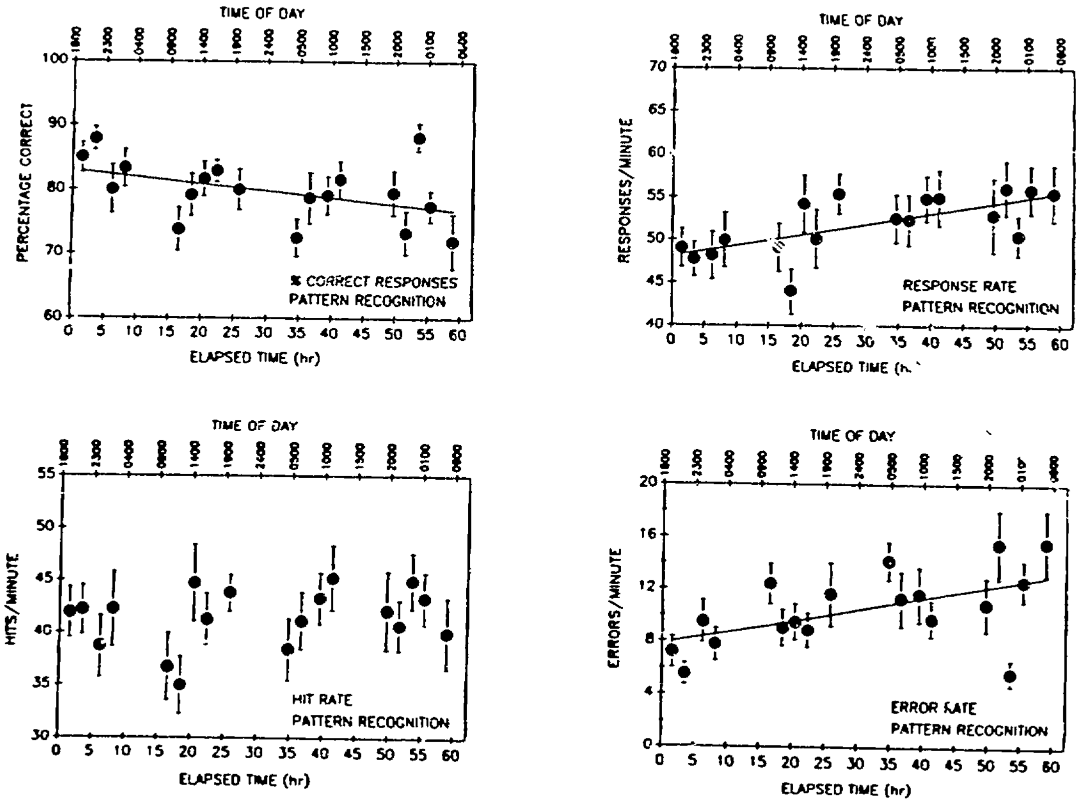

FIGURE 2 Representative data for the Pattern Recognition Test. Percentage of correct responses, response rate, hit rate, and error rate are all shown as a function of time elapsed from the beginning of the experiment. Regression lines are fit based on trend analysis results. Error bars represent ± 1 SEM.

