(12)

Report SAM-TR-80-49

FATIGUE STRESSORS IN SIMULATED LONG-DURATION FLIGHT
Effects on Performance, Information Processing, Subjective Fatigue, and Physiological Cost

Layne P. Perelli, Captain, USAF

December 1980

Final Report for Period March 1977 - January 1980

Approved for public release; distribution unlimited.

USAF SCHOOL OF AEROSPACE MEDICINE
Aerospace Medical Division (AFSC)
Brooks Air Force Base, Texas 78235

NOTICES

This final report was submitted by personnel of the Crew Performance Branch, Crew Technology Division, USAF School of Aerospace Medicine, Aerospace Medical Division, AFSC, Brooks Air Force Base, Texas, under job order 7930-10-37.

The voluntary informed consent of the subjects used in this research was obtained in accordance with AFR 169-3.

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

LAYNE P. PERELLI, Captain, USAF
Project Scientist

William F. Storm
WILLIAM F. STORM, Ph.D.
Supervisor

Roy L DeHart
ROY L. DEHART
Colonel, USAF, MC
Commander

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

(11) 1. REPORT NUMBER
SAM-TR-80-49

2. GOVT ACCESSION NO.
AD-A105 484

3. RECIPIENT'S CATALOG NUMBER

(6) 4. TITLE (and Subtitle)
FATIGUE STRESSORS IN SIMULATED LONG-DURATION FLIGHT.
Effects on Performance, Information Processing, Subjective Fatigue, and Physiological Cost.

(9) 5. TYPE OF REPORT & PERIOD COVERED
Final Report.
March 1977 — January 1980

~~6. PERFORMING ORG. REPORT NUMBER~~

(10) 7. AUTHOR(s)
Layne P. Perelli, Captain, USAF

8. CONTRACT OR GRANT NUMBER(s)

9. PERFORMING ORGANIZATION NAME AND ADDRESS
USAF School of Aerospace Medicine (VNE)
Aerospace Medical Division (AFSC)
Brooks Air Force Base, Texas 78235

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
(17)
62202F
(16) 7930-18-37

11. CONTROLLING OFFICE NAME AND ADDRESS
USAF School of Aerospace Medicine (VNE)
Aerospace Medical Division (AFSC)
Brooks Air Force Base, Texas 78235

(11) 12. REPORT DATE
December 1980

13. NUMBER OF PAGES
186

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
(12) 188

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Fatigue; information processing; human performance measurement; workload in flight; flying performance; response blocking; work-rest cycles; subjective measures; field measurement technology; adaptive measurement; sleep deprivation; physiological cost; task complexity; body temperature; circadian rhythms; heart rate variability; and long-duration flight.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The purpose of this study was to assist in the development of flight duration parameters for scheduling aircrew work-rest cycles. Twenty-four Airmen were selected to represent the USAF pilot population on the basis of flight aptitude scores, class II flight physicals, and personal interviews. Each received an intensive 7-day flight training program in Link GAT-1 Trainers. Then half of the subjects were randomly assigned to a schedule that regularly alternated 12-hr duty days with 12-hr rest periods for 4 days. The remaining subjects followed a schedule with duty days of 12, 24, 24, and 12 hr, with a 12-hr rest

DD FORM 1 JAN 73 1473

317000

UNCLASSIFIED LB
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT (Continued)

period between duty days. During each duty day, all subjects flew two 4.5-hr simulated flights. Flying performance was evaluated every hour by both simple and complex flight maneuver tests, administered and scored by a PDP-12 computer. The subject's threshold of information processing speed was determined hourly by a Discrete Information Processing Test (DIPT). (The DIPT is an adaptive, computer-controlled task developed for eventual use in the actual flight environment.) Continuous heart rate (HR), heart rate variability (HRV), fatigue, sleepiness scores, and sleep logs were collected throughout each mission.

All measures demonstrated significant fatigue effects. HR increased with flight task complexity only during extreme fatigue. HRV decreased with increasing task complexity for both groups, for all three performance measures. Performance on the less complex flight task declined most during extreme fatigue. Disrupted circadian rhythm during night flights seemed to cause greater performance decrement than cumulative fatigue. Performance measures, subjective fatigue, and sleepiness reports were significantly correlated with each other and with changes in rectal temperature. Subjects recovered from intense fatigue effects by the fourth duty day.

SUMMARY

The purpose of this study was to assist in the development of flight duration parameters for scheduling aircrew work-rest cycles. This aim was achieved by using objective measures of flying performance, subjective reports of fatigue and sleepiness, and three physiological indicators: heart rate (HR), heart rate variability (HRV), and rectal temperature. The second purpose was to demonstrate the correlation among these measures and relate them to a Discrete Information Processing Test (DIPT) developed for eventual use in the actual flight environment. The DIPT is a computer-controlled, 5-choice reaction-time task that adapts its stimulus presentation rate to the subject's response accuracy and continuity, up to a point at which the subject can no longer keep pace. Through an iterative process, the subject's threshold of information processing speed can be determined.

For this report, a literature review initially provides historical perspective for various concepts of fatigue, and evaluates past research concerning performance decrement accompanying sleep deprivation, physiological cost indices of fatigue, circadian rhythm changes, and flying skill assessment. The unique requirements for a device to assess performance in the field are presented as the rationale for the development of the DIPT. The DIPT is described within the framework of adaptive technology.

The hypotheses were that each dependent measure would change significantly in the presence of three fatigue stressors experienced during simulated flight: (a) time awake prior to flying (1 hr vs. 12 hr); (b) daily flight duration (9 hr); and (c) total mission duration (4 days). HR and HRV were predicted to be significantly related to the arousal value of each task, based on its complexity; and an interaction was predicted to occur between task complexity and fatigue level. Disrupted circadian rhythm was predicted to be more detrimental to flight performance than was cumulative fatigue.

To represent the USAF pilot population, 24 Airmen were selected on the basis of flight aptitude scores, class II flight physicals, and personal interviews. Each received an intensive 7-day flight-training program in Link Trainers. Then, 12 subjects were randomly assigned to a schedule that regularly alternated 12-hr duty days with 12-hr rest periods for 4 days. The remaining subjects followed a schedule with duty days--of 12 hr; 24 hr; 24 hr; and 12 hr--each duty day being separated by a 12-hr rest period. During each duty day, all subjects flew two 4.5-hr simulated flights, separated by a 1-hr rest. Flying performance was evaluated by a time-on-target tracking score derived from heading, altitude, airspeed, turn rate, turn coordination, and vertical velocity errors. Each flight hour, a PDP-12 computer (Digital

---

EDITOR'S NOTE: Available, at the close of this publication, is a selective list (plus definitions) of the "Abbreviations, Acronyms, and Symbols" used throughout.

Equipment Corp., 200 Forest St., Marlboro, Mass.) administered and scored a simple straight-and-level test, a complex flight maneuver test, and the DIPT. Continuous HR and rectal temperature, subjective fatigue and sleepiness reports, and sleep logs were collected throughout each mission.

All measures demonstrated significant fatigue effects. HR increased with flight task complexity, only during extreme fatigue. HRV decreased with increasing task complexity for both groups, for all three performance measures. Performance on the less complex flight task declined most during extreme fatigue. Disrupted circadian rhythm during night flights indeed seemed to cause greater performance decrement than cumulative fatigue. Performance measures, subjective fatigue, and sleepiness reports were significantly correlated with each other and with changes in rectal temperature. Subjects recovered from the intense fatigue effects by the fourth flight duty day.

CONTENTS

CONTENTS ("Results," Cont'd.) Page

Figures

Figure No. Page

(CONTENTS, Con'd.)

Tables

CONTENTS (Cont'd.)

CONTENTS (Cont'd.)

FATIGUE STRESSORS IN SIMULATED LONG-DURATION FLIGHT

Effects on Performance, Information Processing, Subjective Fatigue, and Physiological Cost

INTRODUCTION

The goal of human factors research is to develop safe, efficient, cost-effective systems which will accomplish their intended missions by maximizing the effectiveness of human behavior within those systems. This type of research is conducted either to support the design and development phase of system acquisition--or (as in the present study) to answer questions during the operational life of that system. The purpose of this study has been not only to assist in the development of flight duration parameters for scheduling aircrew work-rest cycles but also to predict the expected flying performance decrement and physiologic cost of an extended duration mission.

In certain emergency situations, Air Force pilots must perform extremely fatiguing missions over a considerable period of time. Although flying schedules of this type are not commonly flown by commercial airline pilots, the airline industry is also constantly concerned about the effects of fatigue on flying performance; and this study will provide some helpful information on the subject. Moreover, some civilian flying activities (such as crop dusting and forest-fire fighting) involve long periods of stressful low-level maneuvers generating fatigue levels similar to those in some military operations.

Piloting high-performance jet aircraft requires complex psychomotor coordination, high rates of information processing, and high-speed decision making. Psychological stress is created by the small margin for error, while long periods of vigilance are necessary in a relatively monotonous environment. Lack of practice or experience increases error rates; and new learning is frequently required of the pilot, often while he is already in flight. Moreover, pilots are required to adhere to schedules over which they have little control. These are dictated, for the most part, by management, weather, and maintenance considerations. In consequence, the combination of these various factors creates situations in which human behavior is highly susceptible to fatigue stressors.

This study has evaluated four sources of fatigue stressors: (a) time awake prior to flight; (b) daily mission duration; (c) total mission duration; and (d) circadian rhythm effects. The first stressor was a function of the overall length of the duty day. The second related to the buildup of fatigue during the flight. The third was a cumulative fatigue effect building up over all the days that the mission lasted. The fourth (the circadian rhythm effect) was due to the specific time of night or day when the flight took place, and was affected by the pilot's normal pattern of working and sleeping. Each stressor was postulated to have an adverse effect on flying performance, subjective feelings of fatigue, and the pilot's physiological functioning. Furthermore, evaluating all of these variables in the same study was necessary in order to understand the concept of fatigue well enough for it to serve as a basis for operational scheduling decisions.

The primary objective was to compare and interrelate the fatigue effects of two 4-day duty schedules for flying crews (12-hr crew rest/12-hr duty day) on the following measures:

1. A Discrete Information Processing Test (DIPT)
2. Two multidimensional tracking tests of flying performance varying in complexity
3. Heart rate (HR) and heart rate variability (HRV)
4. Urine components
5. Body core temperature
6. Subjective reports of fatigue and sleepiness.

Past studies have: failed to use objective measures of pilot performance; used mission lengths either too long or too short to generate operationally significant levels of fatigue; used only an analogue of flying behavior (e.g., a psychomotor tracking task); used measures too insensitive to find a decrement; or failed to interrelate systematically the fatigue effects on performance decrement, subjective feeling, and physiological costs of pilot behavior. The present study has avoided these past deficiencies by using a recently developed simulator-based physiological and flying performance measurement system and simulating a long-duration mission scenario.

A secondary objective of this study was to help develop the measurement technology necessary to evaluate the quality of human performance in the field. Despite the obvious need for this type of research, very little has been done to develop this technology because of the difficulty of collecting data in operational settings, and because of the lack of reliable measures even in a laboratory.

In the following review of the literature, the concept of fatigue is analyzed from the standpoint of the three classes of measures used historically to detect its effects: objective performance decrement, subjective report, and physiological cost indicators. Fatigue is then related to the concept of workload and reserve capacity, and the use of response blocking as an indicator of fatigue is also examined.

Because of a dearth of research on long-duration flight, the sleep deprivation studies and research involving continuous nonflying performance lasting longer than a normal duty cycle have been used as additional sources of information on fatigue effects. Presented next is a review of HR, HRV, and urinalysis research which relates to flying performance, fatigue, and the arousal hypothesis.

Then, because studies analyzing behavior over extended periods of time must consider changes due to daily variations in performance, a review of circadian rhythm research (pertinent to the present study) is also provided. Thereafter a review of the previous methods used to measure pilot performance objectively was undertaken to gain insight into past problems and to identify those techniques most sensitive to flight fatigue.

In the subsection on "Developing an Adaptive RT Task for Field Studies," the unique requirements a task must meet for its use in operational situations are identified and these requirements provide the rationale for the development of the DIPT. A description of this Reaction Time (RT) task is given and analyzed in the framework of adaptive technology. In the two closing subsections ("Statement of Hypotheses," and "Analysis of Variables") the hypotheses tested in the study are presented and the independent and dependent variables are defined.

The Concept of Fatigue

The following analysis of the concept of fatigue is taken in part from: Bartley and Chute (15), Bills (22,26), Cameron (42), Chambers (46), Floyd and Welford (73), Gartner and Murphy (80), Ryan (185), and Welford (210,211). The basic question of the present study (how long can a person continue to work effectively--and concomitantly, what does one measure to determine this limit?) dates back to early studies of industrial work output and productivity (Bills: 24; and Ryan: 185). The studies were based on the notion that the output of the worker was limited by the buildup of fatigue, and that any means whatsoever of reducing the fatigue would increase productivity. Thus, these studies dealt with such matters as the length of workday and workweek, shift arrangement, illumination, ventilation, workplace design, and plant layout. Little theoretical interest existed in the concept of fatigue as long as the technique employed increased productivity (Chambers: 46). From about 1940, the second and third focal points of interest in fatigue centered around the activities of flying and driving (Cameron: 42). The concern had shifted from corporate profit to safety and prevention of loss of life.

The inability to develop an adequate definition of fatigue has always plagued this type of research (Muscio: 151; Welford: 212; Cameron: 42; and Gartner and Murphy: 80). Part of the problem lies in trying to define fatigue as that which is being measured. Previous research has had difficulty interrelating the three diverse classes of fatigue measurement: performance decrement, subjective feelings of fatigue, and physiological changes. Another part of the problem stems from a reluctance to view fatigue simply as a hypothetical construct which is used to explain changes in foregoing measures, or to give fatigue an operational definition based on type of activity and time engaged in the activity (Cameron: 42). The classic Cambridge Cockpit Studies of Davis (55), Bartlett (13), and Drew (59) relied primarily on decrement measures of flying performance obtained during 2- to 6-hr simulated flights in a modified Spitfire simulator as measures of fatigue. Bartlett (14) provides one of the clearest expressions of this point of view as a definition of fatigue:

> "Fatigue is a term used to cover all those determinable changes in the expression of an activity which can be traced to the continuing exercise of that activity under its normal operational conditions, and which can be shown to lead, either immediately or after delay, to deterioration in the expression of that activity, or, more simply, to results within the activity that are not wanted."

One of the earliest attempts to set appropriate standards of operations to preclude excessive flying fatigue is a comprehensive report to the U.S. (CAA) Civil Aeronautics Administration (202) from a committee headed by Morris S. Vitales. Most of the research reviewed by the committee showed little, if any, significant performance decrement; but reports of subjective feelings of fatigue received strong emphasis. A general theme in their report was that fatigue was a major contributing factor to accidents by causing impaired performance, but that quantification of this effect would be extremely difficult. In follow-on research requested by this committee, Hartman and Fitts (93) used an observational measurement technique to study alertness in air-traffic controllers. Their hope was that the term "alertness," as an inverse analogue to fatigue, could be defined more specifically in order to overcome the imprecision in the term "fatigue." In addition, the implicit assumption was that changes in alertness--defined as: the sensitivity of the operator to job-associated stimuli--were directly related to his level of job performance. Job performance was defined as: the relative adequacy of the operator's interpretation of job-associated stimuli, and the effectiveness of his responses on the job. In this approach, job performance was not measured directly but was inferred from reliable observations of generally relevant behavior.

In 1949, Carmichael and Kennedy (45) developed an alertness monitor which basically monitored the muscle tone of a subject's forehead and was capable of predicting, due to momentary lapses of attention, when long RT's would occur. Although the apparatus was purported to be successful, it was never put into widespread use either in research or in the field, primarily for practical reasons (according to Kennedy: 17). The equipment was bulky and heavy, and required relatively sophisticated electronics for its day; the subject had to wear a headband, and especially trained technicians were necessary. Furthermore, the device had only a limited power supply. Interestingly, Kennedy questioned the usefulness of the device, even though these shortcomings could have been overcome.

In 1884, Mosso (148) performed the first experimental study of muscular fatigue. Although he noted the changes in the subjective feelings of his subjects, usually in the form of increased irritability, his primary measure was performance decrement as measured by an ergograph. In 1917, however, Dodge called for reporting of "sensations of weariness" in all fatigue oriented experiments as possible "subjective indicators of real fatigue" (57: p. 110). In 1944, Ryan (184) attacked the prevailing objective view of fatigue, which was then defined as work decrement--because many studies failed to find output decrements when the subjects were obviously fatigued, and because variability of performance made authentic performance changes difficult to establish. He also noted that work decrement was not as objective as purported, since the decrements could be due to factors other than fatigue. He thus proposed that the definition of fatigue focus on the subjective report of various bodily conditions. Instead of providing unity, however, this concept would have consisted of several varieties of subjective fatigue.

Not until the year after the CAA Report (202) did Bartley and Chute (15), in one of the most comprehensive analyses of fatigue as an explanatory concept, make subjective feelings of fatigue the primary emphasis of the definition. These researchers distinguished between three distinct facets of the problem, and offered a term for each. The term "impairment" was to identify the true reduction of physical capacity which resulted from an accumulated oxygen debt in muscle tissue. "Work decrement" was to describe deterioration in quality of performance for reasons other than sheer physical incapacity. The term "fatigue" was to describe only "subjective feelings of lassitude and disinclination towards activity." Whiting and English (215) had used subjective fatigue in this sense, as an explanatory concept for work decrement. Fatigue was supposed to create a "negative emotional appetite [for continued performance].... Fatigue does not directly cause a work decrement but raises the threshold at which certain work motives are effective" (215: p. 49). (See also Weiskotten and Ferguson: 209.) In their research, Whiting and English found no performance decrements; thus, they supposed feelings of subjective fatigue to occur before work decrement.

In a review of the literature, Johnson and Naitoh (109) reached the same conclusion: Subjective fatigue serves primarily as a warning that biological resources are being overtaxed, but performance can be maintained if extra effort is expended.

Workload, Reserve Capacity, and Fatigue

A modern explanation of why performance decrement is often not observed, even though performance requirements or workload is increased, uses the concept of reserve capacity (Brown: 38). This theory proposes that man rarely works at his maximum output level, and can thus momentarily absorb additional cognitive or physical workload and maintain his baseline performance, providing his reserve capacity is not exceeded. In the present study, the increase of fatigue is viewed as reducing man's reserve capacity, and the concepts of workload and fatigue are considered to be highly interrelated.

This relationship between workload and fatigue has recently been thoroughly analyzed by Gartner and Murphy (80). As with fatigue, much confusion arises over the concept of workload because it has been defined along three broad lines: objective task demands, subjective estimates of ability to cope with task demands, and the metabolic load imposed by task demands. Just as the four fatigue stressors in the present study involve the concept of time, the concept of workload involves performing a given amount of mental or physical work at a given rate per unit of time.

In this study, the performance decrement observed due to fatigue is considered to be a function of both the time at task and the total amount of work performed in a given unit of time. As such, the present study could be conceptualized as a study of the effects of workload on pilot performance. Metabolic load factors are due to energy expended during physical work and are measured by physiological indicators, such as oxygen consumption, respiration rate, and heart rate. Mental load factors are due to the amount, speed, and complexity of information required to be processed, transferred, or held in memory in a given period of time. Both mental and physical load factors are subject to overload which can cause performance decrement. This type of decrement is indistinguishable from that due to fatigue: they are both caused by excessive task demands during the time required for the task. However, overload conditions can be viewed as accelerating the buildup of fatigue and reducing one's reserve capacity. Usually, significant or excessive time at task is required to induce fatigue, whereas overload can occur in a very short period. By viewing fatigue as a result of an excessive workload condition which reduces reserve capacity, one may gain some insight into why performance decrement apparently does not occur under these conditions--even though subjective fatigue does.

Subjective Fatigue and Performance Decrement

The results of the study by Poffenberger (171) have often been quoted as demonstrating that no relationship exists between subjective fatigue and work output; but, as Barmack (11) has pointed out, Poffenberger did not intend to deny this relationship completely. From his data, Poffenberger noted that:

> "When the individual cases are examined there is at least the suggestion that those who show the greatest falling off in output also show the greatest changes in the feelings, and those who show the least falling off in output also show the least change in the feelings" (171: p. 467).

Pierson (169) measured subjective fatigue and simple RT up to the subject's limit of endurance, as defined by signs of obvious physical distress or refusal of the subject to continue. He concluded that: (a) The subjective experience of fatigue is not a valid criterion of the ability to perform speed or endurance-type muscular work. (b) Fatigue and endurance cannot be measured by work decrement. (c) Fatigue, endurance, and work decrement are independent variables. However, even though all subjects subsequently complained of tenderness in the area of the brachioradialis (which was taken as an indication that a near-maximal effort was obtained), the average total time that subjects performed the task was only about 30 min. It is questionable that a very high degree of fatigue was induced, and the effect induced may have been more physical pain than the type of tiredness associated with fatigue.