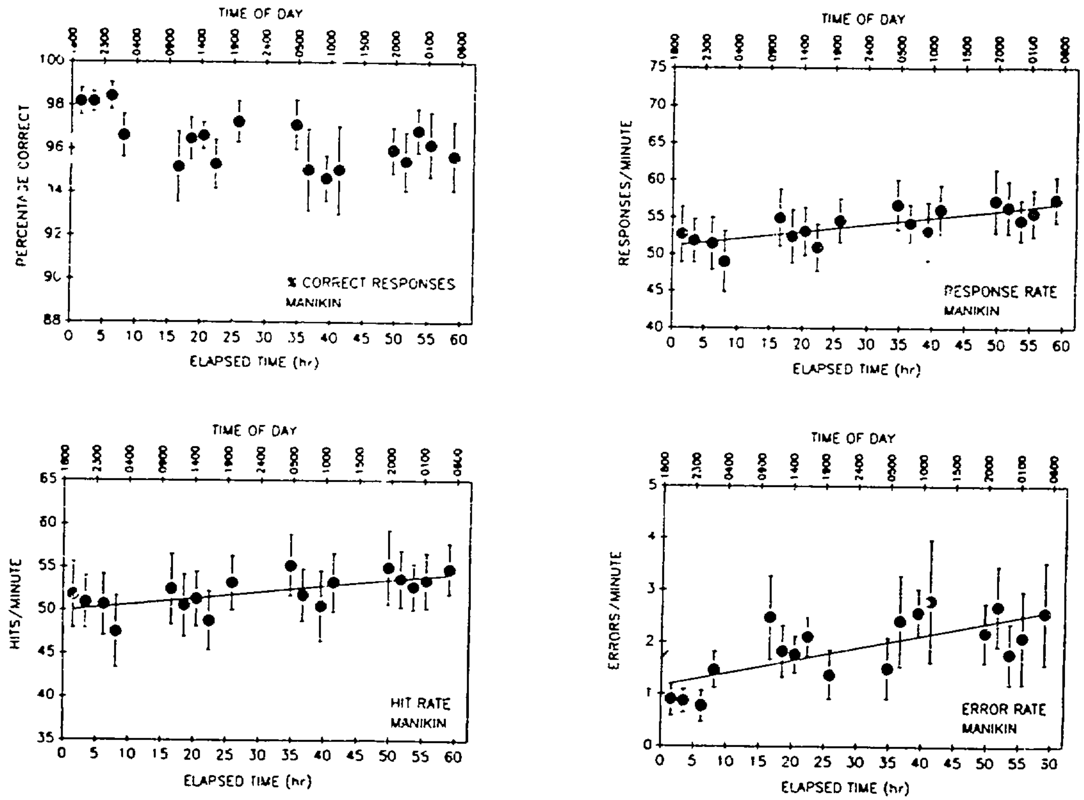

FIGURE 3 Representative data for the Manikin Test. Percentage of correct responses, response rate, hit rate, and error rate are all shown as a function of time elapsed from the beginning of the experiment. Regression lines are fit based on trend-analysis results. Error bars represent ± 1 SEM.

in accuracy. The Time Estimation Test showed no significant change in estimated arrival time of the brick (Table 1).

Linguistic/Symbolic Manipulation Tests

The Grammatical Reasoning Test showed a linear decrease in RT and a linear increase in response rate (Table 1). Unlike the results of the spatial-processing tests, the increase in response rate consisted entirely of a linearly increasing hit rate. Error rate did not change significantly with time. However, increasing hit rate was not enough to significantly increase response accuracy over time.

The Serial Add/Subtract Test showed quadratic changes in RT and response rate as the experiment progressed (Table 1). Again, unlike the spatial tests, change in response rate was solely due to a quadratically changing hit rate. Both measures followed a pattern of higher values early and late in the study with a dip in rate in between. Error rate remained essentially unchanged.

Input, Detection, Identification Test

The Four-Choice Serial Reaction Time Test did not show significant changes on any of the five response measures. Consequently, we did not conduct trend analyses. Possible reasons for this finding are discussed later.

Planned Comparisons

Before the experiment we predicted that, because subjects were able to obtain 14 hr of rest in the first 46 hr of the experiment, significant fatigue-related performance decrements would not occur until the second simulated mission (Hour 46 and beyond). A possible fatigue-related effect, the linear increase in error rate just described, occurred for only two of the three spatial processing tests. We performed planned comparisons on the Pattern Recognition and Manikin tests to detect any significant changes in performance between the beginning of the experiment and the five G-PAB administrations during the second mission. Specifically, means of the five response measures for the first administration of the tests were separately compared to the means for each of the five administrations occurring during the second mission. This resulted in five comparisons per measure: Time 1 versus Times 14, 15, 16, 17, and 18 (see Figure 1). Although these contrasts were not orthogonal, they were appropriate because they provided important information unobtainable from other orthogonal planned comparisons (Keppel, 1973, pp. 92-93). The results are shown in Table 2. A significance level of .10 was used for the same reasons cited previously. All