Still, other studies of longer duration have failed to find a relationship between subjective fatigue and performance decrement. Pearson (166), using the Multidimensional Pursuit Test (in which subjects attempt to center four randomly moving meter needles by means of a stick, rudder, and throttle control system), found no relationship between subjective fatigue and performance decrement over a 3-hr period. Again, however, it is questionable as to whether the task situation introduces sufficient fatigue for a valid test of the hypothesis. By using the USAFSAM Subjective Fatigue Checkcard (Pearson and Byars: 166), which has a range of 0 to 20, the lowest mean score obtained was approximately 9 (see Appendix A, Fig. A-1). Subjective reports are very susceptible to range effects (Poulton: 174), and the subject tends to spread out his responses to fit the entire response range offered. He tends to overrate his feelings of fatigue until he had been exposed to the full range of fatigue--presumably, complete physical exhaustion--which would require several days of heavy activity. The fact that, even with this effect operating these subjects had used only half the scale, is indicative that fatigue in that study was mild.

By definition, subjective fatigue has no obvious, objective set of referents. In order to determine whether subjective fatigue can be used to predict performance decline, the intensity of the fatigue experience must be sufficient to elicit a strong fatigue response on the part of the subject. Schreuder (187: p. 41) provides a good definition of such fatigue:

> "Pilot fatigue is that subjective sense of weariness resulting from the duties of piloting an aircraft which is considered to be excess of the expected normal fatigue and which is cumulative and of such amount as to alter the pilot's judgment and ability, compromise safety, and from which the pilot does not recover in a rest period of reasonable duration.

He makes a distinction between: the ordinary tiredness resulting from a normal day's work, and fatigue. Specifically, when the tiredness is such that it does not dissipate during normal rest, fatigue has set in. He further states that fatigue, as thus defined, is not a common occurrence in the airline pilot. To support this contention, Schreuder (187) also cites data showing that airline pilots as a group do not develop more illness or cardiovascular disease than the normal population. Schreuder would say that one is not subjectively fatigued unless one's judgment and ability have so declined that safety is compromised. That is, performance decline must occur with true subjective fatigue. The problem with this approach is that, depending on the task, the performance decline may not be evident until a dangerously high level of subjective fatigue has been experienced.

A way out of this conceptual dilemma was provided by Hartman (89) by suggesting three categories of fatigue: <u>acute fatigue</u>, which normally occurs between a pair of sleep periods; <u>cumulative fatigue</u>, which accumulates over a period of days or weeks owing to the inadequate recovery from successive periods of acute fatigue; and <u>chronic fatigue</u>, a

> "psychoneurotic syndrome characterized by difficulty in committing oneself to an aggressive course of action, and by a generalized withdrawal or retreat from conflict which is intolerable for situational or personality reasons. Chronic fatigue, in this sense, is rarely seen in the military pilot" (89: p. 94).

In this classification, <u>acute fatigue</u> would cause negligible performance decrement or physiological response, yet a feeling of fatigue would be admitted. During the acute phase, only tasks requiring the highest degree of performance capability would show a decrement. These tasks would not normally be found in a real-world work environment. (If they were, presumably the performance decrement would be acceptable and have no adverse consequences or compromise safety.) During <u>cumulative fatigue</u>, the most operationally significant performance decrement would occur. However, for sustained work periods lasting longer than a normal duty period, detecting the transition from the acute to the cumulative phase is difficult. The buildup of fatigue would be large enough to effect significant performance decrement, even though the subject would still technically be in the acute phase.

The primary practical reason for studying excessive workload and fatigue is to minimize its unwanted effects on behavior and health. Mohler (145) provides an extensive list of clinical symptoms and indicators of both physical and mental fatigue, along with a thorough list of possible fatigue-related causes of serious performance decrement in aviation activities. Gartner and Murphy list aircraft accidents and severe impairment of individual health or well-being as the primary effects to be avoided, but note that "workload and fatigue are seldom cited as direct causes or even major contributing factors to these tragic circumstances" (80: p. 38). Hartman (89) reports that fatigue has been identified as a factor in only 2.7% of USAF aircraft accidents over a 5-yr period, but notes that practicing flight surgeons are aware that fatigue is a common contributor to pilot error. On occasion, however, fatigue is recognized for its contribution to accidents. Barlay (10), for example, describes a fatal aircraft accident in which fatigue was ranked as second in 11 contributing factors; but the pilots were within the officially established limits of duty and rest times. On the other hand, reduction of subjective fatigue is pursued only to the extent that it is highly correlated with undesirable operational consequences, such as accidents, illness, high absenteeism, or severe morale problems. The U.S. Air Force has an additional operational consequence of possible mission failure.

Gartner and Murphy (80) have listed 4 major categories of unwanted effects of fatigue, in order of increasing operational impact:

1. motivation decrement,
2. skill or proficiency decrement,
3. psychological stress, and
4. performance decrement.

Welford (211) has listed 4 general ways in which performance decrement might manifest itself:

1. impairment of sensory or perceptual functions,
2. slowing of sensory-motor performance,
3. irregularity of timing, and
4. disorganization of performance.

Fatigue resulting from flying operations has also been shown to reduce learning ability. Welford, Brown, and Gabb (213) administered a series of learning problems to two groups of radio officers on civilian aircrews. One group was tested first within 1/2 hr of returning from a flight, and was then retested after a stand-down of at least 8 days. The second group was tested the day before departing for a flight, and retested immediately upon return from trips lasting between 6 and 21 days. It was found that solving electrical problems requiring measurement of resistances with a meter took longer, and required more readings be taken for the first group when tested immediately after return from a flight. This effect implied a possible short-term memory disturbance. When flights were divided into "easy" and "hard" by aircrew consensus, solutions were found to be less accurate after "hard" flights. Crews tested before a flight showed no impairment when retested after a fatiguing flight. However, crews tested first in the fatigued condition continued to show impairment when retested after stand-down. Welford et al. (213) have suggested that the secondary impairment may have been caused by fixation to an inadequate performance strategy that the crews had learned while fatigued.

In summary, the concept of fatigue is best defined as a subjective feeling; and, provided that the intensity of the fatigue stressor is severe enough, the subjective report of fatigue should be expected to be correlated with a wide range of performance decrements.

Response Blocking as an Indicator of Fatigue

Because performance has never been shown to decline linearly with time, early researchers have had difficulty in demonstrating an inverse relationship between fatigue and performance. The major result of fatigue is to increase the variability of performance (i.e., even though someone is very fatigued, usually his performance will appear normal; but, infrequently, it will be substandard). From a work-output standpoint, the overall decline in productivity may be slight. But, from an air-safety standpoint, this increased variability is hypothesized both to relate fatigue to accidents and to impair flying performance. This phenomenon appears to have been recognized first by Woodworth and Wells (232), and has subsequently been referred to as "fluctuations" or "oscillations" in behavior (Entwistle: 66; Flügel: 74; and Philpott: 168). Bills (22, 23, 27) has made a thorough investigation of this effect during short-term fatigue and has referred to it as "blocking."

Bills studied five forms of mental, self-paced, continuous work: alternate addition and subtraction, reversible perspective of ambiguous figures, color naming, naming 100 opposites to words, and substitution of letters for numbers. He defined a "block" as: a pause in responding, equivalent to the time of two or more average responses. Blocks were considered involuntary rest pauses which increased in frequency and duration with the buildup of fatigue. Initially, the tasks lasted 7-10 min; later he extended their length to 1 hr. Blocks occurred approximately 3 times a minute, but with large individual variation. Blocks diminished with practice, but increased in frequency and duration with fatigue after approximately 1 hr. He found that individuals who responded most rapidly tended to produce fewer and shorter blocks, and that errors tended to occur with blocks.

The earliest theoretical formulation of fatigue was developed from a concept borrowed from physiology, the neural refractory period. After the passage of a neural impulse along the axon of a single neural fiber, it is impossible to produce a second impulse until a short period of time has elapsed. This interval is referred to as the absolute refractory period. During the following relative refractory period, an impulse can be produced but a stronger stimulus is required. In 1913, Vernon (203) had shown that repetitive stimulation of an isolated neural fiber lengthens its refractory phase. Dodge (57) applied this concept to the fatiguing of mental processes. Telford (196) had supposedly demonstrated the existence of a minimum interval (approximately 5 sec) within which successive auditory stimuli could not be detected. In later studies, this interval became known as the psychological refractory period in which an inverse relationship is found between: reaction time to the second stimulus of a closely spaced pair of stimuli, and the time between the two stimuli. Welford (211) has called the analogy to the neurological refractory period unfortunate; for the minimum interval required for response is now believed probably due to the time required by central information processing mechanism to translate from stimulus to response, and not to the neural phenomenon. No known physiological refractory periods last as long as those noted in research on psychological refractory periods. However, some recent auditory evoked-potential data have been presented as physiological evidence of a psychological refractory period for processing closely spaced stimuli (Surwillo: 193).

Bills thought that the rest pause occasioned by blocking allowed fatigue to dissipate and functioned similarly to the cumulative refractory period observed in nerves. This explanation has never been satisfactorily demonstrated and has been challenged by Barmack, who attributed blocks to "a motivated shift of attention from the work serving as an antihypnotic and as a means of temporarily escaping from an unpleasant task" (11: p. 79). Broadbent (36) has explained blocks in a similar manner, using his filter theory. He suggests that the mechanism which selectively extracts pertinent data from the task at hand becomes temporarily ineffective and allows irrelevant information to gain attention, causing either a loss of speed or accuracy or a missed signal. This momentary failure will produce temporary spurts in performance and, because of this compensation, no overall performance decrement will be observed; however, increased distractability has often been observed in fatigued subjects (Broadbent: 35).

Welford (211) has proposed that peripheral stimuli, arising from feelings of subjective discomfort, may increasingly compete with task-relevant stimuli and capture central processing mechanisms, thus momentarily producing decrements in performance. He does not feel that a failure of central mechanisms is necessarily an explanation of the blocking phenomenon.

Using simple visual and auditory RT tasks, Foley and Humphries (75) found that blocks occurred uniformly throughout 14- to 60-min test periods; but no evidence was found that the occurrence of a block dissipated fatigue buildup, for no gradual increase occurred in RT up to a block and no abrupt decrease in RT after the block. However, since RT did not increase from the beginning to end of the trials, errors were not measured and no subjective fatigue data were collected, so these researchers (75) had no evidence of any fatigue during this type of task. In contrast, Bertelson and Joffe (28)--using a four-choice serial RT task lasting only 30 min--found that RT rose sharply during the four or five responses preceding a block, and fell abruptly following a block, thus supporting Bills' hypothesis that the function of blocks is to dissipate fatigue buildup. Regardless of the causes or function of blocking, it does seem to be the most reliable performance decrement observable as a result of fatigue.

Bills and Shapin (28) found that manual and vocal color-naming produced more blocks in a self-paced mode than when stimulus presentation was automatically controlled. Sampson (186) found that an Experimenter-paced (E-paced) test of serial addition caused subjects to block rather than give wrong answers with increased pacing speeds.

Other researchers (Bartlett: 13; Davis: 55; Drew: 59; Hartman: 90; Hartman et al.: 95; and Orr: 158), while not specifically measuring blocks in their studies, have (from observation) described the changes of the subject's behavior during increased fatigue as momentary lapses of acceptable performance similar to blocking.

The blocking phenomenon has been observed during psychomotor tracking behavior. Craik (53), Hick (101), and Vince (204) described man as an intermittent-correction servo, since their research showed that correction between target and cursor was not made continuously but at discrete intervals of about 0.5 sec. This type of discontinuity was subsequently related to the psychological refractory period, however--not to blocking. While not measuring blocking directly, Collins (49) described the single-axis, compensatory tracking performance of subjects with 34 hr of sleep deprivation as containing the lapses noted by Bills and others. Collins (49) concluded that 1 night of sleep deprivation was sufficient to cause decline in his aviation-related tracking task.

Teichner and Olson (195) have developed a preliminary theory to explain the effects of task and environmental factors on human performance. Their theory utilizes the concept of response blocking as the criterion which reflects changes in the integrity of the internal behavioral processes on which observable performance depends. They define a task as the transfer of information between man and machine in a system. Tasks are classified into five activities: sensing, searching, coding, switching, and tracking. Tasks can then be characterized by three components: (a) complexity, i.e., the number of different possible signals involved in the task; (b) duration, which amounts to a fatigue variable (also, a distinction is made between time at task and time exposed to an environmental stressor); and (c) the rate of signal input and response output requirements (corresponding to a workload factor).

Teichner and Olson (195) postulate five information handling processes within the central nervous system: (a) sensing, which consists of data-getting, and man-machine search tasks involving sensory and receptor-orienting activities; (b) short-term memory; (c) attention, serving as a varying bandwidth filter, whose characteristics are pre-set by long-term memory to maximize the acceptance of desired signals and minimize the acceptance of irrelevant inputs; (d) activation, representing the various excitatory and inhibitory processes which are associated with the reticular and limbic systems (Routtenberg: 180); and (e) tuning, the adjustment of physiological regulatory processes by various control centers of the central nervous system (see also Teichner: 194). Increases in activation are supposed to narrow the attentional bandwidth and increase the receptor scanning rate, thus reducing errors. Override mechanisms triggered by the activating mechanism can reverse the direction of ongoing compensatory physiological reactions. Reversal of bandwidth, from narrow to wide, causes a response block which is defined as the interruption of responding during a high-speed continuous task. Reversal phenomena are assumed to occur when the ratio of afferent input rate to re-afferent output rate of any bodily subsystem reaches a critical value. When the response block occurs, this ratio is altered and the bandwidth again narrows. These intermittent reversals of bandwidths will be observed as irregularly fluctuating performance. Task duration eventually reduces activation level, increases the attentional bandwidth, and alters the critical value for reversal.

Teichner and Olson (195) have analyzed the five types of task activities in terms of signal detection theory; and using Bills' 1937 data (25) demonstrating increased blocking during hypoxia, they have made an initial attempt to determine the point at which a variety of environmental stressors will produce unacceptable performance due to response blocking. Thus, the concept of blocking has proven useful as a generalized indicator of performance decrement sensitive to a wide range of stressors other than fatigue.

As shown in the foregoing literature review, response blocking is one of the behaviors most likely to index a relationship between subjective feelings of fatigue and flying performance decrement.

Performance Decrement and Sleep Deprivation

Comprehensive reviews of contemporary sleep deprivation studies have been provided by Johnson (108); Johnson and Naitoh (109); Lubin (136); Naitoh (152); Naitoh and Townsend (153); Wilkinson (223, 224); and Woodward and Nelson (230). Sleep deprivation studies, and continuous performance studies which induce sleep deprivation indirectly--both meet the requirement of producing a sufficient intensity of the fatigue stressor to demonstrate its effects on performance. Sleep deprivation provides, in essence, an operational definition of fatigue. Even so, most early sleep deprivation studies did not show significant performance decrement because of the nature of tasks that were employed and because of the relatively small sample sizes that were used. In many cases, no statistical tests were performed; and the purpose of early studies seemed to be simply to determine the gross limits of man's ability to stay awake! Unfortunately, the lack of an obvious endurance limit was interpreted to mean that, with moderate sleep deprivation, performance was not impaired to any practically relevant degree. During the early 1900's (when these studies were being performed), there were relatively few jobs which demanded the kind of performance which would be susceptible to fatigue, and fewer still in which performance decrement could result in catastrophe.

During the first experimental study of sleep deprivation, Patrick and Gilbert (164) observed a more severe form of blocking which they referred to as a "mental lapse." Three subjects, kept awake 90 hr, reportedly did not show any remarkable performance decrements until about the 72nd hour, at which time they began falling into involuntary micro-sleep periods after which their performance would immediately improve. Still, many early studies on sleep deprivation failed to find significant performance decline, until a minimum of 36 hr of sleep deprivation, and often a maximum 60-72 hr, in tests of: addition and multiplication, RT, learning, memory, hand steadiness and grip strength, ball tossing, tapping, reading letters, color naming, aiming, telegraphic code transmission, and the pursuit meter (Edwards: 64; Laslett: 131; Lee and Kleitman: 132; Kleitman: 121, 122; Robinson and Herrmann: 176).

In 1937, Warren and Clark (208) used addition and subtraction and color-naming tests (similar to those of Bills) to measure blocks, latency, and accuracy during 65 hr of sleep deprivation. Blocking was found to increase greatly after a period of prolonged sleeplessness, but the usually employed measures of error scores and RT showed no relationship to sleep loss. These researchers stated that, to measure the effects of prolonged sleeplessness, not only should tests be given at regular intervals and be of sufficient length to diminish the effects of compensation, but also that tests should be devised which would make compensation difficult or impossible.

Bjerner (30) used a serial RT test to measure blocks during 2 consecutive nights of sleep deprivation, and also recorded changes in alpha rhythm and pulse rate. He found that the longer RT's, defined as "blocks," were significantly associated with the disappearance of alpha activity and lowered pulse rate. He used the term "lapse" to describe the brief periods of sleeplike activity, and concluded that blocks lasting more than 5 sec were transient phenomena of the same nature as sleep.

In 1959, following Bjerner's approach, Williams, Lubin, and Goodnow (230) presented the results of 2 comprehensive studies of the effects of lapses generated from 72 hr and 98 hr of sleep loss on: various RT tests; auditory, visual and vibratory vigilance tests; and complex behaviors. The term "lapse," used interchangeably with Bills' term "block," was defined--for the RT test, as: any RT lasting longer than twice the baseline response; and, for the vigilance tasks, as: both errors of omission and commission. These researchers developed what became known as the Walter Reed lapse hypothesis: Fatigue causes increasing unevenness of performance, rather than gradual and continual decline, and is seen as brief periods of no response which increase in frequency and duration with continued mental work. Performance between these blocks will be maintained close to the initial level of response. Contrary to Bills' usage, their term of "lapse" (230) is usually meant to refer to an involuntary micro-sleep episode, as Bjerner (30) used it, brought on by excessive sleep loss. Following Broadbent (35), they also define a lapse as simply a "period of no response." In addition, they recognize that:

> "It may be more appropriate to regard lapses as periods during which a subject is less responsive to a stimulus, and in which the gap may be appropriately filled by a response which is highly practiced and less dependent on stimulus discrimination, or inappropriately filled by a competing response" (230: p. 23).

These data showed that sleep loss produced an increase in RT but that, even when extremely fatigued, on some trials the subjects are capable of coming very close to their best performance during baseline; however, a striking increase occurred in long RT (blocks). A progressively greater unevenness of performance was noted, as well as a reliable increase in the duration and frequency of lapses. While their tests (230) were classified as either Subject-paced (S-paced) or Experimenter-paced (E-paced), in all cases except for an addition test, the experimenter determined when the next problem would be presented. On S-paced tasks, performance between lapses could be maintained at close to normal level; and the task-set was such that subjects would sacrifice speed for accuracy, usually maintaining a baseline error rate. For E-paced tasks, impairment took the form of an increase in errors. RT was found to be an increasing monotonic function of task duration with increasing sleep loss; however, variations in duration and regularity of the RT foreperiod gave inconsistent results. Hence, Williams et al. (230) concluded that the critical factor for blocking appeared to be the total time, without interruption, that the same stimuli occur and the same response is required. Knowledge of results (KR) caused slight but inconsistent improvement. They noted that this effect was contrary to other findings (e.g., Mackworth: 137), and felt it might be explained for the S-paced RT task because the subjects received only their actual RT, with no additional exhortations to improve or compete with other subjects. However, competition and encouragement were applied during the KR condition for the visual and vibratory vigilance tasks, and still no conclusive results were obtained. Finally, since tasks involving three sense modalities and a large variety of tasks all showed lapses, they concluded that blocking was a central rather than a peripheral phenomenon.

Williams et al. (230) also recorded EEG alpha amplitude during the auditory vigilance task, and found a decline with increasing sleep loss as did Bjerner (30). In addition, errors of omission were consistently associated with less alpha than were other responses; and they concluded that, during lapses, the subject is in a state close to sleep. Further studies by this group (Williams et al.: 227; Williams and Lubin: 226) have borne out these findings.

The lapse hypothesis appeared useful in explaining most impairments due to sleep loss, except that of memory. Williams, Gieseking, and Lubin (228) demonstrated impairment of immediate recall of word lists due to 31 and 55 hr of sleep deprivation; but, since subjects were required to write down each word prior to attempting to memorize it, these researchers ruled out lapses as a possible cause of impairment. However, Polzella (172), using the short-term memory paradigm of Wickelgren and Norman (216), demonstrated that the occurrence of lapses prevented the encoding of items in short-term memory. He argued that lapses could have occurred after the subjects wrote down the stimulus word in the Williams et al. (228) study, and could account for that decrement also. Thus the lapse hypothesis seems to have a general explanatory power for most types of performance decrement due to fatigue.

R. T. Wilkinson, of the Applied Psychology Research Unit at Cambridge University, England, has carried out an exhaustive series of sleep deprivation studies (218-222) using the 5-choice serial RT test developed by Leonard (133). The apparatus consists of 5 stimulus lights and associated metal contacts. The subject holds a stylus and taps the contact corresponding to the lighted stimulus. This action extinguishes the light and another randomly illuminates. This self-paced procedure is usually continued for 30 min. The test accumulates three scores; errors (tapping a nonlighted stimuli); corrects; and gaps (scored every time a stimulus light remains on for more than 1.5 sec). Gaps are analogous to Bills' blocks, since the normal response time to each stimulus light is approximately 600 ms. In all experimental situations, gaps were found to be the most significant performance measure affected by sleep deprivation, even for periods as short as 30 hr. When 30-sec rest pauses were given every 5 min during the test, gaps still occurred even without sleep deprivation (133). Other researchers (notably Mackworth: 37) have shown that 5-minute rest pauses will, in vigilance tasks, reduce performance decrements due to fatigue. Even though a block lasting only 2 - 3 sec appears to dissipate fatigue in an ongoing task, a 30-sec rest pause is apparently not long enough to dispel any cumulative fatigue buildup.

In contrast to Williams et al. (230), Wilkinson (220) found that feedback of KR and motivational effects of competition interacted with 30 hr of sleep loss. KR reduced the production of gaps more during the no-sleep condition than the sleep condition. He explained this effect in the context of arousal theory (Malmo: 138) and in the inverted-U function which relates performance and arousal level (Duffy: 61). This study (Wilkinson: 220) was conducted over three 2-week periods and, even though subjects had sufficient time to recover between sleepless intervals, production of gaps increased over the 6 weeks. Thus, while practice may initially be found to reduce blocking, this measure seems sensitive to repeated fatigue periods. Each of 12 subjects maintained his relative ranking of performance impairment from one 2-week period to the next, indicating stable individual differences in susceptibility to fatigue. In one of the few studies of the aftereffects of 34 hr of sleep deprivation, Wilkinson (218) found that performance impairment could still be detected after 1 night of restorative sleep. On the day after sleep deprivation in the morning, but not the afternoon, gaps--but not errors or corrects--were significantly more frequent than at baseline. The aftereffect was evident early in the test and changed little with increasing time at task, instead of being more prevalent in the latter half of the tests (as is usually found by the direct effect of sleep deprivation).

Since Bills' response-blocking concept had been incorporated, it is not surprising that the 5-choice serial reaction test has been shown to be sensitive to stresses other than lack of sleep, such as: high noise level (Broadbent: 35); excessive warmth (Pepler: 167); anoxia (Bills: 25); drugs (Steinberg: 192); and alcohol (Wilkinson and Colquhoun: 225).

Murray (149) reviews the effects of sleep deprivation on personality, and theorizes that personality changes result from conflict between the drive to sleep and personal and social drives to stay awake due to task demands and factors in the laboratory setting. Task performance is then influenced by the balance of these motivational variables. Adjustment to this conflict generally results in a mood of apathetic depression in which the subject attempts to reduce social interaction and conserve energy. Frustration is created to the extent that he is forced to interact and perform because of situational pressures, and leads to aggression and irritability which is usually detrimental to task performance. To the extent that the subject uses repression to cope with the conflict, his denials of sleepiness may cause lack of correlation between subjective fatigue and performance measures. Note that this effect would be opposite the trend reported in earlier references: Subjects normally report increased feelings of fatigue before performance declines. In the present study, separate measures of fatigue and sleepiness were collected to determine if these constructs were correlated and which one was more highly correlated with performance and physiological changes.

Murray (149) points out that, although sleep deprivation has not been known to cause true schizophrenia, some individuals appear to have traits which predispose them to a transient psychotic state during sleep deprivation. Morris, Williams, and Lubin (147) found that two groups of highly motivated subjects who experienced 72 and 98 hr of sleep deprivation, respectively, tended to deny their sleepiness but emphasize their fatigue. They became listless and apathetic and, while they could be aroused by external stimulation, they avoided tasks and games requiring attention or concentration. As sleep loss increased, brief intermittent pauses or lapses in ongoing behavior increased in frequency, duration, and depth. Transient perceptual and cognitive changes occurred which resembled a mild psychotic state and generally coincided with the occurrence of a lapse. Self-report of visual misperceptions, ranging from illusions to hallucinations, significantly increased with sleep loss. A significant correlation was noted between reporting dreams during lapses, and later experiencing visual hallucinations. Temporal and cognitive disorganization significantly increased with sleep loss. Also noted were changes in memory, perceptual thresholds, concept of self, and speech. All of these factors would have detrimental effects on flying performance.

Air Force pilots are often required to fly when they normally would be sleeping, and moderate sleep loss can be expected to occur occasionally, even during routine flying operations. From the range of debilitating consequences presented in the foregoing literature review, we concluded that the most common behavioral effect of sleep deprivation is that of lapses, or response blocks. This finding thus supports the use of a response block measure to index flight-crew performance decrement.

Physiological Cost Indices of Fatigue

The basic physiological approach to fatigue involves the measurement of energy expended in performing a given amount of work. Waller and De Decker (206) were able to relate increases in carbon dioxide production to reduction in work output during a night's work. They used the term "physiological cost" to describe the increased metabolic demand resulting from increased fatigue and consequent lowered performance. Page (161,162) suggested that the concept of fatigue be replaced with the concept of metabolic cost. Physiological measures employed were usually heart rate, oxygen consumption, and carbon dioxide production. Unfortunately, such early researchers found these measures insensitive to any behavior in which gross muscular work is minimal, and thus unsuited for measuring "cerebral work" or mental activity. In an attempt to unify two of the three classes of fatigue measurement, Bitterman (29) suggested that fatigue be defined as "reduced efficiency resulting from continued work, and reversible by rest." Efficiency was defined as "the ratio of performance to expended effort"; effort was to be determined from metabolic cost indices. However, this definition has not resulted in a practical methodology or follow-on research to establish objective measures of efficiency.

Physiological cost concepts are related to Selye's concept (188) of the General Adaption Syndrome, in which any stress to which the body is exposed creates an overall nonspecific systemic reaction to cope with or reduce the stress. As has been demonstrated (188), some definite stress reactions generate physiological patterns seen in many diseases. Tissue trauma (such as hemorrhage, burns, and frostbite) and environmental variables (such as heat, physical exhaustion, bed rest, drugs, hypoxia, and diet) also elicit well-established physiological responses. Selye has theorized that fatigue also creates a stressful condition to which the body tries to adapt--and thus produces an abnormal set of physiological indicants which could be used as an index of the severity of the fatigue stressor (188).

After having reviewed several fatigue studies showing no significant or dramatic performance decrement and one study with a performance increase, Cameron (40) concludes that performance measures are too erratic and unreliable to serve as indicators of fatigue. He states that "fatigue" should be used as no more than a descriptive term for a generalized stress response over a period of time. In fact, he notes that the only unique aspect of fatigue is time. He feels that the best index of the acute and chronic fatigue effects would be the time required for biologic emergency mechanisms to return to a normal arousal level. However, how this time would be measured is not clear; for he states that "indirect measures such as physiological indicators of activation level are likely to be inconclusive because they are not specific to fatigue." On the other hand, he admits that: "It may, of course, be necessary to employ physiological measures in order to determine the point in time at which a normal homeostatic equilibrium is achieved, and recovery may be said to have occurred" (40: p. 646).

Along the lines of Cameron (40), Harris and O'Hanlon (86) have reviewed the recovery of man from exposure to such adverse conditions as sleep deprivation, abnormal work-rest cycles, prolonged physical work, and environmental and situational stressors. Their purpose was to determine if recovery functions could predict both how long a man could maintain effective performance before he had to be relieved during continuous military operations, and how long a rest period would be required before he would again be ready to perform effectively. They concluded that, although such predictions cannot be made on the basis of the knowledge now available, the following list of potential physiological failures seems most important to consider--and reversal of these impairments may provide practical indications that recovery has taken place. This list includes: degraded physical working capacity, inadequate iron reclamation, myocardial "fatigue," paroxymal cerebral cortical activity, impaired carbohydrate metabolism, thiamine deficiency, involuntary hypohydration, glycogen exhaustion, increased susceptibility to infection, imbalanced protein metabolism, and adrenal cortical and medullary exhaustion. Harris and O'Hanlon (86) also feel that changes due to fatigue will become apparent in the physiological systems before performance degrades. Their implication is that, even though a given schedule of work has not yet produced performance decrement, work-rest cycles should be so structured that severe changes in the physiological systems would be prevented. They also note, however, in their recommendations for future physiological research, that every attempt should be made to relate any physiological changes observed to changes in task performance and in subjective feelings.

The physiological cost of fatigue is generally not an immediate problem for the individual, provided he receives sufficient recovery time. This cost is fatigue's only operational consequence except for that of performance decrement. Gartner and Murphy (80) cite 4 difficulties in using physiological indicants for fatigue:

1. Response generality. Measures are not specific to various states conceptionalized as fatigue or anxiety (or any other hypothetical construct).

2. Response patterning. Patterned physiological reactions are often stimulus-specific and/or peculiar to an individual.

3. Response intensity. Poor correspondence between the intensity of psychological (behavioral) reactions and associated physiological changes.

4. Temporal relationships. Some physiological responses occur almost instantaneously (e.g., heart rate, GSR), whereas others can be observed only after hours, or days (e.g., excretion of urinary metabolites) (80: p. 32).

However, physiological indicants offer a major advantage in their promise as unobtrusive and objective measures which could be used in operational situations. Due to the complexity of the problem, a multidimensional approach is necessary to provide a better understanding of the fatigue concept.

According to Hartman and Cantrell (91), the best approach to maintaining man's capacity for skillful work is to engineer the system so that physiological degradation is eliminated. This approach implies that, if physiological indicators known to be associated with stress reactions are found to be within normal limits, then it is presumed no performance decrement of operational consequence has occurred. The problem is to so quantify these physiological limits in relation to a criteria of performance degradation that system managers will design, man, and use the system in such a way that the limits are not exceeded.

The physiological cost concept has most recently been applied in a series of studies, by the USAF School of Aerospace Medicine, attempting to relate urinary metabolites, catecholamines, and steroids to the duration and stress factors of Air Force flying operations. Hale et al. (43) viewed physiological and psychological responses to long-duration missions as biologically incurred costs which must be repaid during a fixed, minimal recovery period. They found the urine battery sensitive to double crew missions, in C-5 aircraft, lasting up to 65 hr. (For a review of work in this area, consult Hartman et al.: 95; and Johnson and Naitoh: 109.)

Urinalysis as a Method for Indexing Fatigue

Of the physiological measures which have been related to fatigue and flying performance, the present study has investigated three: urine components, cardiac activity, and body temperature. (Temperature is discussed in the report section on "The Circadian Rhythm Hypothesis.")

F. N. Dukes-Dubos (62) has prepared a summary of the problems arising from individual variability, explanations for the seeming lack of consistent relationship between urinary components and fatigue, and a broad review of the many substances analyzed for their possible relationship to fatigue. In the present study, the urine components analyzed were: sodium (Na); potassium (K); nitrogen (N); urea; steroids or 17-hydroxycorticosteroids (17-OHCS); and catecholamines (epinephrine and norepinephrine). The first four serve as crude indices of general metabolic activity--urea being specifically related to protein metabolism.

The following general review of 17-OHCS and catecholamines is taken from Guyton (83). Steroids (cortisol) are secreted into the blood stream from the adrenal cortex in response to a wide variety of stresses, and enable the body to cope with stress through its effects on carbohydrate, fat, and protein metabolism. Moreover, steroids cause a stimulation of gluconeogenesis by the liver and a decrease in glucose utilization by the cells, thus (in turn) raising the blood glucose concentration. At the same time, the secretion causes a reduction in protein stores in all parts of the body except the liver. Blood amino acid concentration goes up, transport of amino acids into extra hepatic cells is diminished, and transport of amino acids to the liver is enhanced. Amino acids are thus mobilized from the tissues to the liver. Finally, fatty acids are brought out of adipose tissue, increasing their blood concentration and (in turn) their utilization for energy. The adrenal cortex secretes steroids in response to adrenocorticotropic hormones from the addenohypophysis which is under direct control of the hypothalamus. With this indirect feedback mechanism, levels of cortisol can continue to rise to very high blood concentrations as long as the stress agent continues to stimulate the hypothalamus in some way. Cortisol fixes to its target tissues in about 20 min after release. The normal blood concentration is about 12 μg/100 ml; its half-life in the blood is 100 min. The normal secretory rate is 15 mg/day, of which approximately 75% is excreted in the urine.

The urinary circadian rhythm for 17-OHCS has been established and found to lag the concentration rhythm in plasma by about 2 hr (Migeon et al.: 143). The maximum value was found during the later hours of sleep, usually around 0600 hr, after which the urinary concentration fell until 2400 hr. The concentration then rose rapidly between 0200 and 0600 hr. In this same study, nurses and watchmen who had worked a schedule of 5 nights a week for at least 6 months were found to have the same general rhythm, thus indicating the rhythm to be extremely stable. Increases in 17-OHCS excretion have been found for various anxiety-producing situations, electro-shock treatment, hallucinogenic drugs (Bliss et al.: 32), and mildly stressful motion pictures (Wadeson et al.: 45). Berkun et al. (19) performed an extensive series of experiments simulating five stressful military situations in which the subject was led to believe that he was in immediate danger of losing his life or being seriously injured, or that by his actions he had seriously injured one of his coworkers. All stress situations elevated 17-OHCS excretion, and the level of increase could be rank-ordered according to the level of stress presumed to be induced for each situation.

Miller (144) provides a review of the many studies in which 17-OHCS has been found to increase due to the stress of military flying. In 1943, Pincus and Hoagland (170) conducted three sets of experiments which related steroid excretion and flying stress. In the first set, subjects performed for up to 6 hr on the Coordinated Serial Reaction Apparatus in which aircraft-type controls were used to "fly" a beam of light, along a prescribed pathway, to extinguish 1 of 5 randomly illuminated photocells. A self-paced trial consisted of extinguishing 50 lights, a procedure which usually took about 40 sec. A composite score consisted of total time to complete a trial and number of deviations from the pathway. The score increased with time on task, and subjects showed significantly increased steroid production with time as compared to a time-matched control group. In addition, individual scores were found to be positively correlated with the level of steroid increase.

In the second set of experiments, the apparatus was modified so that the target was a randomly moving airplane model which had to be tracked; this experiment was supposed to simulate flying in formation. As an added stressor, the tests were carried out in an altitude chamber creating partial pressures of oxygen from 21% to 12.6%, simulating altitudes from sea level to 13,000 ft (3,960 m). Results were similar to the first condition except that, from 5,000 ft (1,524 m), performance decreased more rapidly with increasing altitude, and steroid production was increased even more.

In the third set of experiments, urine samples were collected from 16 instructor pilots before and after 152 flights which lasted, respectively, from 1 to 4 hr. Increases in steroid production were found to be related to the amount of time the pilot was airborne. Independent ratings, given by the pilots' squadron commander on their individual susceptibility to fatigue, were also found correlated to absolute steroid production. When these steroid data were compared to those of 7 test pilots over 56 flights, the test pilots were found to secrete more steroids by a constant amount than the instructor pilots, regardless of length of flight. This finding was interpreted to mean that the test-pilot flights were more stressful. Similar results, while not always as dramatic and clear-cut, have been found in later studies; for example, 20-hr missions in B-52 training flights (Marchbanks: 140).

Some evidence shows that a maximum of 120 hr of sleep deprivation with minimal workload, while admittedly stressful, does not cause increased adrenocortical activation (Kollar et al.: 125). A followup study concluded that 205 hr of sleep deprivation resulted only in mild, if any, excess secretion of either steroids or catecholamines (Rubin et al.: 181).

The relationship between the adrenal medulla and threatening situations was first demonstrated by Cannon and de la Paz (43). Catecholamines are secreted by the adrenal medulla in response to stimulation from the sympathetic nervous system. While proportions can vary depending on physiological conditions, the average secretions are 75% epinephrine and 25% norepinephrine. The effects of catecholamines on the body are the same as those caused by direct stimulation of the sympathetic nervous system; but the effects last about 10 times longer, because circulating catecholamines are only slowly removed from the blood. The sympathetic nerve endings secrete norepinephrine; but, in a matter of seconds, it is reabsorbed or destroyed at the cellular level by O-methyl transferase or monoamine oxidase. These enzymes are similar to cholinesterase which destroys acetycholine, the agent secreted by the parasympathetic nervous system. Also, while both sympathetic nervous system and the secretions of the adrenal medulla have general nonspecific effects, the catecholamines stimulate and increase the metabolic rate of every cell in the body. However, circulating catecholamines do not readily pass the blood-brain barrier (Rothballer: 179). Stimulation of the sympathetic nervous system mobilizes the body for action. Norepinephrine causes general vasoconstriction, increased cardiac activity, increased basal metabolism, sweating, inhibition of the gastrointestinal tract, glucose release from the liver, decreased kidney output, and adrenocortical secretion. Epinephrine, which has similar effects, has a greater stimulating effect on cardiac activity and basal metabolism and less constricting effect on blood vessels in muscles. The normal resting secretion rates are 0.2 μg/kg of body weight per minute for epinephrine, and 0.07 μg/kg per minute for norepinephrine.

According to some indications, catecholamines are excreted due to stress; but they are generally released in relation to the overall activity level (Pátkai: 163) and performance level (Frankenhaeuser et al.: 76). In a review of catecholamine response to various activities, Euler (68) reports that *mental stress* associated with anger, aggression, or exhilaration will increase *norepinephrine* excretion--while *emotional states* characterized by apprehension, discomfort, or painful or unpleasant feelings will increase *epinephrine* excretion.

Fiorica et al. (see "Author's Note," below) found that 84 hr of sleep deprivation caused an elevation in rectal temperature but that catecholamine excretions were elevated only at night relative to a control group, thus only reflecting differences in activity levels. During the day, when their activity levels were equal, the catecholamine excretions were about the same. They concluded that lack of sleep in itself does not elicit increases in adrenal medullary activity. They measured psychomotor performance every 4 hr on the "Kugelsmachine," which is supposed to estimate attention, visual estimation, decision-making, and manual skill. Subjects were required to select steel balls of various sizes and place them into the correspondingly sized hole of a continuously rotating cylinder. The test lasted 10 min. Clear differences from the control group were observed starting from the second day of sleep deprivation.

Euler and Lundberg (70) found that urinary epinephrine levels were elevated in: pilots, during 1-hr moderately stressful flights; and in inexperienced passengers, in flights lasting 1.5 hr. This effect was also noted for norepinephrine in the pilots, but *not* in the passengers.

Melton and Fiorica (142) found that epinephrine and norepinephrine excretions were elevated in private pilots with less than 100-hr flying experience on cross-country flights lasting 3.5 hr and 9 hr, but that the level of excretion was not related to the length of the flight.

In a recent study by Krahenbuhl, Marrett, and King (126), various phases of Air Force flying training in the T-37 jet aircraft were examined for their effect on catecholamine production. The emergency procedures phase given in a Link Trainer was essentially nonstressful, but both epinephrine and

---

AUTHOR'S NOTE: For detailed information, please consult--Fiorica, V., E. Higgins, M. Lategola, A. Davis, Jr., and P. Iampietro. Physiological responses of men during sleep deprivation. FAA AM 70-8. Oklahoma City, Okla.: FAA Civil Aeromedical Institute, May 1970.

norepinephrine were significantly elevated from control values during spin, solo, and check flights. Epinephrine appeared to be more responsive than norepinephrine. Performance ratings during spin maneuvers, which were found most stressful, appeared to follow an inverted-U relationship with epinephrine excretion. That is, the flights appeared to evoke an optimal stress reaction, in terms of catecholamine response, which was related to maximum performance.

According to a study by Fröberg et al. (78)--in which subjective fatigue, shooting performance, and catecholamines were measured during 75 hr of sleep deprivation--epinephrine excretion reached its peak in the afternoon, and norepinephrine in the early morning hours. Furthermore, self-ratings of fatigue were negatively correlated with epinephrine, and positively correlated with norepinephrine. Shooting performance showed high negative correlation, with fatigue ratings and norepinephrine; and positive correlation, with epinephrine. Correlations increased even more when psychological variables were adjusted for the lag in catecholamine excretion.

No functional relationship appears to exist between the adrenal medulla and adrenal cortex. However, Broverman et al. (50) have attempted to differentiate the effects of short- vs. long-term stress on two classes of behavior due to the interaction of catecholamine and steroid effects on the body. Short-term stress is hypothesized to: facilitate performance on serially repetitive, overlearned tasks; and impair performance on novel tasks requiring perceptual restructuring. Long-term stress is hypothesized to have the opposite effects. Their review of the literature gives some support to this contention (50). They attempt to account for these findings by arguing that, during short-term stress, behavior is dominated by the sympathetic nervous system. With increasing exposure of the central nervous system to the stress-elicited adrenal hormones, however, dominance shifts to the parasympathetic system and thus causes an overall depression of activity.

The following mechanism is proposed to cause this shift. Because of the blood-brain barrier, levels of norepinephrine in the central nervous system are limited to those within the adrenergic neurons. Levels of norepinephrine are controlled in part by the intraneural metabolic activity of the enzyme monoamine oxidase. Chronic elevations of monoamine oxidase diminish neural stores of norepinephrine, but suppressed monoamine oxidase activity allows the neural accumulation of "false neurotransmitter" monoamines which displace norepinephrine and then act as inefficient neurotransmitters. According to some evidence, 17-OHCS suppresses monoamine oxidase activity (50). Since 17-OHCS readily penetrates the blood-brain barrier, and circulating norepinephrine does not, long-term stress is hypothesized to cause not only an increase of central false neurotransmitters but also a net depression of sympathetic activity leading to enhanced parasympathetic system dominance.

Because of the extensive research history involving flying stress and urinalysis, and because of the potential explanatory power of this type of measure, the author decided to collect urine samples throughout the present study. An attempt was made to relate changes in this measure to changes in flying performance, subjective fatigue, and other physiological data.