TABLE 2
Planned Comparisons Between First Administration of Spatial Tests and the Five Administrations During the Second Mission

	Comparison Times (p)				
Task	*1905–1915 1–14*	*1905–2110 1–15*	*1905–2305 1–16*	*1905–0100 1–17*	*1905–0430 1–18*
Pattern Recognition					
RT	ns	**	ns	**	*
Percentage correct	ns	**	ns	*	**
Response rate	ns	**	ns	*	*
Hit rate	ns	ns	ns	ns	ns
Error rate	ns	**	ns	**	**
Manikin					
RT	ns	ns	ns	ns	*
Percentage correct	ns	*	ns	ns	*
Response rate	ns	ns	ns	ns	*
Hit rate	ns	ns	ns	ns	ns
Error rate	*	*	ns	ns	*

*$p \leq .10$. **$p \leq .05$.

significant comparisons were consistent with the trend analysis results (i.e., faster RTs, less accuracy, greater response rate, and greater error rate by the end of the experiment). We observed one significant difference between Time 1 and Time 14 at $p \leq .10$. By Time 15, 6 of 10 comparisons were significant at $p \leq .10$. Oddly, Time 1 and Time 16 were not significantly different; but at Time 17, 4 of the 10 comparisons were significant (all for the Pattern Recognition Test). By the last test—Time 18—8 of 10 comparisons reached significance.

DISCUSSION

The most interesting results were those obtained with the two main spatial-processing tests and the two linguistic/symbolic manipulation tests. The decrease of from 100 to 500 ms in RT for correct responses is a positive result that may well prove operationally noteworthy as faster RTs in the cockpit can be critically important in emergencies. The decrease in RT is not likely to be a practice effect because subjects had 11 administrations of the G-PAB before the experiment. Examination of the training data revealed that the training sessions resulted in asymptotic performance and were comparable to the 12 used in a similar study employing the WR-PAB (Gillooly, Smolensky, Albright, Hsi, & Thorne, 1990).

Somewhat surprising, response rate (composed of hit and error rates) increased linearly for three of four tests as the simulated SUSOP progressed (Table 1). The linear increase in hit rate for the Manikin and Grammatical Reasoning tests indicates subjects continued to perform well on these tests throughout the experiment. None of the tests showed a decrease in hit rate—another positive result.

Spatial-Processing Tests

For the spatial-processing tests, the error rate also increased linearly with time. For these tests, increasing response rate was due, in part, to an increase in error rate. For the Pattern Recognition Test, the error rate increase was enough to cause a significant decline in accuracy (Table 1 and Figure 2). This is an important finding because errors in the cockpit can have dire consequences and, depending on their severity and timing, may not always be correctable. As was stated, the increase in error rate was linear. If additional cycles and missions were added in this experiment, as they may well be in the fleet, the linear error rate increase could be a cause for concern.

Why does this increase in error rate occur? One possibility is that subjects are in a state of high arousal. According to Hockey (1986), the characteristics of a high-arousal state include increased selectivity of attention in dual-component tasks, reduced working-memory capacity, and increased speed with decreased accuracy in rapid decision-making tasks. The latter characteristic was present in this data set for the Pattern Recognition Test. This provides some support for the conclusion that the observed performance degradation was related to increased arousal from the experimental situation. It is also consistent with the informal observation that the subjects were highly motivated and very competitive. On the other hand, informal observation also indicated no outward signs of arousal. One drawback of a laboratory SUSOP experiment is the difficulty in reproducing the cyclical arousal pattern (an almost sinusoidal variation between the extremes of fear and boredom) present in an actual long-range attack mission.