Heart Rate and Heart Rate Variability Research

The cardiac activity indices of heart rate (HR) and heart rate variability (HRV) have been used extensively to analyze pilot activity in flight, because the data can thus be collected without interfering with flight activities. This technique has an advantage over urinalysis in that data reduction can potentially be performed much more quickly for HR than for urine, because lengthy chemical analysis is not required. In addition, HR can be measured for specific segments of performance during relatively short time spans. In no way can the relative contribution of any segment of behavior during the urine collection period be precisely determined; and since these segments must, for practical reasons, extend over periods of 3 - 4 hr, urine analysis is confined to relatively gross estimates of when performance decrement has occurred. Finally, research on HR and HRV has proved to be more closely related to activity levels and performance quality than was urine analysis.

Following is a review of research in which cardiac data have been found to be related to specific phases of flight and pilot workload. Ruffell Smith (182) found that civilian airline pilots' HR increased 5% - 10% above baseline during preflight checks, and about 50% during takeoffs. During approach and landing, HR increased but was much more variable, depending on the particular difficulty of the flight. During a flight, HR also increased to a varying extent, depending on perceived hazards and stresses encountered.

Roman et al. (177) investigated the relation of HR and landing errors for cockpits with restricted fields of view. While landing quality was not reduced for even the narrowest fields of view used, HR was found to be inversely correlated with landing performance. Hasbrook and Rasmussen (96) found significant HR increases for each of 10 simulated instrument approaches flown by 10 pilots. However, the overall mean HR level decreased 11 beats per minutes (bpm) on successive approaches, thus indicating the pilots were adapting to the stress. The mean HR changes were consistent for all approaches, increasing 5.2 bpm per approach. The variation within approaches averaged 7.9 bpm.

In a study by Bateman et al. (16), the HR's for commercial pilots--on routine flights, upgrade training flights, and simulator flights--were found to be very similar and higher than resting rates. However, basic training flights were found to be significantly higher still. HR increased, not only when pilots were subjected to specific inflight stresses but also when they were demonstrating maneuvers requiring a high degree of skill.

The concept of additional HR (HR increases not accompanied by corresponding increases in oxygen consumption) has been found to be a reliable and valid indication of "psychological activation" during flight (Blix, Stromme, and Ursin: 33). More sophisticated HR analysis techniques have been used to differentiate the workload of a pilot during several phases of flight. Opmeer and Krol (157) found that increases in HR and decreases in a measure of beat-to-beat variability matched the predicted order of increasing difficulty of four phases of flight: baseline, level flight, takeoff, and approach. When pilots were required to fly realistic flight plans in a simulator, the same relative increases were found. HRV was found to be a more sensitive measure than HR; and Opmeer and Krol (157) concluded that HRV was more related to cognitive tasks, where HR alone was more responsive to anxiety-inducing tasks.

Roscoe (178) has demonstated that HR is a useful tool in evaluating pilot workload changes created by new aircraft instrumentation, such as automatic landing devices, heads up displays, and advanced control systems. HR was found to vary as changes in weather conditions and different runways created more stressful landings. Inflight HR patterns were found to be more consistent than preflight baseline measures, due to the relative uniformity of the tasks and environmental stimulation.

While inflight cardiac indices have yielded some information on cognitive workload and stress levels experienced by pilots, laboratory studies in which the stimulus presentations can be more precisely controlled have been much more successful in relating these indices to performance and workload. The normal resting HR exhibits a relatively large degree of beat-to-beat irregularity referred to as sinus arrhythmia. Most of the variation during sinus arrhythmia is attributed to the cyclical effects of respiration on cardiac activity. In the present study, HRV is used to refer to beat-to-beat variability which is not solely a function of sinus arrhythmia.

Generally, HRV is found to be more sensitive to cognitive workload than HR. For example, HRV, but not HR, was found to differentiate rest periods from task periods during mirror tracing (Obrist, Hallman, and Wood: 156). However, Kalsbeek (112) reports that increased attention and information processing (or mental load) greatly reduces this irregularity without appreciably changing the mean HR. He points out that static and dynamic physical work also reduces irregularity, but increases HR. In studies where physical work was varied with mental load, HRV was not able to differentiate increases in mental load when the HR rose above approximately 130 bpm. However, respiration rate was found to vary with mental load up to at least 155 bpm.

Ettema and Zielhuis (54) found that sinus arrhythmia was significantly depressed, and HR, blood pressure, and respiration rate were significantly increased as the rate of presentation for a binary choice RT task increased up to 50 signals/min. They concluded that the effect was due not only to a change in the breathing pattern but also to a rise in vagal tone and sympathetic nervous activity induced by the mental workload. Boyce (34) found essentially the same increase for HR and decrease in HRV for increasing mental load from a single to a double digit mental arithmetic task. However, as the physical exertion required to provide an answer increased, both HR and HRV also increased. The author attributed this effect to the method used in scoring HRV, since the steady increase in HR during the physical task artificially increased the variability about the mean. A series of studies by Thackray has shown HRV to be a useful measure for separating mental work from rest periods (197-199). For two-dimensional compensatory pursuit tracking, HRV, HR, blink rate, respiration rate, respiration period variability, and skin conductance were capable of differentiating rests from work periods (Thackray: 197). For a simulated radar control task, HRV was found to be higher for subjects reporting high boredom (Thackray, Bailey, and Touchstone: 198). In addition, the performance of the subjects in the higher boredom group also significantly declined over the test period. This finding suggests that HRV reflects a level of attentiveness which is related to overall performance capability.

For the group taken as a whole, neither performance nor HRV changed during the 1-hr task.

Using a 4-choice serial reaction test, Thackray, Jones, and Touchstone (199) found that HRV increased, HR decreased, and response variability and errors increased over a 40-min test; however, mean RT did not change. HRV was significantly correlated with response variability during the last 4-min period of the test. HR showed no correlation at all. When the groups were separated into introverts and extroverts by means of the Eysenck Personality Inventory, extroverts but not introverts showed a significant continuous rise in response variability and a significant correlation with their HRV score.

A more comprehensive view of relationship between HR indices and performance has been stated in the broader framework of arousal theory. Since the early work of Yerkes and Dodson (233), the level of performance quality has been found to be related to the degree of arousal or activation level of the organism by an inverted U-shaped function. This relationship implies that an optimal level of activation produces maximum performance. (For a review, see Kahneman: 110.) Duffy (61) has proposed two basic dimensions for behavior: activation level and direction. The activation level is usually defined in terms of physiological measures, such as cortical activity, muscle potentials, cardiac and respiration rates, and skin conductance. Levels of activation, especially the sleep-wakefulness dimension, are viewed as being mediated through the reticular activating system located in the brain stem (Hebb: 98; Lindsley: 134; and Malmo: 38).

While much research has demonstrated the existence of a generalized pattern of activation, Lacey (129) has reviewed many studies in which physiological indices of activation were disassociated from each other during certain situations of high behavioral arousal. Lacey has termed the change of one or more physiological indicators, in a direction opposite to that predicted by arousal theory, as: "directional fractionation." His studies have shown repeatedly that attentive observation of external stimuli causes cardiac deceleration, while other indicators are showing an increase in sympathetic activity as would be expected by the arousing nature of the situation. Environmental rejection, such as would occur to aversive stimuli or during processing of internalized information, would be predicted to cause cardiac acceleration. Lacey has hypothesized that the decreased HR accompanying attentive observation facilitates the intake of information due to the visceral afferent negative feedback from heart to brain. Obrist and his colleagues have recognized the existence of directional fractionation, but do not admit any causal relationship between HR deceleration and facilitation of the attentive response (Obrist et al.: 155, Obrist, Hallman, and Wood: 156). Their position has been referred to as: the Cardiac-Somatic hypothesis.

Much of the support for Lacey's position comes from RT studies in which cardiac deceleration is observed during the foreperiod. Evidence indicates that the degree of deceleration observed is related to the speed of RT (Lacey and Lacey: 128, p. 548). Eason and Dudley (57) demonstrated that heart rate, skin conductance, evoked cortical potentials, and muscular tension simultaneously reflected both generalized arousal and directional fractionation. Arousal level was manipulated by having subjects perform a simple visual RT task with and without threat of shock, or just passively observe the stimuli. HR was found to decelerate during the foreperiod in all conditions; but, since HR was highest during the shock threat condition, when the subjects' attention was presumed highest, the authors concluded that the cardiac deceleration-attention hypothesis was not supported. Lacey would not predict that overall HR would be lower in the high-attention condition, but only that the magnitude of deceleration would be greatest. Unfortunately, no direct test existed for magnitude of deceleration changes between conditions; but, from inspection of the data presented, the magnitude appears greatest for the high arousal condition.

Kahneman et al. (111) found evidence of directional fractionation, but their data appear to be in opposition to the cardiac deceleration-attention hypothesis. Cardiac acceleration occurred during the information-intake stage of a mental arithmetic task, and deceleration occurred during the information-processing stage of the task.

Dahl and Spence (54) had several tasks independently rated for their complexity along an information-processing dimension, according to Bergum s (18) taxonomic analysis of performance; both the direction of the subjects' attention, inward or outward, and the degree of required concentration were rated according to Lacey's analysis (129). The task-demand characteristics were considered an operational definition of level of activation. Increasing task demand and concentration level were highly correlated with mean HR. The direction of attention required by the task was not. However, the results are in line with previously reported research (129) showing an increase in HR with mental load; for both task-demand characteristics and concentration can be conceptualized as components of mental load.

Gaillard and Trumbo (79) measured HR and HRV for a serial reaction test in which task difficulty was manipulated by using either a fixed or variable foreperiod during a 3-hr test session. Subjective activation level was either increased with a stimulant (amphetamine) or depressed with a sedative (barbiturate). In the high arousal condition, HR increased and HRV decreased due to amphetamine; however, in the low arousal condition the barbiturate caused HRV to increase, but HR increased also. This effect may be explained by the fact that, as subjects became drowsy, the HRV increased; but, as they attempted to fight off the depressive effect, the extra effort caused HR increase. Mental effort was assumed maximal during the variable foreperiod tests; and, in this condition, HRV was reduced. HR was higher for variable foreperiod tests but only in the low arousal condition, thus indicating a greater increase in effort. Generally, HR decreased and HRV increased during the 3 hr at the task, thus indicating adaptation to the situation and possibly a loss of concentration. This pattern of autonomic activity is indicative of a loss of activation level. Unfortunately, no attempt was made to relate performance to HR changes.

Porges (173), using both fixed and variable foreperiod RT task conditions, had subjects either respond or passively observe the RT stimuli. He measured both HR and HRV during four phases of the test: prior to trial onset, beginning of foreperiod, end of foreperiod, and during the response. In the passive control group, no changes occurred in either HR measure. In the RT groups, HR and HRV increased following the RT signals, and HRV decreased in anticipation of the response phase. In the fixed foreperiod RT group, neither

HR measure was related to RT; but, for the variable foreperiod RT group, subjects showing a greater mean reduction in HRV from pretrial to preresponse phases had faster mean RTs. HR decreased and RT significantly improved over the 10 trials, but were not correlated with each other.

During increasing levels of fatigue, HR can be expected to decrease as the subject's level of arousal falls--or to increase as he puts forth extra effort to stay awake. The paradoxical increase in HR with fatigue normally occurs with physical exertion. HR will continue to increase under vigorous exercise, up to the point of collapse. Thus, the task demands must be taken into account in predicting the arousal level of a long-duration flight. Most sleep deprivation studies have found increases in HR when even moderate task demands are made on the subject. Malmo and Surwillo (138) found HR increases from the first half to the second half of a 60-hr vigil for 2 out of 3 subjects, and concluded that sleep deprivation had the effect of raising the level of activation. Corcoran (52) attempted to separate the concept of arousal from effort by requiring minimal activity from subjects during a 60-hr period without sleep. In this case, both HR and performance on an unarousing, nonphysical, 30-min vigilance task fell consistently. Corcoran argued that performance would follow the inverted-U with decreasing arousal, and that arousal would fall with lack of sleep or increased fatigue; but the effort to remain awake, which was what was being measured by physiological indicators, would be a function of task demand and subjective motivation to remain awake.

In a 60-hr sleep-deprivation study without structured activities, HR was found to increase up to about 36 hr of sleep deprivation, and to fall thereafter (Fenz and Graig: 71). This change tended to reflect a change in the subject's desire to remain awake, since the authors report that: "while during the early stages of sleep-loss the effort was mainly on the part of the *S* to keep awake, in most cases it was the effort on the part of the *E* to keep *S* awake during the last hours of deprivation" (71: p. 553).

Bergström (17) compared HR changes during a 75-hr vigil to that of a control group, and found that the mean HR generally decreased with sleep deprivation but increased somewhat for the controls. During a shock-threat tracking task, HR rose for both groups, but the increase was significantly higher in the control group. Either the threat was not perceived as arousing during sleep deprivation as during normal conditions, or the subject's cardiac response capability was reduced due to the physiological cost of sleep deprivation.

From the foregoing research, the following changes in HR and HRV have been predicted for long-duration flights. First, HR and HRV would tend to increase with moderate levels of fatigue. With very high levels of fatigue, HR would be expected to fall and HRV to increase still further. Also, tasks creating greater levels of arousal, due to their complexity or concentration, may be initially more resistant to fatigue effects. Accordingly, it can be hypothesized that straight and level periods of flight requiring minimal control input and instrument monitoring should show greater performance decrement with fatigue than periods when maneuvers must be performed. Thus a task complexity by time-at-task interaction would be predicted. The higher the arousal value of the task, HR should be higher and HRV should be lower. Tasks requiring maximum levels of information processing and concentration should show least performance decrement, greatest HR increases, and greatest HRV decreases. To the extent that the subject cannot concentrate due to extreme levels of fatigue, arousal level will be relatively lower; and performance decrements would probably be correlated with increases in HRV.

The Circadian Rhythm Hypothesis

Most, if not all, biological processes are rhythmic or cyclical in nature; and these rhythms in turn influence other processes, both physiological and behavioral. A comprehensive overview of the general nature of biological rhythms and mathematical descriptions of them based on cybernetics is given by Sollberger (190). While most circadian rhythm research has been carried out on infrahuman species, the existence of circadian rhythms in man is well documented. Conroy and Mills (51) reviewed human research confirming the existence of diurnal rhythms for temperature, endocrine systems, the kidney, the cardiovascular system, hematopoiesis, respiration, metabolism, electrolytes, the digestive tract, the introvascular pressure and pupil size of the eye, and cerebral function.

The circadian performance hypothesis states that the quality of performance is, in part, a function of the time of day during which it is measured. Generally, performance has been found to improve rather steadily from morning to afternoon or early evening, then plateau, and next decline, reaching its lowest level around 0400 hr. Thorough summaries of research in this area have been made by Colquhoun (50), Hockey and Colquhoun (62), and Kleitman (121).

This cycling of performance is one of the reasons for difficulty in demonstrating the effects of fatigue on performance. Because of the more or less naturally occurring rise and fall in performance, early studies gave paradoxical results depending on when performance was measured. In fatigue studies lasting more than 24 hr, performance would improve after reaching its low point early in the morning, even though no rest had been allowed. This performance cycling increased variability and, if not taken into account in the experimental design, made difficult the detection of significant decrements in performance. Coupled with the fact that small sample sizes were usually employed, early research in this area, like that of fatigue, was somewhat inconclusive. For example, Freeman and Hovland (77) reviewed early studies of daily muscular and mental work output and found that these studies showed either a continuous rise, a continuous fall, a morning rise and afternoon fall, or a morning fall and afternoon rise. While they recognized diurnal variations in performance, early researchers were not able to detect a consistent pattern to the cycle.

The physiological rhythm most closely associated with performance cycles is body temperature. While temperature regulation is critical for warm-blooded mammals, diurnal cycling has long been recognized. As early as 1845, Davy (56) had recorded the daily rise and fall in human body temperature. While the time of the occurrence of the precise minimum and maximum varies from individual to individual, within an individual both the phase and amplitude of the cycle is usually consistent (Conray and Mills: 52). The minimum is usually around 0400 hr and the maximum, somewhat more variable, occurs between 1700 and 2100 hr. However, Kleitman (121) gives examples of individuals with extreme deviations from this periodicity.

One of the first clear demonstrations of a relation between the circadian rhythm for both simple and choice RT and body temperature was by Kleitman, Titelbaum, and Feiveson (124). The results of this study led Kleitman to conclude "that there is probably no RT curve independent of temperature" (121: p. 154). One of the tests they used was a modified color-naming apparatus which measured long pauses in RT as defined by Bills. Additional verification of the blocking phenomenon was provided by this apparatus; and blocking was found to exhibit a cycling effect, being lowest in the afternoon and highest in the evening, with morning tests being in between.

Just as extreme fatigue has been found to cause an overall reduction in performance, sleep deprivation has also been found to generally depress body temperature. Unless a new sleep cycle is initiated, the diurnal rhythm will be maintained during the sleepless period. Body temperature for 5 subjects, measured at 1100 hr every day, was found to fall from a mean of 97.6° F to 96.3° F (36.4° C to 35.7° C) after 123 hr of sleep deprivation; however, after only 72 hr, it appeared slightly elevated (Ax and Luby: 8). Since only one measurement was taken a day, it could not be determined if the rise was simply due to a circadian shift. Eight subjects underwent 87 hr of sleep deprivation with continuous monitoring of rectal temperature, and a consistently lower temperature from baseline, 0.5° to 0.7° F (0.9° to 1.3° C) was found (Krieder: 127).

Murray, Williams, and Lubin (150) collected oral temperatures and 4-point subjective fatigue scale data 10 times every 24 hr from 15 subjects over a 97-hr sleep deprivation period. Body temperature, although maintaining a persistent diurnal rhythm, showed an overall decline. The subjective fatigue ratings also showed a daily rhythm, but generally increased with sleep deprivation. Body temperature and fatigue ratings were both significantly correlated with sleep loss. However, when sleep loss was held constant, the partial correlation between subjective fatigue and body temperature still showed a strong inverse relationship (-.76). The group mean temperatures for the four successive 24-hr intervals, analyzed by Kleitman (122), were 98.16°, 97.76°, 97.71°, and 97.68° F (36.76°, 36.53°, 36.51°, and 36.49° C). The drop of 0.39° F (0.18° C), from the first interval to the second, was found statistically significant. Likewise, the increase of group self-ratings of fatigue, from the first 24-hr interval to the second and from the second to the third, were also significant.

Harris et al. (87) found that the oral temperature rhythm was lower during a 54-hr mission in a C-141 aircraft than postflight. Double crews worked on schedules of either 4-hr on, 4 off; or, 16-hr on, 16 off. The 4-on, 4-off work-rest cycle had the greater depressing effect on temperature, possibly indicating that the cycle was the more fatiguing. Moreover, crew members occupying key positions had lowest temperatures both during flights and during recovery periods.

Kleitman and Jackson (123) investigated the relationship between oral temperature and performance in 9 subjects housed on a large Navy ship following 8 different work-rest cycles over a 5-month period. Work-rest cycles varied from normal (8-hr on, 16-hr off) to rotating shifts (24-hr duty in 96-hr time) to unusual schedules (e.g., 4 on, 8 off). An important aspect of this study is that one of the performance measures was a 20-min test in a Link Trainer to score air speed, pitch, bank, rate of turn, and the total time all four instruments were on course concurrently. The other two performance tests were color-naming and an odd-even choice RT task. In general, the less the new routine deviated from the usual, the better the adjustment to it. The diurnal body temperature rhythm was maintained throughout all schedules; however, the ranges dropped from about 2° F (1.11° C) in April to 1.3° F (.72° C) in July and August. Each performance test showed a diurnal variation in performance. Usually, the higher the body temperature, the better the performance. During the entire experiment, subjects maintained their relative performance rank for each test; but performance among tests was poorly correlated. The subjects were divisible into two groups on the basis of body temperature. One group had a higher and less variable daily mean than the other. The group with the higher body-temperature levels also had the highest Link Trainer scores. This relationship did not hold for the other two tests.

Aschoff et al. (6) administered a battery of tests lasting about 15 min every 3 hr over a 4-day period for several different groups, and found that the subjects' performance rose and fell in close synchrony with rectal temperature. The tests consisted of: grip strength, digit cancellation, addition, addition and subtraction according to certain rules, tapping, 4-choice RT to colors, subjective estimate of a 10-sec interval, and a 7-point self-rating scale of alertness. One group followed a normal living routine but had their sleep interrupted twice for testing. When compared with a similar group without night-sleep interruption, the daytime group performance was found to be only slightly impaired, if at all.

Any characteristic of the external environment, as opposed to internal biological factors, which tends to maintain rhythm synchrony has been termed a "Zeitgeiber." To eliminate the light-dark cycle as a Zeitgeiber, Aschoff et al. (6,7) kept a third group in continuous darkness for the 4 days, yet the circadian rhythms persisted. For three additional groups undergoing 1 or 2 nights of sleep deprivation, continuous wakefulness diminished the nightly fall in rectal temperature and reduced daily performance levels. In general, 1 night of uninterrupted sleep gave complete recovery of performance.

According to a followup analysis of the data (Aschoff et al.: 7), the rectal temperature rose slightly with each sleep interruption while still maintaining an overall decline until morning. In addition, the rise occurred with each test session just prior to its start, throughout the day. They (7) felt the rise was due to anticipatory emotional and mental stress, since no test required appreciable muscular work.

Kleitman (121) views body-temperature variations as causing performance changes. He suggests that body temperature reflects the degree of chemical activity in the brain and that:

> "Either (a) mental processes represent chemical reactions in themselves or (b) the speed of thinking depends upon the level of metabolic activity of the cells of the cerebral cortex, and, by raising the latter through an increase in body temperature, one indirectly speeds up the thought process" (121: p. 160).