The same pattern of responding also can indicate a fatigue-induced change in strategy to one involving a greater acceptance of risk. Subjective measures taken in a related study using the same tests and experimental design indicated subjects were moderately fatigued by this SUSOP scenario (Shappell et al., in press). Holding's (1983) COPE model predicts fatigued subjects will make more risky choices when given the opportunity. According to this model, prolonged work leads to less active control over behavior and the selection of easy but risky alternatives (Hockey, 1986). There were no obvious response alternatives in our experiment. However, for the spatial tests, the linear trends toward responding at faster rates and committing more errors are consistent with greater risk acceptance. Thus, subjects may be tolerating more errors to save time. If so, this behavior pattern could have negative operational consequences.

We emphasize that these data do not indicate subjects cannot continue functioning at relatively high levels. In fact, the increasing hit rates and decreasing RTs indicate otherwise. This type of result is not unique. Clear decrements in performance associated with extended operations and sleep deficit are hard to find (Johnson & Naitoh, 1974). For example, in a study of extended operations in a helicopter simulator, subjects continued to fly well even after 5 days with little sleep (Krueger, Armstrong, & Cisco, 1985). However, they made occasional cognitive and judgmental errors. These results may parallel ours. Our subjects presumably became more fatigued as the SUSOP progressed. Some supporting evidence is found in the pattern of significant contrasts in Table 2. The two spatial tests degraded somewhat by a little over 50 hr (Time 15) into the experiment. Although still operating within acceptable limits, subjects' performance was consistent with a change in cognitive strategy—tolerating more errors to proceed more rapidly.

The magnitude of the fatigue-induced performance changes described here was not large. The suggestion that subjects may have changed their strategy to one involving higher levels of risk is based on consistent but subtle changes in the data. Effects of this type and magnitude are not unusual in studies involving fatigue. Traditionally, experiments that induce fatigue have only occasionally produced significant effects, which Holding (1983) attributed to three factors. First, any change in the experimental protocol may be as effective as a rest period. In our study, subjects alternated between a variety of tests, perhaps enabling output to remain generally high. Second, motivated subjects are able to overcome fatigue on clearly defined or highly structured (as opposed to open-ended or self-paced) tasks. The tests employed in our experiment were highly structured, and the Marine subjects appeared highly motivated. Third, primarily central, cognitive changes (rather than peripheral, end-organ ones), often indicated by an aversion to effort, are expected from fatigue. These subtle changes may have contributed to the data of our experiment.

Time-of-day effects also may have contributed. These effects can even outweigh those of sleep loss and exercise (Englund et al., 1985). However, their role in our experiment is unclear. Although there was apparent cyclical variation to the data (see Figures 2 and 3), there is no obvious relationship between performance and time of day. For example, Table 2 reveals that performance changes between Time 1 and Times 17 and 18 may be due to an early morning circadian trough. The six differences at Time 15, however, are not so easily explained. Because time-of-day effects were not a central issue in this investigation of a specific military scenario, statistical analyses to detect rhythmic changes in the data were not performed.

Regardless of the factors contributing to the performance changes, it is the nature of those changes that are of greatest interest to us. The real concern regarding fatigued but motivated aircrew may not be related to changes in traditional measures of their efficiency as much as changes in their approach to work. The changes may be primarily central in character (Holding, 1983), leading to more risk-taking behavior. Also, because flying relies so heavily on spatial processing and because the results with the spatial tests were replicated (Shappell et al., in press), this finding warrants further attention.

Linguistic/Symbolic Manipulation Tests

There was no similar pattern of error rate increase for the linguistic/symbolic manipulation tests. Reasons for this are unclear. It can be due to the different types of processing involved or less difficulty in the tests chosen to represent the linguistic/symbolic manipulation category.

Input, Detection, Identification Test

The Four-Choice Serial Reaction Time Test did not reveal any changes in response measures over time. This may have been because the performance decrements were linked to fatigue, and perceptual fatigue is most evident in tasks with strong central components (Holding, 1983). The version of the Four-Choice test employed here likely involves little central processing (Perez et al., 1987).

To the extent that fatigue induced the changes on the spatial tests, even in the relatively benign laboratory environment, countermeasures are worth investigating to prevent or lessen performance degradation. Based on these data, any potential countermeasures should be closely examined for their effectiveness on accuracy measures and error rates. A successful countermeasure should enable subjects to resist any tendency to take risks in exchange for saving time or effort and to focus on the task at hand. Our results also indicate that any countermeasures should be introduced to this experimental paradigm shortly after 50 hr, after significant performance degradation has occurred according to the planned comparisons.

ACKNOWLEDGMENTS

We thank Dr. D. Reeves, whose efforts were critical in the design of the experiment. We also thank Ms. K. Winter for computer-programming support and Mr. L. G. Meyer, Dr. P. Dahms, and Dr. R. R. Stanny for assistance in data collection. The assistance of T. Green, R. Jay, D.

Charles, V. Foti, E. Collins, and M. Benjamin in data collection is also gratefully acknowledged.

Trade names of materials and/or products of commercial or nongovernment organizations are cited as needed for precision. These citations do not constitute official endorsement or approval of the use of such commercial materials and/or products. Volunteer subjects were recruited, evaluated, and employed in accordance with the procedures specified in the Department of Defense Directive 3216.2 and Secretary of the Navy Instruction 3900.39 series. These instructions are based on voluntary informed consent and meet or exceed the provisions of prevailing national and international guidelines. The views expressed in this article are ours and do not reflect the official policy or position of the Department of the Navy, Department of Defense, or the U.S. Government. This work was sponsored by the Naval Medical Research and Development Command under Work Unit 62233N MM33P30.001-7001.

REFERENCES

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

Babkoff, H., Caspy, T., Mikulincer, M., & Sing, H. C. (1991). Monotonic and rhythmic influences: A challenge for sleep deprivation research. *Psychological Bulletin, 109,* 411–428.

Babkoff, H., Mikulincer, M., Caspy, T., Carasso, R. L., & Sing, H. (1989). The implications of sleep loss for circadian performance accuracy. *Work and Stress, 3,* 3–14.

Babkoff, H., Thorne, D. R., Sing, H. C., Genser, S. G., Taube, S. L., & Hegge, F. W. (1985). Dynamic changes in work/rest duty cycles in a study of sleep deprivation. *Behavior Research Methods, Instruments, and Computers, 17,* 604–613.

Baddeley, A. D. (1968). A 3-minute reasoning test based on grammatical transformations. *Psychonomic Science, 10,* 341–342.

Bartlett, F. C. (1943). Fatigue following highly skilled work. *Proceedings of the Royal Society (B), 131,* 147–257.

Benson, A. J., & Gedye, J. L. (1963). *Logical processes in the resolution of orientation conflict* (Report No. 259). Farnborough, UK: Royal Air Force Institute of Aviation Medicine.

Coleman, R. M. (1986). *Wide awake at 3:00 A.M.* New York: Freeman.

Czeisler, C., Moore-Ede, M., & Coleman, R. (1982). Rotating shift work schedules that disrupt sleep are improved by applying circadian principles. *Science, 217,* 460–463.

Englund, C. E., Reeves, D. L., Shingledecker, C. A., Thorne, D. R., Wilson, K. P., & Hegge, F. W. (1987). *Unified tri-service cognitive performance assessment battery (UTC-PAB), I. Design and specification of the battery* (Report No. 87-10). San Diego: Naval Health Research Center.

Englund, C. E., Ryman, D. H., Naitoh, P., & Hodgdon, J. A. (1985). Cognitive performance during successive sustained physical work episodes. *Behavior Research Methods, Instruments, and Computers, 17,* 75–85.

Gillooly, P. B., Smolensky, M. H., Albright, D. L., Hsi, B., & Thorne, D. R. (1990). Circadian variation in human performance evaluated by the Walter Reed Performance

Assessment Battery. *Chronobiology International, 7,* 143-153.

Heslegrave, R. J., & Angus, R. G. (1985). The effects of task duration and work-session location on performance degradation induced by sleep loss and sustained cognitive work. *Behavior Research Methods, Instruments, and Computers, 17,* 592-603.