However, some research has disputed this causal relationship. Williams, Kearney, and Lubin (229) reported no consistent relation between vigilance tasks varying in signal uncertainty and oral temperature over 64 hr of sleep loss. However, the task and temperature were measured only once a day. Uncertain tasks were found to be more susceptible to sleep-loss effects; but, since only a nonsignificant decline in temperature was recorded over the 2 nights, the lack of correlation between temperature and performance was not surprising. This finding does not provide an exception to the relation between performance and temperature.

A more serious challenge to this hypothesis has been provided by Rutenfranz, Aschoff, and Mann (183). They administered to 12 subjects a 5-choice RT test to color, lasting 4 min every 4 hr over 4 separate periods of rotating watchkeeping during sea voyages. Each period lasted 4 - 8 days. The abnormal shifts had no adverse effect on RT; however, significant diurnal variation was observed for both RT and body temperature. RT did increase during those sections of the voyage which exposed the subjects to tropical heat. RT at night was inversely proportional to the duration of the preceding period of sleep or wakefulness; but, in all cases, RT was longer than during the day. Thus the range or oscillation of the diurnal rhythm of RT was largely dependent on the amount of sleep prior to being tested during the night. They (183) did not report this kind of analysis for body temperature. No suggestion was given as to what might control the amplitude of temperature. Body temperature was found to be correlated to RT if all pairs of readings of both variables, obtained at different times of day, were combined. However, no correlation was found when 24-hr means of each variable for different days were correlated, even though both exhibited a range of scores. Similarly, readings obtained at each test time showed no correlation when analyzed separately. In other words, RT could not be predicted from body temperature. Thus Rutenfranz, Aschoff, and Mann (183) have concluded that body temperature and RT are not causally related, but only appear to depend on each other because of their simultaneous control by Zeitgeibers.

A test of letter cancellation has been found to be significantly and positively correlated with oral temperature for separate tests given at 0800, 1030, 1530, and 2100 hr (Blake: 31). Wilkinson (in Colquhoun: 50, p. 229) has offered the following explanation. Unlike the short test given by Rutenfranz et al. (183), Blake's test lasted only 30 min; so reason indicates that, if body temperature and performance are causally related, a more prolonged period of work would be necessary to demonstrate the relationship.

Kleitman has also postulated a basic rest-activity cycle (BRAC), of about 90 min, which is supposed to occur continuously, day and night (121). The evidence for the BRAC comes from extrapolations of known 50- to 60-min periods occurring during sleep-wakefulness cycles in infants, and from the 90-min cycling of sleep stages observed in man. Kleitman feels the BRAC may account for variations in alertness which occur throughout the waking day. Sleep deprivation studies in which lapses are noted, such as those of Williams et al. (230), have never attempted to relate their occurrence to a 90-min cycle; but, presumably, these occurrences would show some correspondence to the BRAC. In the case of Bills' blocking, these lapses occurred much more frequently than every 90 min. Possibly, however, the frequency and the duration of the blocks increase and decrease according to the BRAC, even though the blocks generally increase in frequency and duration with fatigue and still follow a general circadian rhythm (as already shown).

Additional support for the BRAC comes from a study by Orr and Hoffman (159) which has demonstrated a 90-min rhythm for HR in bedrest subjects. During 21- to 44-hr continuous vigilance experiments the 90-min cardiac rhythm was at first wiped out, but it returned by the end of the vigil (Orr, Hoffman, and Hegge: 160). In this study (160) signal detection performance tended to exhibit a 90-min period also. Both HR and performance measures showed a distinct rise in amplitude over the 90-min period, thus indicating a cumulative stress response.

The importance of circadian rhythms to the conduct of air operations has recently been recognized in several areas. Since around-the-clock flying operations are now commonplace, questions have been raised as to the detrimental effect of flying at night when the performance cycle is on its decline. In addition, the added stress factor of disrupted circadian rhythms, caused either by extended operations or by flying through several time zones, has also received research attention. Hauty and Adams (97) conducted studies to determine both the extent of time lag for rephasal of circadian rhythm shifts and the extent of performance decrement induced during transition from one time zone to another. Four subjects were evaluated for several physiological indices, simple and 3-choice auditory and visual RT, decision time (the difference between the simple and 3-choice RT), and subjective fatigue (Pearson and Byars: 166) on alternate days during a 1-week baseline, an 8- or 12-day layover after a transcontinental flight, and a 1-week recovery period upon return. Assessment sessions lasted 25 min and were made at 0700, 1100, 1500, 1900, and 2300 hr local time. Subjects were carried as passengers on all flights and had only minimal activity requirements between test sessions. No records of sleep quantity or disturbances were reported. The directions traveled on the three flights--east-west (Oklahoma City - Manila), west-east (Oklahoma City - Rome), and north-south (Washington, D.C. - Santiago)--were chosen because they provided somewhat comparable flight durations (23.5, 15.5, and 18 hr, respectively), yet represented 3 distinct conditions of time-zone displacement (+10, -7, and 0 hr, respectively). Presumably, different subjects were used in each condition, since order effects could be expected from the flight experience. Hauty and Adams (97) have cautioned that their conclusions were tentative; for the schedules were not identical, a small number of subjects was used, and some slight unavoidable differences were obtained between conditions (as is common in field studies of this type).

For the east-west flight, the primary phase shift for rectal temperature and HR was 4 days. For the west-east flight, rectal temperature required 4 to 6 days to shift; and HR, 6 to 8 days. The north-south flight did not cause a phase shift. The rephasal times for the east-west flight were only about 1 day for both rectal temperature and heart rate; but, for the west-east flight, rephasal was not complete for either measure by the end of the fifth day of postflight assessment. All three flights initially produced a significant but slight increment in subjective fatigue lasting only 1 day, with the east-west flight producing the most and the north-south flight the least. For the return flights, only the west-east flight produced a slight increase in fatigue; but collection of the north-south data was made the day after arrival, so no conclusion could be drawn about that condition.

All RT data were combined for analysis. For the initial east-west flight, both RT and decision time were degraded; and this effect also occurred to a lesser extent on the return flight in this condition. The return flight required 34 hr, and the first day of postflight assessment followed the day of arrival. Neither portion of the west-east flight or the north-south flight evidenced any significant performance decrement.

Klein and his colleagues (118) have conducted studies using an approach similar to that of Hauty and Adams (97) in which baseline circadian rhythms were measured, the subjects transported to different time zones, and rephasal of circadian rhythms measured. After the subjects returned home, the second rephasal period was obtained. In an initial study, they determined the effects of 8 hr of time-zone displacement on pilot performance in a supersonic simulator (Klein et al.: 118). Pilots repeated a standard circular instrument flight, lasting 12 min, every 2 hr for 25 hr during each test-period day. After 2 days of baseline test periods, subjects were flown as passengers from Germany to the United States, where 4 postflight test periods were conducted on days 1, 3, 5, and 8. The subjects returned to Germany 9 days later; and 4 recovery test periods were conducted again on days 1, 3, 5, and 8. Resynchronization of the baseline circadian rhythm required about 8 - 9 days for the outgoing flights, and approximately 5 days for rephasal after return. The magnitude of 24-hr flying performance decrement was greater after the eastward, return flight, than after the outgoing flight. This result is attributed to the nature of airline schedules, which are set up so that a tired traveler to the U.S. arrives in the evening when he may go to bed without delay--whereas the west-to-east traveler arrives in Germany in the morning and must wait about 12 hr to go to sleep if he is going to adjust to local patterns.

Klein, Wegmann, and Hunt (119) measured, on the "Kugelsmachine," the circadian rhythms for rectal temperature, urine, simple visual RT, symbol cancellation, digit summation, and psychomotor performance. The psychomotor test had been previously shown to have a clear-cut circadian rhythm. In this study, direction of travel was reversed. Tests were administered every 3 hr over a 25-hr period: on the 3 days prior to the outgoing flight from the U.S. to Germany; and on days 1, 3, 5, 8, and 13 postflight, both outgoing and return. Local time difference was 6 hr. Testing required about 40 min; and no appreciable workload or fatigue stress was present in the study, other than that caused by travel and the loss of sleep during the two night test periods.

In contrast to the findings of Hauty and Adams (97), rephasal times took much longer. Body temperature required 11 - 12 days to rephase after the westbound flight, and 14 - 15 days after return home. Simple performance measures rephased more quickly. For example, to return to baseline, RT required 6 days after the westward flight, and 9 days after the eastward flight; but the complex psychomotor task required about 12 days for the first rephasal and 10 days for the second. During the rephasal period, the amplitude of the rhythm was reduced or flattened for all measures. Nonsignificant 24-hr mean decrements were found during the first day of each rephasal period: about 1.7% after the westbound flights, and 2.8% after the eastbound. Therefore, upsetting the circadian rhythm apparently did not cause performance to fall lower than that noted during baseline at any point in time. However, the occurrence of low points during rephasal was shifted to times which otherwise showed higher performance during baseline.

Klein et al. (119) concluded that the loss of sleep and increased fatigue due to the differences between the east-west flight schedules (i.e., length of transit times) caused the net decrement rather than either the absolute number of time zones crossed or the disruption of the rhythm per se. They also concluded that desynchronization was not appreciably different whether one was "outgoing" or "homegoing." However, the anecdotal evidence (Preston and Bateman: 175), that the westward flights are less tiring overall than the eastward, appears to be confirmed. They have reasoned that adjusting one's sleeping pattern when one goes through an advance shift (westward flights) is easier because staying awake is less fatiguing (while waiting for local time to catch up) than losing sleep as usually occurs as a result of a delay shift (eastward flight).

In review of these and other similar studies, Klein et al. (120) stated that pilots with the higher performance level were also those with the smallest circadian range of oscillation, a finding along the lines of Kleitman and Jackson (123). According to Klein et al. (120), the flattening of the cycle was related to the extra effort required to maintain performance levels; high motivation levels have sometimes caused an overall elevation of the 24-hr performance curve. Sleep deprivation was concluded to cause an increase in the oscillation and an overall decrease in the 24-hr mean of performance. Thus, for this factor, decrements below baseline levels would be expected. They also noted that, if the subjects had been engaged in continuous work (i.e., had been crewmembers instead of passengers), the oscillations might have been even more pronounced.

Klein et al. (120) found that body temperature gave a good approximation of performance in most studies. However, this need not always be the case. Body temperature has a restricted range, fixed by physiology; performance does not. Also, while the general circadian pattern shows an increase of body temperature during the day, fatigue brought on by a high workload (as opposed to sleep deprivation) would generate performance decrements out of phase with temperature, until temperature started to fall, around 2100 hr.

Klein et al. (120) also concluded from the individual differences for resynchronization among their subjects that approximately 25% - 30% of transmeridian travelers would have no trouble adjusting, while a like number would probably not adjust at all during short stays. The speed of resynchronization would also be a function of the degree of the subjects' interaction with local Zeitgeibers. Outdoor activity and intense social contact could reduce resynchronization time to 6 - 7 days.

While this circadian research indicates that the normally occurring performance cycle will not cause operationally significant performance decline, this cycle must be considered in the design of fatigue experiments, and may have more serious consequences when interacting with additional fatigue stressors.

From the foregoing literature review, body temperature was not only chosen as the best index of circadian cycling but was also evaluated as a predictor of flying performance decrement in the present experiment.

Long-Duration Continuous Performance Research

Although a single flying mission may extend over several days and crewmen may be required to provide continuously high levels of performance over longer than normal duty periods, virtually no research has been performed to investigate pilot performance in these situations. A review of the interaction of work-rest cycles and circadian rhythms has been provided by Trumbull (201); however, no studies specifically related to flying performance are referenced. Most sleep deprivation studies have not imposed continuous work requirements, but only tested the subjects periodically.

The initial coninuous performance studies were concerned with determining the effect of various work-rest cycles for sustained military operations and long-duration space flight. Hartman and Cantrell (91) investigated the effect of a work-rest cycle--of 2 hr on duty, 2 off, 2 on, and 2 off, followed by 8 hr of sleep--on simulated missile system operations for 30 days; and they found essentially no performance decrements for this type of 8-hr workday.

Chiles et al. (48) and Alluisi (5) have carried out an extensive 8-yr program on various work schedules, using the Multiple-Task Performance Battery (MTPB) to determine: the minimal number of men required to maintain around-the-clock operation; the optimum duty periods for such work; the total number of days that performance can be maintained without decrement; and the schedules which make the least demand on man's performance reserves. The MTPB consists of: three passive tests, to assess auditory and visual vigilance and probability monitoring; and three active tasks, to assess memory and information processing, sensory-perceptual stimulus discrimination, and simple group-dependent problem-solving functions. The MTPB requires 40 - 48 hr of practice for asymptotic levels of performance, and evidences clear circadian periodicity when measures are taken on an around-the-clock basis. Their general conclusions (5,48) were that: (a) Two men can handle 24 man-hours of work per day very satisfactorily for at least 30 days. (b) Three men can handle 48 man-hours of work per day up to approximately 15 days. (c) Duty periods of no more than 4 hr are desirable for tasks which are inherently uninteresting. (d) The 4 hr on, 4 off, work-rest schedule of 12 hr of work per day was no more demanding than a normal 8-hr split-shift workday without confinement. (e) A total of 16 hr of work per day on a 4 hr on, 4 off schedule was more resistant to the effects of 44 hr continuous work than was a 4 hr on, 2 off schedule including 40 hr continuous work. To maintain performance following the 4 hr on, 4 off schedules, all external demands must be eliminated and provisions must be made for all physical needs.

In a followup study, Morgan, Brown, and Alluisi (146) determined the effects of 48 hr of continuous work on the MTPB following a baseline schedule of 4 hr on, 4 off, 4 on, and 12 off. Decrements first occurred after about 18 hr of continuous work, and performance decreased to about 82% of baseline during the early morning hours of the first night. Performance then improved to about 90% of baseline during the following day, thus indicating that performance was still significantly influenced by the primary circadian cycle. That night, performance fell to 67% of baseline. After a 24-hr rest period, performance then returned to essentially 100% of baseline. This research demonstrates that performance will not decline linearly with increasing amounts of fatigue or sleep deprivation if continuous performance is required of the subject past at least one circadian cycle. Performance must be sampled throughout the work period by measures sensitive enough to detect circadian shifts if the minimum performance level is to be detected.

Hartman and Cantrell (91) investigated the effects of 64 - 72 hr of sleep deprivation on three work-rest cycles: 4 hr on, 4 off; 4 on, 2 off; and 8 on, 16 off. The experimental conditions simulated a spaceflight of 12 days' duration. This study essentially replicated the work of Chiles et al. (48), except that the sleep deprivation period was extended and a normal schedule, of 8 hr on and 16 off, was included for direct comparison. They used the synthetic work technique, but five different tasks were employed: the Complex Coordinator (a World War II discontinuous 2-axis tracking device, using an aircraft-type stick-and-rudder control system); the Multidimensional Pursuit Test; a multiple RT task; a complex discrimination task; and the Neptune (a task battery consisting of tests of vigilance, short-term memory, arithmetic, and tracking). The findings were similar to those of Chiles et al. (48) except that no differences were found between the schedules prior to the sleep deprivation period. The 8-hr-on and 16-off cycle was most resistant to sleep deprivation effects, and provided a more substantial advantage during recovery. Circadian cycling was observed in a gross sense but was not significant.

Circadian variation has also been shown to occur on the Stressalyser, a step-input pursuit tracking task in which a subject uses a reverse-linked control wheel to position a pursuit pointer over one of five randomly illuminated lights (Buck: 39). The task is paced with a new stimulus occurring after 200 ms of correct target alinement. A trial consists of presentation of 100 target stimuli, and lasts 2 - 3 min. Scores consist of RT between target presentation and initiation of control movement, movement time, error correction time, and overshoot and error (pursuit movements initiated away from target). The device has the advantage of being somewhat portable, and subjects are able to test themselves at home or work once every 4 hr over a 16-hr period. From morning to evening, RT and movement time generally decreased but error rates and overshoot rates increased. Thus, Buck (39) concluded that the speed vs. accuracy tradeoff, used to maintain performance with increasing fatigue, also had an associated circadian rhythm. KR on the Stressalyser affected the level of performance, but not the circadian rhythm.

The Stressalyser has been shown to be somewhat sensitive to two nights of sleep loss when three trials are administered and the results averaged (Buck: 41). Tests were given at 0830, 1230, 1630, 2030, and 0030 hr, each test lasting about 12 min, including rest pauses.

Finally, Buck (39) administered the Stressalyser to pilots and crew members before and after flights--between Vancouver and Tokyo, and between Toronto and Rome--to assess the time required for psychomotor adaptation to new time zones of +7 hr and -5 hr, respectively, including either 24-hr or 7-day layovers. Speed scores tended to relate to time of day where testing occurred; and he interpreted the results to indicate an immediate adaptation of performance rhythm from one time zone to another, in contradiction to most other researchers.

The synthetic work technique of Alluisi and Chiles, while according continuous activity to the subjects, provides a somewhat mild task demand from a workload standpoint. Generalizations to continuous performance situations, with great increases in stress or in mental or physical workloads, must be made with caution. Other studies of continuous work for 48-hr periods have reported wide ranges of performance impairment, depending on the workload imposed, measures employed, and situational variables and physical activity levels required. For example, Drucker, Cannon, and Ware (60) found significant performance decrements both for driving simulator performance and for a target detection task with 15-min breaks every 1.5 hr for 48 hr. Decrements followed the circadian rhythm cycle. In a 48-hr simulated combat field exercise, Ainsworth and Bishop (4) found that tank crews exhibited significant decrement in passive surveillance and driving tests--but not in communication and maintenance functions. In another field study of 48 hr continuous operations, Banks et al. (9) found no performance decrements for soldiers on target acquisition tests, rifle firing, and grenade throwing.

Much more severe impairment has been found for shorter continuous work periods. Orr (158) had two subjects perform continuous vigilance psychomotor and math tests and found that, even though they were well motivated and received monetary incentive, their performance showed almost total impairment by 21 hr. Each task lasted 30 min but, while not specifically stated, virtually no rest between tasks must have been allowed; for the tasks themselves did not impose workloads that could have been intense enough to generate such catastrophic fatigue-induced impairment. Actually, the fatigue stress was somewhat greater than it appeared. Testing began at 1600 hr and subjects had been awake since 0800, so the total time awake was 31 hr. Furthermore, the fact that the training session was conducted from 1600 to 1730, just prior to the start of the experimental session, meant that the continuous performance period was actually 23 hr. The training session would be expected to be much more fatiguing than work in a well-learned task. Nevertheless, this research points out that certain conditions can generate much greater performance decrements than those in most fatigue research.

Just as correlating performance decrement with subjective fatigue has presented difficulties, relating physiological changes to performance decrement has also been difficult. Indeed, Weybrew (214) has concluded (from a review of military studies of operational fatigue induced from sustained performance situations) that these factors are not significantly related and that psychophysiological changes do not occur in all individuals in a given stress situation. However, some investigators have been able to demonstrate a relationship between all three classes of fatigue indicators--physiology, performance, and subjective fatigue.

Grandjean et al. (82) found that tests of grid tapping, critical fusion frequency (CCF), and self ratings of subjective fatigue--administered to air traffic controllers over the course of their 10-hr work day--indicated increasing fatigue, especially during the last 4 hr of work. Significant intercorrelations were found for all three measures. In addition, urinary catecholamines were found to increase significantly more during air traffic control work than during routine office work. No attempt was made, however, to relate these measures to job performance.

The studies in all three domains, by Hauty and Adams (97) and Klein et al. (120), were not designed as fatigue studies per se but to assess circadian rhythm changes. The study by Fröberg et al. (78) induced high fatigue levels due to the somewhat continuous nature of performance requirements over 75 hr, and was able to demonstrate significant relationships among all three classes of measures.

In summary, most long-duration performance research has focused on work-rest cycle effects, using tasks which did not usually demand the continuous high workload in many flying operations. Still, performance decrements have been detected; and, according to some indications, during very long work periods all three classes of fatigue indicators can be shown to be related.

Flying Performance Research Involving Stress and Long-Duration Missions

The first studies of objective pilot performance during fatigue by Bartlett (13), Davis (55), and Drew (59) were undertaken to determine the types of errors made and the causes of performance decrement due to fatigue. These researchers were able to detect statistically significant deteriorations of performance in flights which lasted from 2 to 5 hr. Graphic recordings of the deviations of the airspeed, altitude, heading, and sideslip indicators from the required values were made during a set of 4 maneuvers lasting 10 min. The remaining straight and level segments were not scored. Seven maneuver sets were administered during the 2-hr flight.

From their observations (13, 55, 59), the following types of behavior were found to contribute to pilot error due to fatigue. Control movements became excessive and sluggish, leading to what Bartlett (13) has termed "skill fatigue." Subjects tended to focus on one part of the task to the exclusion of others. Unreasonable mistakes were attributed to lapses of attention. The subject's report as to the quality of his performance was found to be unreliable. According to some indications, deterioration on successive tests was due to cumulative fatigue carryover from one day to the next. Large increases in irritability and emotionality were found with increased performance decrement. Pilots became increasingly aware of physical discomfort. They also noticed an "end-spurt" effect, a tendency for subjects to improve markedly when they thought the flight was almost over. Generally, as fatigue set in, the subjects tended to lower their standard of performance. In more experienced pilots, the onset of fatigue signs was delayed 2 - 4 hr. There was no simple relation between mission duration and errors. Progressive deterioration was not observed in any pilot; the number of errors decreased with time, but their duration and magnitude increased. "Bad patches," 15-min segments in which the error for any instrument was twice the baseline, tended to increase with time; but these patches were interspersed between periods of normal performance throughout the flight (55).