Hockey, G. R. J. (1986). Changes in operator efficiency as a function of environmental stress, fatigue, and circadian rhythms. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Vol. 2. Cognitive processes and performance* (pp. 44-1 to 44-49). New York: Wiley.

Holding, D. H. (1983). Fatigue. In R. Hockey (Ed.), *Stress and fatigue in human performance.* New York: Wiley.

Jerison, H. J., & Argintenu, J. (1958). *Time judgments, acoustic noise, and judgment drift* (Report No. WADC-DR-57-454). Dayton, OH: Wright Air Development Center.

Johnson, L. C., & Naitoh, P. (1974). *The operational consequences of sleep deprivation and sleep deficit* (Report No. AGARD-AG-193). Neuilly-sur-Seine, France: NATO Advisory Group for Aerospace Research and Development.

Keppel, G. (1973). *Design and analysis: A researcher's handbook.* Englewood Cliffs, NJ: Prentice-Hall.

Krueger, G. P. (1989). Sustained work, fatigue, sleep loss and performance: A review of the issues. *Work and Stress, 3,* 129-141.

Krueger, G. P. (1991). Sustained military performance in continuous operations: Combatant fatigue, rest and sleep needs. In R. Gal & A. D. Mangelsdorff (Eds.), *Handbook of military psychology* (pp. 244-277). New York: Wiley.

Krueger, G. P., Armstrong, R. N., & Cisco, R. R. (1985). Aviator performance in week-long extended flight operations in a helicopter simulator. *Behavior Research Methods, Instruments, and Computers, 17,* 68-74.

Mullaney, D. J., Kripke, D. F., Fleck, P. A., & Johnson, L. C. (1983). Sleep loss and nap effects on sustained continuous performance. *Psychophysiology, 20,* 643-651.

Mullaney, D. J., Kripke, D. F., Fleck, P. A., & Okudaira, N. (1983). Effects of sustained continuous performance on subjects working alone and in pairs. *Perceptual and Motor Skills, 57,* 819-832.

Naitoh, P. (1981). Circadian cycles and restorative power of naps. In L. C. Johnson, D. I. Tepas, W. P. Colquhoun, & M. J. Colligan (Eds.), *Biological rhythms, sleep and shift work* (pp. 553-580). Jamaica, NY: Spectrum.

Neri, D. F., & Gadolin, R. E. (1990). *Sustained/Continuous Operations Subgroup of the Department of Defense Human Factors Engineering Technical Group: Program summary and abstracts from the 5th semiannual meeting* (Report No. SR90-1). Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Nicholson, A. N., & Stone, B. M. (1987). *Sleep and wakefulness handbook for flight officers* (Report No. AGARD-AG-270 (E)). Neuilly-sur-Seine, France: NATO Advisory Group for Aerospace Research and Development.

Perez, W. A., Masline, P. J., Ramsey, E. G., & Urban, K. E. (1987). *Unified tri-service cognitive performance assessment battery: Review and methodology* (Report No. AAMRL-TR-87-007). Dayton, OH: Armstrong Aerospace Medical Research Laboratory.

Ryman, D. H., Naitoh, P., & Englund, C. E. (1984). Minicomputer-administered tasks in the study of effects of sustained work on human performance. *Behavior Research Methods, Instruments, and Computers, 16*(2), 256-261.

Shappell, S. A., Neri, D. F., & DeJohn, C. A. (in press). Simulated sustained flight operations and performance, Part 2: Effects of dextro-methamphetamine. *Military Psychology.*

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

Wever, R. A. (1981). On varying work-sleep schedules: The biological perspective. In L. C. Johnson, D. I. Tepas, W. P. Colquhoun, & M. J. Colligan (Eds.), *The twenty four hour workday: Proceedings of a symposium on variations in work-sleep schedules* (Publication No. 81-127, pp. 51-86). Washington, DC: U.S. Department of Health and Human Services.

Wilkinson, R. T., & Houghton, D. (1975). Portable four-choice reaction time test with magnetic memory. *Behavior Research Methods and Instrumentation, 7,* 441-446.