Davis (55) did not attribute the performance decrement to fatigue, but to anticipatory tension or anxiety which was generated as a result of the subject's failure to meet the set performance standards. The tension and ensuing impairment were supposed to be a function of the difficulty of the task. If the subjects could not take effective action to reduce the threat of failure, Davis theorized that even more tension would be built up and performance degraded still further, in a vicious circle. Although some additional studies showed that flying success was related to level of neuroticism and that more difficult tasks showed greater impairment, these findings did not rule out the possibility that time at task (an operational definition of fatigue) could have been related to performance decrement.

The next two investigations of long-duration instrument flight were performed at the Air Force Aerospace Medical Research Laboratory, Wright-Patterson AFB, Ohio. In 1952, McIntosh, Milton, and Cole (141) conducted a study during actual flights. A C-47 aircraft was instrumented to record the time that airspeed, altitude, heading, vertical velocity, inclinometer, rate of turn, bank, and pitch instruments were held within specified tolerance limits. The same basic flight pattern--consisting of alternating 3-min straight and level maneuvers, and 30-min precision maneuvers--was scored for all flights.

Each of three different pilots flew a single mission lasting either 10 hr, 15 hr, or 17 hr, with one refueling stop midway in the flight. During the 17-hr flight, an unplanned landing was required to replace batteries. Tests of addition, illusion, and reading comprehension were administered before, during, and after the 10-hr flight. RT data were collected during the 15-hr flight. Readings from an alertness indicator (as described by Carmichael and Kennedy: 45) were taken every 15 min during the 17-hr flight. Time of takeoff for the 17-hr flight was 0700 hr and was preceded by a full night of rest; the takeoff time for the 10-hr flight was not reported. The takeoff time for the 15-hr flight was 1600 hr, after a full day's work, thus making this flight (objectively) the most fatiguing. Without statistical analysis, the authors concluded from an inspection of the data that:
(a) Objective pilot performance measures gave no indication of a decrement, and therefore decrement was not a function of the length of the flight.
(b) Pilot performance between straight and level and maneuver phases was equal. (c) Performance on auxiliary tests did not change during or after the 10-hr flight. (d) RT measures taken during the 15-hr flight did not change.
(e) The alertness indicator provided some evidence of a change in alertness near the end of the flight. All pilots believed that, although they felt extremely tired by the end of the flights, they could have coped with any critical situation. At no time did the aircraft appear not under proper control or not being flown safely. However, the safety pilots' reports indicate that the pilots became extremely irritable and exhibited some abnormal behaviors.

The fact that this study was unable to detect performance decrements may be due to several reasons. First, the error tolerances for the instruments scored were extremely narrow. For example, altitude had to be held within ± 20 ft. Thus, error measurements taken at the beginning of the flight would be very high. Secondly, only the pilot flying the 10-hr mission was given any practice on the flight plans--a 6-hr flight in a Link Trainer, using the same flight plan performed the day before his actual flight. This training was deemed unnecessary and was dropped for the other two pilots. Thus, for the more fatiguing flights, the baseline scores included error attributable to learning the flight plan. No method was available to estimate that the pilots had plateaued, and no comparisons were made between them. Since the pilots flew the same flight plan over and over, the practice effect was confounded with any decrement due to fatigue. Finally, although a turbulence meter was used to estimate the similarity of each flight period, weather changes throughout the flights could not be controlled.

In 1955, Chiles (47) had four subjects undertake a 56-hr vigil. They performed two vigilance-type simple RT tasks in an aircraft cockpit from the 19th to the 37th hr, and again during the last 15 min of the vigil. During the other times they were permitted to relax. The start time of the vigil and the prior activity of the subjects was not reported. The subjects apparently were allowed to sleep when they wanted, up through the 19th hr. The only conclusion that Chiles could draw from these data was that the measures reflected considerable variability in alertness, but little indication of an overall downward trend in performance.

At the conclusion of the vigil, two of the subjects, who were pilots, were brought to a Link Trainer and then made three instrument landings. Their performance was judged "satisfactory." The total flying time was not indicated. This study was cited by Cameron (42) as an example of the futility of demonstrating performance decrement with the passage of time. As in all cases when conclusions are drawn from negative results, however, the lack of sensitivity of measures employed, deficiencies of the experimental design, and lack of proper controls may have failed to provide a sufficient test of the hypothesis.

These results were contradicted by a study by Jackson (107), in which each of 10 pilots flew four 15-hr flights on alternate nights. Takeoff was at 1700 hr. During each flight, two pilots alternated 2-hr watches. Altitude and heading deviations were recorded for 40 min during a 1-hr segment of manual straight and level flying during each watch. Turbulence level was also recorded. Pilots were permitted to sleep when not on watch. Performance deteriorated over the 40-min segment and progressively deteriorated over the first three watches, but improved somewhat by the fourth. Improvement was due either to the end-spurt effect or to a diurnal rhythm effect, because this watch always occurred after dawn. Jackson concluded that a 2-hr watch was of satisfactory length for this type of flying. Performance did not change from flight to flight over the week; but, from his own observations, he concluded that a fifth flight would not have been advisable.

Hartman (88) investigated the performance decrement occurring during a 24-hr flight in both a C-124 and a C-133 simulator. Four pilots flew a series of 11 legs, each lasting approximately 2 hr, and terminated with an instrument landing. The crew consisted of a pilot who flew the entire flight, a copilot, and a flight engineer. Time-lapse photographs taken evey 20 sec recorded the airspeed and altitude throughout the cruise portion of each leg and were scored for deviations from specified values. No significant differences were found across time. The instrument landings were scored from the ground-track record plotted during each approach. At the end of the 22nd hr, performance showed a significant and substantial drop but returned to normal for the last landing, thus indicating an end-spurt effect. In this study, performance was sustained at normal levels for 20 hr but then, without warning, showed a considerable decrement. These results point out the difficulty in predicting future flying performance from efficiency measures alone.

Recently the Link General Aviation Trainer-1 (GAT-1) has been modified in various ways to provide estimates of pilot performance in simulated flight. The GAT-1 simulates a light single-engine aircraft (such as the Cessna 172) with motion in roll, pitch, and yaw axes. A GAT-1 modified to score heading durations from predetermined flight path has been used to determine the effects of high temperature on performance, heart rate, and body temperature during 50-min flights at temperatures of 25.0° C (77° F), 43.3° C (110° F), and 60.0° C (140° F) (Iampietro et al.: 106). Both of the higher temperature conditions showed significant deviations from the flight path during various legs of the flight.

Utilizing a GAT-1, Gold and Kulak (81) were able to show significant decrement at the 0.005 confidence level for airspeed, heading, vertical velocity control, and deviations from localizer and glide-slope paths during landings at simulated altitudes of 15,000 ft (4572 m). Some of the data indicated that glide-slope control performance was affected at levels of 12,300 ft (3750 m).

Billings, Gerke, and Wick (21) directly compared the effects of secobarbital on instrument landing approach scores from the same pilots both in a Cessna 172 airplane and in the GAT-1. Measures consisted of lateral and vertical angular deviations from the localizer and glidepath centerlines and indicated airspeed. Because of limitations in each system, the two conditions were not identical; the simulator had somewhat lower cockpit workload and communication requirements than the aircraft. Although the aircraft missions could not be completely duplicated from day to day, due to weather conditions, these variables were constant in the simulator. The simulator data were found to be more consistent, had only half the error variability of the aircraft data, and were more strongly associated with the drug-level administered. Billings, Gerke, and Wick (21) concluded that the effects of pharmacological stressors, especially at low doses, were more readily apparent in a simulator than a real aircraft. Higher arousal levels in actual flights may have overcome, in part, the depressant effects of the drug, thus making the aircraft a more conservative test. Overall, the comparability of the data indicates strong similarities between pilot performance in the simulator and real aircraft.

A GAT-1 has been used for a pilot performance evaluation system developed by the USAF School of Aerospace Medicine, Brooks AFB, Texas, to assess the effects of various environmental and chemical stressors (Henry et al.: 99; and Henry et al.: 100). Deviations from prescribed altitude, heading, airspeed, vertical velocity, turn rate, and turn coordination were automatically recorded during a simulated, 1-hr cross-country flight consisting of 37 maneuvers. The first and last segments of the flight were extensively practiced during training, and the pilots were provided cue cards and a display panel to present maneuver instructions. The middle segment, inserted only during a test flight, consisted of 1 of 4 different orders of 19 maneuvers. Cue cards were not provided for this portion, and pilots had to interpret flight commands from the display panel. Takeoffs and landings were not scored. A single score, obtained for the entire 1-hr flight, consisted of the total seconds that each scored instrument reading was outside the predetermined error tolerance. Error feedback lights on the display panel were illuminated whenever any of the scored instrument readings was outside of the error tolerance.

Henry et al. (100) demonstrated, for a small sample of enlisted Airmen, the sensitivity of the 1-hr flight-test to ethyl alcohol dose levels as low as 0.3 gm/kg body weight. This finding corresponds to a blood alcohol level of approximately 25 mg %, which is substantially below the limit used to define legal intoxication for automobile drivers.

These subjects also performed for two consecutive, 25-min periods on either the "Multidimensional Pursuit Test" or the "Complex Coordinator," two World-War II psychomotor tests previously found sensitive to a wide variety of stressors. Performance decrement was comparable to that recorded in the GAT-1 for higher dose levels of 0.6 and 0.9 g/kg body weight. The fact that the GAT-1 task had higher arousal value than the simple tasks did not reduce its overall sensitivity to the stressor, yet the greater complexity of the task may have made it slightly more sensitive to the drug.

In a followup study using 12 USAF instructor pilots, Henry et al. (99) found essentially the same results except that only the higher doses of 0.6 and 0.9 g/kg body weight caused significant decrement. However, the magnitudes of the decrements were approximately equal for the respective pilots and for the previous subjects with no flying experience. Therefore, Henry et al. (99) concluded that, to obtain a conservative estimate of stress effects, nonpilots could be used for most studies, thus providing a considerable cost savings.

In the foregoing study (99) an additional flying performance measure was evaluated in which the pilot used the GAT-1 to track a laser target on a large screen in front of the trainer for 5 min. While this measure was as sensitive to decrement as the 1-hr flight plan, the pilots rated it as a less effective overall test of flying performance than the maneuver test. Additional validity for the maneuver test was obtained by having 4 Air Force flight examiners subjectively rate the performance of 3 pilots in each condition. Although no correlations were computed, a plot of their scores with the maneuver test scores indicated a strong relationship.

The most recent study of long-duration pilot performance and subjective fatigue ratings utilized a fixed-base GAT-2 helicopter simulator that measured flight path and altitude deviations during 3- to 8-hr flights which included additional stressors of noise and 17-Hz vibration (Stave: 191). All flights used the same instrument flight plan, containing 4 takeoffs and landings/hr conducted after the subject's normal day's work. Four flight conditions were investigated: (a) a 4-hr flight, with three 4-min rest pauses/hr and one 8-min rest pause/hr; (b) a 3-hr flight, with no rest pauses; (c) two 3-hr flights, back-to-back with no rest pauses; and (d) two 4-hr flights, with three 4-min rest pauses/hr and one 8-min rest pause/hr (during the 8-min rest, the subject could leave the simulator).

Unfortunately, little statistical analysis was performed; and the basis for most of these conclusions is not clear. At any rate, from a correlational analysis of flying scores and flight duration, noise, and vibration, the following results were reported. Noise levels up to 100 db had no effect on any scores; if anything, they only added to the realism. Some indication existed that the higher the vibration intensity, up to 0.3 g, the better the performance score.

Although fatigue ratings vs. the time of flight were not statistically analyzed, a graph of these two variables shows a general increase of fatigue over time; but the final fatigue levels for the 3-, 4-, and 8-hr missions are approximately the same. The 6-hr mission, however, yields an apparently higher maximum fatigue level. The author concluded that this effect is a result of the lack of rest pauses during the 6-hr mission. He also reported that many individuals would show a rapid increase in their fatigue rating and decline in performance when they realized that the session was almost over. This finding contradicts the end-spurt effect usually reported in such studies.

Stave (191) reported that, as the fatigue ratings increased, the performance score improved. The correlation was "significant" (+0.56) but the level was not reported. He explained that this unexpected improvement could have been a result of the increased effort put forth by the subjects to overcome their feelings of fatigue. He also noted that tasks of greater duration might be needed to demonstrate a significant relation between fatigue and performance. However, since the subjects were flying exactly the same flight plan over and over, the improvement could have also been the result of learning, or perhaps relearning; for the data-collection period covered 1 yr.

He also noted (191) the occurrence of lapses, which were classified as: performance scores exceeding three standard deviations from the mean of scores for that hour. From "examination of the lapse data," he concluded that lapses became less frequent as the flight progressed. For individual runs, no clear-cut pattern emerged.

Lapses were found to be highly correlated with fatigue ratings (+.84). Since Stave (191) did not indicate if the fatigue rating occurred before or after a lapse, the function of the lapse as a dissipator of fatigue was not clarified. Moreover, since he concluded that fatigue ratings increased with flight time, and that fatigue ratings and lapse occurrence were highly correlated, it is not clear why he also concluded that lapses became less frequent with flight time. Nevertheless, he then related the lapse phenomenon to Bills' blocking and to pilot error as a possible causative mechanism for aircraft accidents (191).

To date, a shortcoming of all pilot-performance evaluation systems is that the flight profiles flown are often extremely simple and highly repetitive due to the nature of the scoring system. Hence this is not a good simulation of actual flight; for pilots rarely, if ever, have enough experience with a given flight plan to commit it to memory--as they might by the end of practice in a laboratory study, or during a repetitive series of flights in a long-duration study. Moreover, the actual workload on the pilot is less with predictable flight plans than in actual flight because, after the subject learns what to do, he does not have to take in external commands and process them.

For the present study, the author decided to modify the GAT-1 measurement system used by Henry et al. (100) to provide the capability of presenting flight profiles which were equivalent for statistical comparability but could not be memorized by the pilot. Therefore, a computer was interfaced with the GAT-1 to provide a large repository of scorable flight instructions. A display panel was added to the GAT-1 to provide flight commands to the subject.

Furthermore, the error-feedback information was eliminated, since it was not available to pilots in actual flight. This information not only unrealistically reduced the workload on the pilots, but also might cause some of them to adapt a control strategy in which they waited for the feedback lights to come on before taking corrective action--instead of performing the more demanding task of continually crosschecking their instruments. Feedback would artificially improve their performance; but, in actual flight, performance would be degraded.

As shown by a recent pilot study, fatigued subjects tended to use intervals when they were not being scored as rest pauses and became very lax in performing their flying duties. The present measurement system was designed to score flying performance virtually continuously, so that subjects would be denied these unrealistic rest periods and fatigue effects would be amplified.

Development of an Adaptive RT Task Suitable For Field Studies

No performance measurement technology currently exists which can be used in an operational situation to demonstrate performance decrement due to continuous work. If job performance can be measured directly, then the research approach is straightforward. However, since modification of operational equipment is generally prohibitive because of cost and the lack of objective job-related variables which can be measured, an attractive approach is to measure performance decrements on an auxiliary task. Then decrements on the auxiliary task must somehow be correlated with the criterion measure: job performance.

If an auxiliary task is to be used, three approaches can be taken. The first is to substitute the task for the job entirely as Alluisi (5) has done in his synthetic work studies, in which the course of job-related performance degradation is assumed to follow that in the task. The second approach is to use the actual operational situation, or a laboratory simulation of the job, and to insert a short auxiliary task at intervals during the work period. If the auxiliary task is performed concurrently with the job operations, this approach is usually referred to as primary-secondary task methodology.

The auxiliary task can also be inserted between intervals of job performance. This third approach uses the operational situation to generate realistic levels of fatigue or workload, and the job is assumed to create generalized performance impairment that will carry over to the auxiliary task. This approach, used in the present study, is an attempt to assess a person's fundamental performance capability as modified by situational or environmental variables. If the fundamental performance capability in an information processing domain could be measured reliably, some generalization to a wide spectrum of specific behaviors or jobs would be expected.

According to the review of the literature: (a) Fatigue-related performance decrement is difficult to obtain without inducing fatigue levels which are operationally unimportant to flying; for example, 64 hr of sleep deprivation. (b) The existing tests are not suitable for inclusion in an operational environment. (c) No attempt has been made to relate existing tests which are sensitive to fatigue; for example, Leonard's 5-choice serial reaction test (133) to flying performance. In the present study an attempt has been made to remedy these deficiencies.

The characteristics of tasks sensitive to sleep loss and continuous performance have been noted by Naitoh (152), Wilkinson (223), and Woodward and Nelson (231). To develop a task sensitive to fatigue stressors, the following factors have been taken into account in the present study:

1. Monotonous tasks and those with high complexity or difficulty but relatively low interest are highly sensitive to sleep loss.
2. Self-paced tasks show little loss in accuracy, but total response time increases with sleep loss.
3. As the time available for making a response increases, the task becomes less sensitive.
4. The longer a task lasts, the more sensitive it is to sleep loss.
5. Tasks measuring blocking or gaps tend to be most sensitive to sleep loss.
6. Tasks requiring continuous performance, not providing breaks or rest pauses, are more sensitive to sleep loss.
7. Tasks requiring minimal physical activity are more sensitive to sleep loss.
8. Sleep loss tends to increase RT, but not consistently.
9. Knowledge of results (KR) increases resistance to sleep loss.
10. Newly acquired skills or tasks, that have not been well practiced or have not reached a plateau in learning, are susceptible to sleep loss.

(Tasks making severe short-term memory demands are also susceptible to sleep loss, but this factor has not been incorporated in the test developed for the present study.)

In order to develop a task which meets the practical requirements for field use, the test should be:

1. Easily and quickly learned to a stable baseline level.
2. As short as possible in duration, to preclude any significant additional workload or fatigue on the operator and to keep to a minimum any interference with ongoing operations.
3. Easy to administer.
4. Resistant to "spoofing" or guessing.
5. As resistant as possible to changes in motivation.
6. Nonauditory in nature because of varying noise levels in operational settings, and varying degrees of hearing loss found in operators.
7. Conducted with apparatus capable of being made highly portable with its own independent power supply.
8. Generally acceptable to the population under investigation (Gartner and Murphy: 80).

To satisfy the operational requirements with a test sensitive to sleep deprivation, the author developed a complex 5-choice visual RT task and incorporated it into a computer-based adaptive logic presentation system. The result has been termed the "Discrete Information Processing Test (DIPT)."

The distinctive feature of an adaptive logic system is that the subject's response is fed back to modify the difficulty level of the next stimulus presentation, on the basis of how well the subject is performing. An early use of this concept was the development of programmed instruction techniques in which the order of training material was established by a predetermined criterion of student progress. Later applications of adaptive technology to performance research have been reviewed by Kelley (115).

One of the first uses of adaptive technology was to improve training on psychomotor tracking tasks (Kelley: 114). In an adaptive tracking task, during a continuous tracking period, the subject's integrated error score was used to adjust the difficulty level of the task and thus maintain the error score at a preset criterion. For several reasons, this technique required less time than fixed tracking to train the subject to a baseline criterion. First, training periods are more productive, because little time is wasted in giving the subject practice at a difficulty level which is either already mastered or entirely above his present capability. Second, and equally important, the subjects maintain high motivational levels on this type of task. They are given an optimum challenge: no matter how hard they try, they cannot succeed; yet they never experience severe failure. The adaptive tracking task was extended to jet flight simulator training by Ellis et al. (65). The turbulence input to a fixed-base Universal Digital Operational Flight Trainer Tool was modified continually as a function of how well the students were able to hold the simulated aircraft within a given criterion of reference altitude. These students, as compared with a group of students conventionally trained, improved more rapidly and made fewer errors. Moreover, rather than having developed problems in controlling a secondary variable (such as airspeed), they seemed to have developed a control strategy which reduced other errors as well.

A variation of adaptive technology has been applied to the secondary task approach to assess reserve capacity (Kelley and Wargo: 97). In this case, the difficulty of the secondary task is adjusted on the basis of primary task performance in an attempt to quantify and control the operator effort expended in maintaining various levels of performance. A similar approach was taken by North and Gopher (154) to attempt to predict flying performance. In their study the adaptive feature was used to maintain dual-task difficulty by giving the subject continuous KR for both tasks relative to the desired criteria.

Wargo (207) has discussed the procedures for developing adaptive measures and described an adaptive task developed for assessing time-estimation skill. Wiener and Keeler (217) have discussed the application of various adaptive strategies to vigilance tasks. In order to maintain criterion performance, the adaptive task became progressively less difficult throughout a vigil.

Adaptive techniques have been used to determine various psychophysical thresholds; for example, the point of subjective equality for visual latencies (Kappauf: 113). Hudson (105) has described a proposed adaptive measurement system for evaluating multiparameter visual thresholds. In this system, the adaptive technology is used to reduce the total number of data points required to predict the threshold. The computer generates hypotheses to predict the subject's response from a minimum number of observations. If the criterion is not met, the computer switches to the next higher order equation and forms a new prediction. This procedure is continued until the error between the observed and predicted values falls to the desired level. This adaptive paradigm differs from the usual, in that the subject's response is not used to modify the difficulty of the stimulus comparison but to modify the prediction of the subject's next response.

In the adaptive task developed and used in the present study, the rate of stimulus presentation was a function of the latency of the prior response time and error rate. The computer increased the rate of presentation until the subject's information processing capacity was overloaded. This presentation rate was taken as the subject's threshold of information processing capacity. This task has two sources of overload: first, the processing threshold at which the subject can no longer keep up with the presentation rate; and second, a block (as described by Bills) in which the subject cannot respond for a short period of time, regardless of the presentation rate. To the extent that the latter block is not indicative of the true information processing threshold, the subject is given a second chance to lower the threshold. However, the predominant source of overload will be the Bills'-type of blocking, as it becomes more frequent with fatigue.

The task consists of the following steps: Initially, the subject must learn an association between five symbols and a corresponding numeral from 1 to 5. Each trial consists of a display of one of the symbols and all 5 numbers in random pattern. The subject must press a button corresponding to the location of the number which has been associated with the displayed symbol. This information processing task requires 4 subtasks--(a) recognition of the symbol; (b) recall of the number-symbol association from long-term memory;

(c) search for the location of the correct number; and (d) the motor response to push the location of the correct button. This task apparently afforded the required complexity level, yet would not require a lengthy training period. KR is not provided to the subject. Guessing is easily detected, since the probability of responding correctly by guessing is only 0.2.

The adaptive logic of the task is as follows: The subject is initially presented displays in a self-paced mode to which he responds as fast as he can. The computer averages the RT for the first four correct responses and, from this point on, the rate of presentation of the panels is computer-paced. The computer presents each panel initially for the duration of the self-paced RT score. The determination of the subject's initial response capability saves task presentation time; for the task, if started at a fixed presentation interval, would have to be slow enough to encompass the slowest initial response time expected from any subject at any level of fatigue.

The task is designed to be simple enough that the subject will almost always respond correctly. Wrong answers are scored primarily to protect against any subject who might try to guess. If he makes 2 errors before he has 4 correct, the presentation rate is reduced by 100 ms. If he has 4 correct before he makes 2 wrong answers, the presentation rate is increased by 100 ms, up to the rate where he blocks. A block is defined as: the presentation of 2 panels in a row to which the subject does not respond. The presentation rate at which this block occurs is recorded. When a block is detected, 2 self-paced presentations are made to allow the subject to recover his concentration. However, no rest pause is provided during this period. The speed of presentation is then reduced by 300 ms, and the adaptive presentation logic is continued. The duration of the task is variable; it lasts only as long as necessary for the subject to make 2 successive blocks within 200 ms of each other. A diagram of the adaptive logic program is presented in Figure 1.

Thus the task contains both paced and unpaced components. It is primarily computer-paced, but responds to the overload condition by slowing the presentation rate just as the subject would if he were in control.

The average response time when blocking occurs has been found to be about 600 ms. Pilot studies have shown that the total task time required for a subject to block within the prescribed limits is normally about 60 sec, and that the subject will normally produce 2 - 4 blocks to reach criterion. Note that the definition of a block is relative in that it is based on the ongoing presentation rate (such as used by Bills), in contrast to a fixed-interval definition (such as the gap measure used by Wilkinson).

Critical to this type of adaptive technology is the use of a high-speed computer; for it permits a continuous task to be presented to the subject, scored, and modified--with delays which are imperceptible to the subject. The computer utilizes overload information to adapt the speed of presentation to the information processing capability of the subject. No subject can keep up with the speed with which the computer can present and score the task, so the computer can always present the displays and score the responses at rates faster than the subject's processing threshold speed. With the use of miniaturized microprocessors, this task could probably be made "field portable."

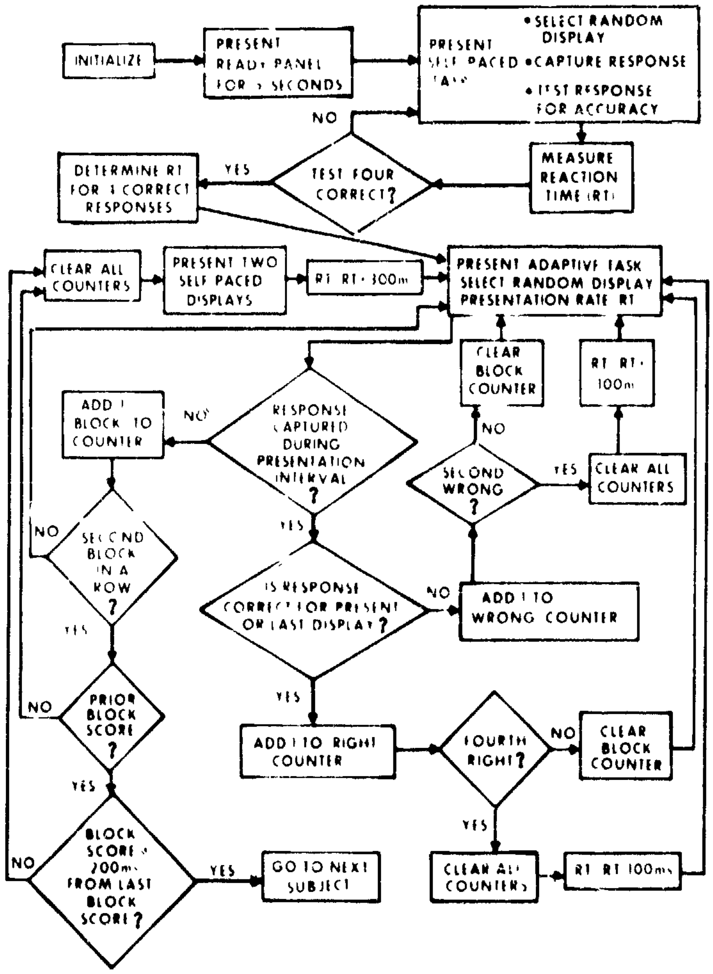


Figure 1. Adaptive flow logic diagram for discrete information processing test.

Wilkinson (224) has maintained that, in order to show mild sleep-loss effects, a test must last at least 30 min and preferably 1 hr. Williams et al. (230) found that a 3-min test of addition was sensitive to 48 hr of sleep loss for the number of additions attempted--but not for accuracy achieved, even up to 84 hr of sleep loss. Donnell (58) gave a 1-hr version of the Wilkinson (224) 2-digit addition test to sleep-deprived subjects and found that: after the first 10 min of the test, the number of additions attempted was sensitive to 32 hr without sleep; and, after only 6 min of testing, 64 hr of sleep loss could be detected. However, this result may have been due to the fact that the subjects knew the test was to last 1 hr. But in 1972, Lisper and Kjellberg (135), using a simple, auditory RT test, were able to detect 24 hr of sleep deprivation during the first 5 min of a test lasting only 10 min. Blocking, defined as RT's twice the period average, significantly increased during the last 2 min of testing of the sleep-deprived condition. The subjects were tested in a highly unarousing situation, however--half-reclining, in a dimly lit and soundproof chamber--so this test would not be suitable for field use. In the present experiment, sleep loss was a maximum of 24 hr; but the high continuous workload due to flying is much more fatiguing than equivalent time in sleep deprivation studies. Thus, even though the adaptive RT task lasted only about 60 sec, the subject with less than 24 hr sleep loss might be expected to show performance decrement on this task.

Rather than being specific to an evaluation of fatigue effects alone, this task would doubtless be sensitive to any stressor which lowered a subject's information processing capability (for example, hypoxia). Since speed of processing is critical to performance in many Air Force jobs (such as Air Traffic Controller), decrements on the adaptive task should generalize to many jobs other than flying.

The continuous flight performance assessment system in the present study provided criteria to determine the relationship of the auxiliary task to flying performance. However, activities requiring discrete information processing will become an increasingly large proportion of the pilots' overall workload in future military aircraft operation. Increased requirements for high-speed decision making, auditory and visual monitoring, and the use of digital data input and display devices--combined with those for computer-assisted takeoff, landing, and auto-pilot systems--will tend to reduce the importance of the psychomotor tracking skill in flying. Thus, the auxiliary task developed may become closer to being the criterion in future pilot proficiency assessment.

Statement of the Problem

From the foregoing literature review, conflicting results were found for the effects of fatigue on performance, depending on the workload demands, task duration, and the measures employed. No study has been performed to determine the cumulative effects of consecutive long-duration missions on flying performance, physiological cost, and subjective fatigue.

The need for additional research in this area has often been recognized. In 1971, Cantrell, Trimble, and Hartman (44) reviewed categories which create stress and may reduce long-term aircrew effectiveness in flying the C-5 aircraft. The unique capabilities of this aircraft can impose severe aircrew workload and work-rest cycle requirements on the crew members, and these authors had hoped that previous research could provide operational guidelines for the C-5 manning requirements and management policies. However, after an extensive review, they (44) concluded that: little research was direcly applicable to their problem; most research concerned short-term rather than long-term stress; the nature of past research permitted little generalizability; and systematic research (in the laboratory or field) was lacking in these areas. At about the same time, the lack of this type of research data applicable to U.S. Army continuous field operations was noted at a symposium chaired by Hodge (104). There, the conclusion was that laboratory research in the past had not addressed these types of problems, and research goals were recommended to bridge gaps in knowledge of long-term performance. Thus, useful research concerning continuous performance is generally lacking for most operational situations, not only those for aircraft.

In order to understand the effects of continuous, long-duration missions on flying performance, the present experiment was performed to interrelate the effects of two fatiguing flying schedules on: two objective measures of flying performance, discrete information processing, subjective fatigue and sleepiness, two heart-rate indices, rectal temperature, and urinary excretions.

Despite widespread disagreement on a satisfactory definition of fatigue, the author feels the concept can be useful in several ways. Of the three classes of measures used to detect fatigue effects, only the subjective report of fatigue appears to be specific as a measure of fatigue itself. This finding is in accord with the position of Bartley and Chute (15). Moreover, one may view the concept of fatigue as a hypothetical construct or intervening variable which accounts for performance decrements and physiological changes, as Bills did. Fatigue can also be used as an independent variable, with its level operationally defined by a set of task demands and the total amount of time spent on the task from a rested baseline. In the present study, fatigue stress was operationally defined by the duration of three variables controlling mission length: (a) the time awake prior to flying; (b) daily duration of flying; and (c) the total number of days (crew duty periods) that the mission entails.

The schedules used for the simulated missions have been designed to evaluate the crew duty limitations allowed by current Air Force regulations. Flight Management, AFR 60-1 (1), is modified by supplemental regulations; but, in general, its requirements are as follows: The maximum crew duty day for dual control aircraft is 24 hr; and, for single control or single pilot aircraft, 12 hr. The total flight time within the crew duty day is not regulated, but a pilot can accumulate up to 330 flight hr per quarter and a maximum of 125 hr per 30 consecutive days. A minimum of 12 hr of crew rest is required between duty days; at least 8 hr of uninterrupted rest must be provided; a requirement for true sleep is not specified. After 3 consecutive maximum duty days, 24 hr of crew rest are recommended. These provisions may be waived only by the commander of a major Air Force command, only when mission priority justifies the increased risk.

In Schedule A in this study, 12-hr crew duty days were alternated with 12 hr of crew rest for 4 consecutive days. This schedule is within the foregoing Air Force regulation and causes no disruption of circadian rhythms. The second schedule (Schedule B) required two 24-hr duty days between two 12-hr duty days. Between all duty days, all crews received 12 hr of rest. During the first 12 hr of the 24-hr duty day, subjects performed light nonflying activity. The second 12 hr corresponded to the crew duty day of the first schedule, which consisted of 2 simulated flights, each lasting approximately 4.5 hr. In the second schedule, the crew performed their second set of two flights during a disrupted circadian rhythm cycle. Since the crew flew their third set of flights during the normal performance period, a determination of circadian effects could be made. The 2 flying performance measures and the DIPT differed in their level of stimulation; thus, a test of the arousal hypothesis could be made.

Traditional psychological measures of performance (such as tracking, reaction time, vigilance, short-term memory, or arithmetical manipulation) are often not sensitive to reduction in performance over a long period of time, because of a human's reserve capacity and spontaneous recovery effects. When the subject is not tested to his performance limits, he is able to compensate on subsequent trials by extra effort or attention which, in turn, may precipitate the detected physiological changes. Thus, on the average, he will maintain his performance at baseline or at that level of perceived response requirements which appears to be satisfactory for task accomplishment. Within this globally acceptable behavior pattern, however, blocks or lapses occur during which his performance is far below task requirements. Only rarely do task requirements so far exceed a subject's performance capability that he does not have the time or the reserve capacity to recover his performance decrement spontaneously. In an operational situation, this resulting lapse (or failure) in the information transfer process is detected and perceived as important only when it leads to an accident, personnel injury, or some other serious mission disruption.

In theory, fatigue has two primary detrimental effects on human performance: (a) to increase randomly the involuntary momentary lapses of attention which decreases information processing capability; and (b) to slow perceptual response speed. According to this study, if a subject is experiencing a stressful situation (as shown by abnormal physiological changes), then psychomotor and perceptual performance decrements (in the form of reduced reserve capacity) are also occurring. However, the appearance and detection of fatigue effects are conceptualized as being partially task-dependent; and thus flying skill, in the form of a multidimensional tracking task, would not be expected to deteriorate as much as a perceptual test whch pushes a pilot to his limit of processing capacity without allowing for a period of spontaneous recovery. In the present study, the adaptive RT task, which pushes him to his performance limits, was designed to be as short as possible in order to induce minimal fatigue and cause minimal interference with the task of flying. Thus the magnitude of decrement in this task was not expected to equal that in the flying performance tasks.

Welford (212: p. 184) has stated that, when valid subjective and physiological manifestations of fatigue occur, an underlying performance decrement is always present if one knows where to look for it. That is also the position taken in the present study: True subjective fatigue and increased levels of physiological cost indicate that maximum performance capability has been reduced--although this decrement may not show up if the task demands are such that adequate (baseline) performance can be maintained, because maximum performance capability is not required.

The workload imposed by the task, the sensitivity of the measurement, and the specific schedule of time at task will affect the onset rate and severity of physiological cost, the subjective feelings of fatigue, and the performance decrement. In general, subjective complaint, which will occur first, will be followed by physiological changes and, finally, performance decrement. Depending on the severity of the fatigue induced, performance should return to baseline first, and then should be followed by subjective fatigue and physiological recovery.

Statement of Hypotheses

The following specific hypotheses were deduced and tested in this study:

1. The psychomotor tracking dimension of flying performance as measured by a Straight and Level Test (SLT) and a Flight Maneuver Test (FMT) will significantly decrease as a result of three fatiguing stressors:

   a. Time awake prior to flying (1 vs. 12 hr, thus creating crew duty days of 12 or 24 hr duration)
   b. Daily flying duty duration (9 hr of flying time)
   c. Total mission duration (4 days)

The fatigue stressor (a) divided the subjects into two groups. Stressors (b) and (c) occurred across all subjects, regardless of time awake prior to flying.

2. The discrete information processing capacity dimension of flying performance as measured by an adaptive test, the DIPT, will significantly decrease as a result of the fatiguing stressors just discussed.

3. Subjective reports of fatigue and sleepiness will significantly increase as a result of these fatiguing stressors.

4. Subjects awake 12 hr preflight (Schedule B) will report significantly longer duration of sleep postflight than those awake only 1 hr preflight (Schedule A).

5. Because of the fatiguing stressors, HR will significantly increase and then decline, and HRV will significantly increase. Confirmation of this hypothesis will support the concept that increasing fatigue reduces the level of arousal.

6. Body core temperature will significantly decline as a result of the fatiguing stressors when circadian effects are controlled. Temperature will rise from morning to evening, then decline until 0400 hr.

7. Urinary excretions of Na, K, Na/K ratios, 17-OHCS, epinephrine, norepinephrine, and urea will significantly increase as a result of the fatiguing stressors.

8. The magnitude of increases in HR and decreases in HRV will be significantly related to the arousal value and to the attention or concentration demands of each performance test. The simple psychomotor tracking demands of the SLT are predicted to have the lowest arousal value, and are followed by the complex tracking task, the FMT. The adaptive RT task is predicted to have the highest arousal value, due to the intense concentration required and the fact that subjects were pushed to their performance limits by this task.

9. Significant interaction will occur between the fatiguing stressors and the two psychomotor tracking tasks. Performance on the simple unarousing tracking task (SLT) will decline more, due to fatigue, than on the complex arousing tracking task (FMT).

10. A fourth source of fatigue stress, the circadian rhythm effect of decreased performance on night vs. day flights, will be significantly greater than the cumulative fatigue effects of the fatigue stressors. On night vs. day flights, flying performance measures, HR, and body core temperature will significantly decrease--and HRV and subjective fatigue and sleepiness will significantly increase.

11. Performance on the SLT, FMT, and DIPT will be significantly and positively correlated with each other.

12. In the presence of the four fatigue stressors, the changes in three measures of flying performance will be significantly related to changes in: subjective fatigue and sleepiness; HR; HRV; body core temperatures; and urinary excretions of Na, K, Na/K ratios, 17-OHCS, epinephrine, norepinephrine, and urea. Together, these changes will improve prediction of flying performance on a long-duration mission.

13. Because of the hypothesis that reports of subjective fatigue and sleepiness measure a similar underlying dimension, predicted changes in subjective fatigue and sleepiness will be significantly correlated in the presence of the four fatigue stressors.

14. In the presence of the four fatigue stressors, reports of subjective fatigue and sleepiness will also be highly correlated with decreases in body core temperature.

Relationships between changes in HR and HRV and changes in subjective fatigue, sleepiness, body core temperature, and urine were not examined in this study.

Analysis of Variables

The independent variables of fatigue stress have been defined in the hypotheses by the time awake before flying, the daily flying duration, the total mission duration in days, and night vs. day flying. The dependent variables are defined as:

1. Multidimensional tracking tests of flying performance.

a. Straight and Level Test (SLT)--the combined total seconds that the airspeed, altitude, turn rate, turn coordination heading, and vertical velocity instruments were outside of prescribed tolerances during a 10-min segment. Before each SLT, the subject was given 2 min to adjust his flight parameters to those displayed to him. This time would normally be sufficient in which to make any changes necessary from his previous flight path. The subject then tried to hold the aircraft on the displayed course during the next 10 min.

b. Flight Maneuver Test (FMT)--the combined total seconds that the instruments cited in (a) deviated outside prescribed tolerances during 8 consecutive flight maneuvers (Table 1). The maneuvers were separated by 25 sec of unscored flight, during which the subject could make any course adjustments necessary to prepare for the next maneuver. All 8 maneuvers were presented for each test but in random order. The total test required 17 min, of which 13 min were scored. The total possible error score for the SLT and the FMT was equal.

TABLE 1. EIGHT MANEUVERS PERFORMED IN RANDOMIZED ORDER IN EACH FLIGHT MANEUVER TEST

Maneuver	Vertical Velocity[a] (ft/min)	Duration (sec)
1. Straight climb	+250	120
2. Straight descent	-250	120
3. 180° climbing turn	+250	60
4. 180° descending turn	-250	60
5. 270° climbing turn	+500	90
6. 270° descending turn	-500	90
7. 360° level turn, left	0	120
8. 360° level turn, right	0	120
	Subtotal:	780
	Initial course adjustment	+65
	Intermaneuver course adjustment (7 at 25 sec each)	+175
	Total:	1020

Airspeed: 70, 80, or 90 mph, assigned to each maneuver at random.

[a] Vertical velocity instrument is calibrated only in ft/min.

2. Discrete Information Processing Test (DIPT)--This adaptive RT score is defined as the threshold of the subject's information processing capability. DIPT is the average, in milliseconds, of the last two presentation rates to which the subject omitted two responses in a row.

3. Subjective Fatigue--The first measure was the score derived from the SAM Subjective Fatigue Checkcard (Pearson and Byars: 166). The second was the score derived from the Stanford Sleepiness Scale (SSS) developed by Hoddes et al. (103) to quantify progressive levels of sleepiness. This 7-point rating scale was sensitive to 24 hr of sleep deprivation, and correlated with performance decrement when performance declined significantly due to sleep deprivation. (Both forms are shown in Appendix A.)

4. Measures of Cardiac Activity
   a. Heart Rate--the time interval between successive R-R intervals, converted to beats per minute and sampled every 5 sec during the period of the SLT, the FMT, and the DIPT.
   b. Heart Rate Variability--the standard deviation of the HR sample during the SLT, the FMT, and the DIPT.

5. Rectal Temperature--the average of the rectal temperature, measured continuously and sampled every 5 sec during the SLT, the FMT, and the DIPT.

6. Urine Battery--the quantitative levels of Na, K, urea, 17-OHCS, epinephrine, and norepinephrine, as determined from the chemical analysis of each urine sample. The respective excretion rates are known to vary during a 24-hr period. To control for this variation, each variable was referenced to 100 mg of creatinine; for the total amount of creatinine output is fairly constant from day to day, and is generally a function of lean body mass (Hale, Ellis, and Williams: 84). Also an Na/K ratio was determined for each sample.

7. Sleep Data--The form used to collect information on sleep duration and sleep quantity is shown in Appendix A. This information was used primarily as control information, so that performance differences could be attributed to differences between schedules and not to differing amounts of sleep between groups. However, subjects in the more fatiguing schedule were expected to sleep more during recovery phases.

METHODS

Subjects

Due to the nonavailability of Air Force pilots, 24 male Airmen who had just completed Basic Military Training at Lackland AFB, Tex., were selected to participate in this study. To obtain subjects with sufficient flying aptitude to provide acceptable performance in the GAT-1, and also to permit generalizations to the Air Force pilot population, a stringent selection program was instituted during their 30 days of basic training.

First, the Armed Services Vocational Aptitude Battery test scores were screened on two groups: Airmen who entered training 8-14 March; and 12-18 April 1977. On their fourth day of training, those who had obtained a G score of 70 or higher and a QT score of 75 or higher were given a briefing on the nature of the experiment and were requested to volunteer. These two scores give an estimate of general aptitude and intellectual ability; the cutoffs established were deemed the minimum required for success in the following phases of the selection process.

Of the 2,368 Airmen screened, 709 qualified and were given the briefing. The primary incentives offered to the subjects were 7 days of convalescent leave at the conclusion of the study, an opportunity to learn about flying, and a thorough medical examination. They were informed at this time that, if selected, they may be required to remain awake up to 36 hr during simulated missions lasting as long as 8 days.

Obviously, due to the expense and time involved in obtaining and training subjects for this study, selection procedures had to be geared toward obtaining subjects with the maximum probability of completing the project. Subjects were asked not to volunteer if they had a fear of flying, or easily became airsick, carsick, or seasick. Other subjects excluded were those who had ever had periods of: unexplained unconsciousness; frequent or severe headaches; sleepwalking or bedwetting episodes, after age 16; epileptic seizures or fits; dizziness or fainting spells; nervous trouble of any sort; piles; or hemorrhoids, or rectal disease. Subjects had to be less than 6-ft 2-in. tall (183 cm) and weigh less than 210 lb (95 kg). These requirements were part of those medical selection criteria used in Air Force pilot selection (3) which were considered relevant to the present study. Subjects under 18 yr of age were likewise excluded. Married Airmen were also discouraged from volunteering; for a young, newly married Airman--who had already been absent from his bride for the 45 days of basic training--would probably have found the additional 30-day period of absence (required by the study) stressful, and would have had difficulty in completing it.

The 275 Airmen who volunteered were then given the pilot selection portion of the Air Force Officer Qualification Test (AFOQT) on their sixth day of training. This test lasts approximately 2.5 hr; and a minimum score of 25 is required for an applicant to enter Air Force flying training. Next, the biographical records of those 188 subjects attaining the minimum score of 25 were screened to eliminate those who, from a combination of past indicators (such as criminal behavior, traffic violations, school or employment problems, or drug usage) appeared to demonstrate a history of immature behavior and would have difficulty completing the study. Also eliminated from the study at this time were subjects whose career fields were in critical demand by the Air Force. The remaining subjects were rank-ordered; then 55 from the first group and 57 from the second, who were judged most qualified by the author, were chosen to receive a class II flying physical. It was administered, on their eleventh day of training, by the Lackland AFB Clinic in accordance with AFR 160-43 (3).

The class II physical examination consists of a thorough self-report of prior medical history, visual and hearing tests, laboratory analyses of blood and urine samples, physical measurements, blood pressure and pulse measurements, a resting electrocardiogram, a dental examination, and a clinical examination by a qualified Air Force flight surgeon. This physical is given yearly to all Air Force personnel who are on flying status.

The class I physical examination is used to qualify officers medically for entry into the Air Force flight training program. Class I differs from the class II physical, principally in having a more stringent visual examination. In the present study, visual acuity requirements were waived, provided that subject had vision correctable to 20-20 by eyeglasses.

The results of the physicals were reviewed by the author; then 23 Airmen from the first group and 26 from the second, who met the necessary medical requirements and who were most qualified on the basis of previous test scores, were selected for a structured, personal interview. The interview occurred on day 22 of the subjects' training. The multiple purpose of the interview was not only to insure that the subject clearly understood the stresses involved in the experiment and what was to be expected of him, but also to obtain his informed consent for participation in the study as required by AFR 80-33 (2). A copy of the consent form is provided in Appendix A (Fig. A-4).

In addition to the interview data, peer ratings were available on all subjects. Subjects were ranked by the author on their motivation, maturity, test scores, and medical records; and the 15 judged most qualifed from each group were selected to receive GAT-1 training at Brooks AFB, Tex. Subjects arrived at Brooks AFB the day after completing basic training. Thus, the 12 subjects in each group who had attained the highest flying training scores were used in the study.

Upon arrival at Brooks, all subjects were given a treadmill stress evaluation of their cardiac condition to insure that they did not have any unusual arrhythmias or conduction defects. In addition, the subjects were given a standard battery of psycnological tests (administered by the USAFSAM Clinical Sciences Division, Neuropsychiatry Branch), which consisted of: the Minnesota Multiphasic Personality Inventory; the Benton Visual Retention Test; the Wechsler Adult Intelligence Scale (WAIS); the Halstead-Reitan Impairment Index; the Edwards Personal Preference Schedule; and the 16 Personality Factor Questionnaire. These tests are used as part of a clinical examination to determine a pilot's fitness to fly, after recovery from an accident or if an emotional disturbance is suspected. The dual purpose of testing the subjects was to provide a more complete sample description and to insure that no subjects possessed personality traits which were obviously deviant from the normal Air Force pilot population. No deviant patterns were noted by the Neuropsychiatry Branch. For the 24 subjects used in the study, the average full-scale WAIS IQ was 118.75, which compared very favorably with the overall Air Force pilot population. For example, the full-scale WAIS IQ of 50 pilots randomly selected from a cross section of 298 Air Force pilots was 119 (Fine and Hartman: 72).

The average number of years of education for the 24 subjects was 12.5, and the average age was 20.1 yr. Their average AFOQT score was 66; their average G score and QT score, 92. Thus the primary differences between the sample in this study and the Air Force pilot population appears to be only that all Air Force pilots have college educations and are officers.

Apparatus

Four GAT-1 moving-base trainers (Model No. B633000 manufactured by the Link Division, Singer-General Precision, Inc., Binghamton, N.Y. 13902) were modified so that voltage inputs which determined the readings of 6 flight instruments were sampled 5 times/sec by a PDP-12 computer. The sampling was done for each trainer simultaneously during every scored segment of the flight. These values were digitized and compared to stored values that corresponded to the flight instructions being given to the subject via the Flight Director Panel (FDP) mounted on the instrument panel in the GAT-1 (Fig. 2). An instrument was scored in error if the comparison resulted in a difference between the indicated and required values which was outside of the following limits--

1. Airspeed: ± 5 mph
2. Heading: ± 2.5°
3. Altitude: ± 40 ft
4. Vertical Velocity: ± 150 ft/min. (FPM)
5. Turn rate: one-half needle-width deviation from the standard-rate turn (3°/sec) markers or the level-flight marker.
6. Turn coordination-movement of the ball to either outside edge of the "coordinated turn" markers.

All 6 instruments were scored for the SLT. During each maneuver of the Flight Maneuver Test (FMT), only those instruments which were not changing in value were scored. Thus, in all turns, the heading indicator was not scored; in all climbs or descents, the altimeter was not scored. The flight scoring system for each GAT-1 was calibrated at the beginning of each of the 6 blocks of data collection.

All flight commands and the DIPT were presented to the subjects via the respective FDP, which was an alphanumeric, 256 (8 x 32)-character, Self-Scan

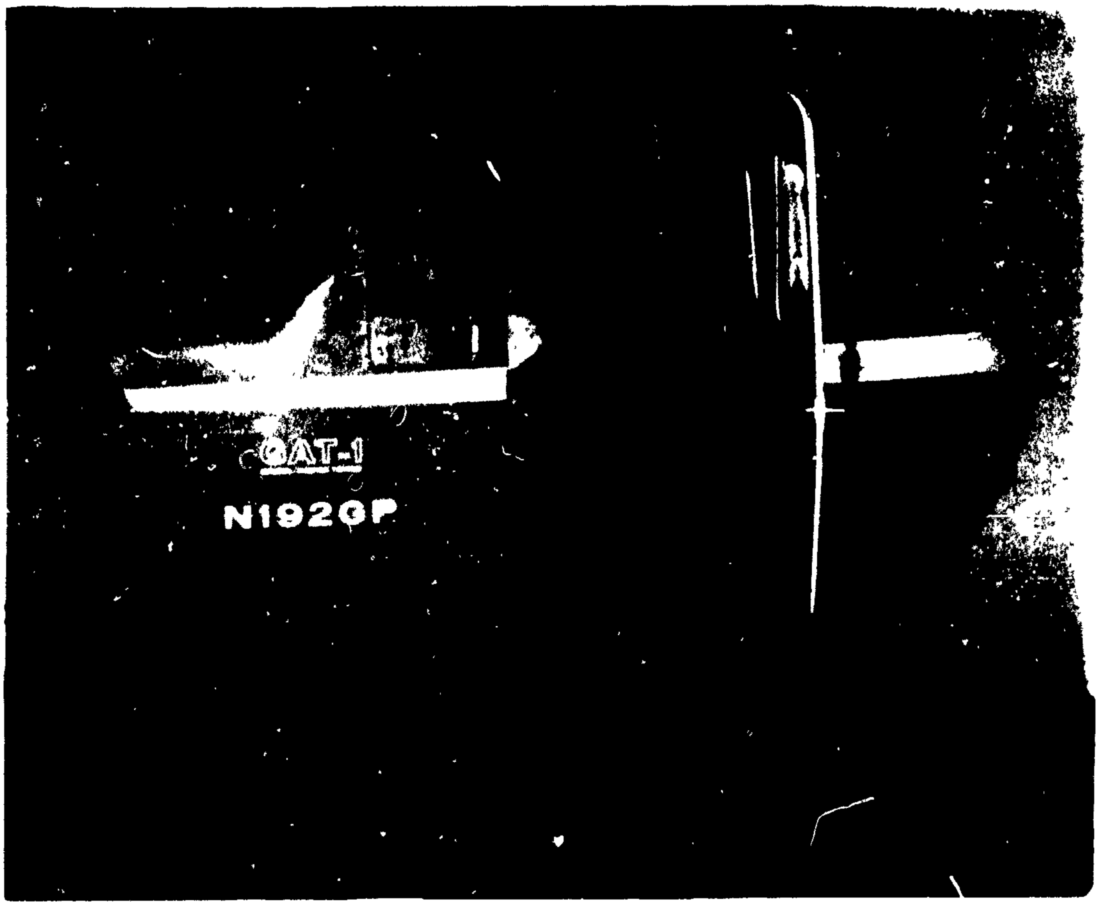


Figure 2. The GAT-1 moving-base trainer. (Note DIPT response panel on lower right-hand quadrant of instrument panel, above which the flight director panel is mounted.)

Panel (Model No. BD540832-200; Burroughs Corp., Plainfield, N.J. 07061) mounted directly over the instrument panel in each GAT-1. The display was 8.6 cm high x 23.3 cm long; each character was formed by a 5 x 7 dot matrix, 5 mm wide x 7 mm high.

Along with every message, excluding the DIPT stimulus displays, a countdown index was displayed on the FDP which decremented 5 counts every 5 sec. This index indicated the total time remaining before the displayed message would change. The countdown index was actually the total time remaining, multiplied by 2. Thus the subject could mentally divide the number displayed by 2 to ascertain how much time he had remaining for a given maneuver. A clock with a second hand was located on the instrument panel, and could also be used to time the maneuvers.

Each maneuver was allotted only the exact time required for its completion. For example, a standard rate turn of 270° required 90 sec; an altitude change of 500 ft at a vertical velocity of 250 ft/min required 120 sec. In order to alert the subject that a new message was about to be displayed on the FDP, a 95 dB alarm sounded for 3 sec at the point when 5 sec remained for the currently displayed message.

The temperature was maintained at approximately 23°C (75°F) in each trainer room throughout the study. The noise for the engine sound for each trainer was adjusted to 80 dB A-scale, with the engine set at 2,000 rpm, altitude 0, brake ON, motion axes OFF. This scale permitted satisfactory communication between subject and technicians when the trainer was operated at full power. All windows of the trainer were covered with frosted plastic to force the subject to rely on instruments for direction and motion cues.

Rectal temperature data were recorded continuously on magnetic tape throughout all 4.5-hr flights. A thermistor probe (Model 401; Yellow Springs Instrument Co., Yellow Springs, Ohio 45387) was inserted 10 cm beyond the anal sphincter and worn by the subject during his entire 12-hr flight duty period. Subjects reported no serious complaints (only minor inconvenience) from wearing the probes for prolonged periods during the study.

The temperature resistance reading was converted to a voltage and amplified in each trainer. The reading was digitized for display to the nearest 0.01°C by a special unit designed and built by the USAF School of Aerospace Medicine. This unit was also used to place a set of calibration signals on the magnetic tape from each GAT-1 just prior to the beginning of each 4.5-hr flight. At certain intervals throughout the flight, the rectal temperature was recorded from the readout system as a check that the measurement system was functioning properly.

The subject's electrocardiogram (ECG) signal was continuously recorded for each 4.5-hr flight from a lead system corresponding to V5 placement. The signal was amplified in the trainer before being recorded on magnetic tape. At certain time intervals during each flight, the ECG signal was simultaneously played back from the magnetic tape to a strip-chart recorder as a check that the system was recording satisfactorily.

At the beginning of each SLT, FMT, and DIPT, the computer generated a 1-volt 300-ms square-wave pulse which was recorded on the magnetic tape with the ECG and rectal temperature signal. This spike pulse allowed precise matching of these data to the corresponding SLT, FMT, or DIPT.

In addition, a time code prerecorded on the magnetic tape was continuously displayed on a time-code translator to assist the technicians in making periodic checks of the physiological data recording system throughout each flight. The time code was later used as an aid in alining the spike-pulse channel to the behavioral events.

In order to permit unimpeded turns past 360°, all connections between the trainer and the computer, tape recorder, and intercom system were made through a set of slip rings mounted in the trainer. The ECG and rectal temperature data and spike pulses for all subjects on a given flight were recorded on the same 1-in. magnetic tape by means of a 14-channel tape recorder (Model No. FR1800-L; Ampex Corp., Redwood City, Calif. 94063).

The same flight commands were issued simultaneously to all subjects on a given flight. Scoring was also done in parallel for the SLT and FMT. The DIPT was administered serially to each GAT-1. The order followed the numbers assigned to each GAT-1 (Fig. 3). While another subject was performing the DIPT, the remainder had a blank FDP and held the course requirements of the preceding SLT. The scores for the SLT, FMT, and DIPT were printed out on the computer teletype and recorded in a data log by one of the experimenters. These data were also automatically recorded on paper punched tape. Provided in Figure 4 is a flow diagram of data links between the various components in the performance and physiological measurement system.

A complete description and listing of all computer software programs used in this study has been given by Threatt and Perelli (200). This information included not only the programs necessary to build and edit the SLT, FMT, and DIPT files for computer storage, but also a list of the maneuvers in all four sets of FMT's and the two sets of stimulus displays used by the DIPT adaptive program. Their report also contained a description and listing of the master program used to control presentation of flight commands, to score flying performance, and to present the DIPT. [Note: The PDP 12 is being replaced with a PDP 11L34 Computer, and all software documentation will be modified accordingly.]

Flight Plans

During each block, lasting 8 days, subjects flew two 4.5-hr flights during each flying crew duty period. Four unique flight plans--A, B, C, D--were constructed, each having the same general flight profile and equal difficulty and workload. All subjects flew their 8 flight plans in the following order: A, B, C, D, A, B, C, D. Due to the nature of the flight plans, subjects could not learn to anticipate the specific requirements of each upcoming maneuver or course change. Subjects were not given any information about the specific flight plans during the course of the study.

A flight plan was broken into 4 legs, each containing 3 SLT's. The legs were constructed so that the cross pattern in Figure 5 was created. The order in which the legs were flown established the 4 (namely, A-D) basic flight plans—

A: 1, 2, 3, 4;
B: 3, 4, 1, 2;
C: 2, 1, 4, 3; and
D: 4, 3, 2, 1.

Each flight plan started with a different leg and, thus, a different heading. The first two flight plans (A and B) followed the cross pattern in a clockwise direction; the second two (C and D) created a cross-pattern traversed in a counter-clockwise direction. Note that the order of SLT's within the leg (Fig. 5) was always: a, b, c. Airspeed for each SLT was always 90 mph. Each heading, 90°, 180°, 280°, and 360°, was held three times per flight plan. Each altitude--2350, 2450, 2650, 3250, 3650, and 3750 ft--was held in two SLT's per flight. These altitudes were chosen and arranged so that it would be possible randomly to assign selected altitudes to the FMT's and to transition from one SLT to another, while remaining within the times allotted to the basic flight plan. The necessary commands for each flight plan were coded and then stored in the computer.

Each SLT consisted of two commands: A NEW COURSE command, containing heading, altitude, and airspeed information which the subject was to attain within a 2-min period--followed immediately by a HOLD COURSE command, containing these same flight parameters, which the subject was to maintain for the next 10 min. Scoring began as soon as the HOLD COURSE message was displayed, and continued until the panel message changed.

Sixteen FMT's were constructed, each containing the eight flight maneuvers listed in Table 1. Four sets of four unique FMT's were constructed by: (a) randomizing the order of the presentation of the eight maneuvers; (b) randomly assigning one of three airspeeds to each maneuver; and (c) counterbalancing turn direction for the 180° and 270° turns within each set. The 360° level turns were counterbalanced for direction within an FMT. The order of the four FMT's within a set was then randomized.

Within each set, the start of each FMT was randomly assigned one of four altitudes (2750, 3000, 3250, or 3500 ft) and headings (90°, 180°, 270°, or 360°). Each FMT ended at the same altitude and heading at which it had started. Maneuver commands for each FMT were constructed to take into account the changes in heading and altitude generated from one maneuver to the next. The commands for each set of FMT's were then stored in the computer as four separate data files.

Each maneuver of the FMT consisted of two commands: The first was a PREPARE FOR MANEUVER command, containing heading, altitude, and airspeed information which the subject was to attain and hold. This order was immediately followed by an EXECUTE MANEUVER command, containing the previous airspeed, the vertical velocity change (if required), direction and magnitude of turn (if required), and the final altitude and heading to be attained.

Each set of four FMT's was randomly paired with one of the four flight plans. The computer program inserted the four FMT's within a set sequentially

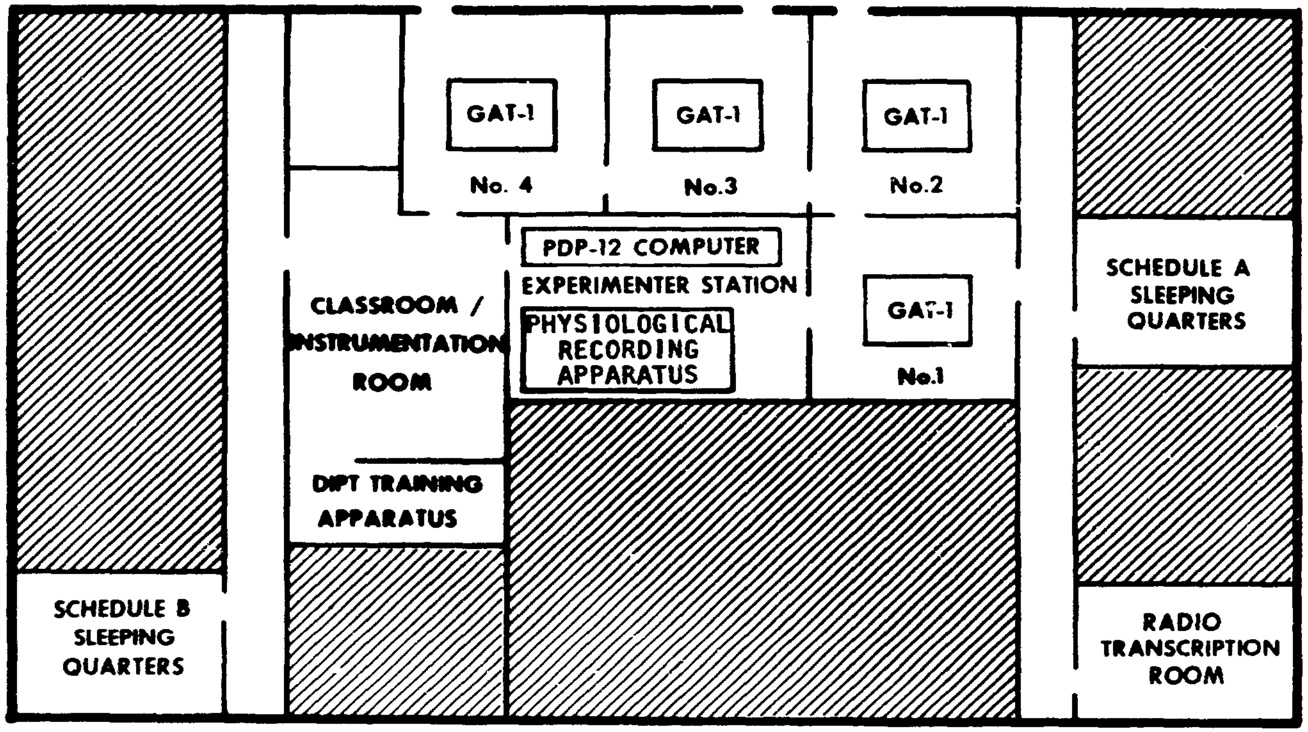


Figure 3. First-floor layout of experimental facility. [GAT-1: moving base trainer, Model B63000; DTPT: Discrete Information Processing Test]

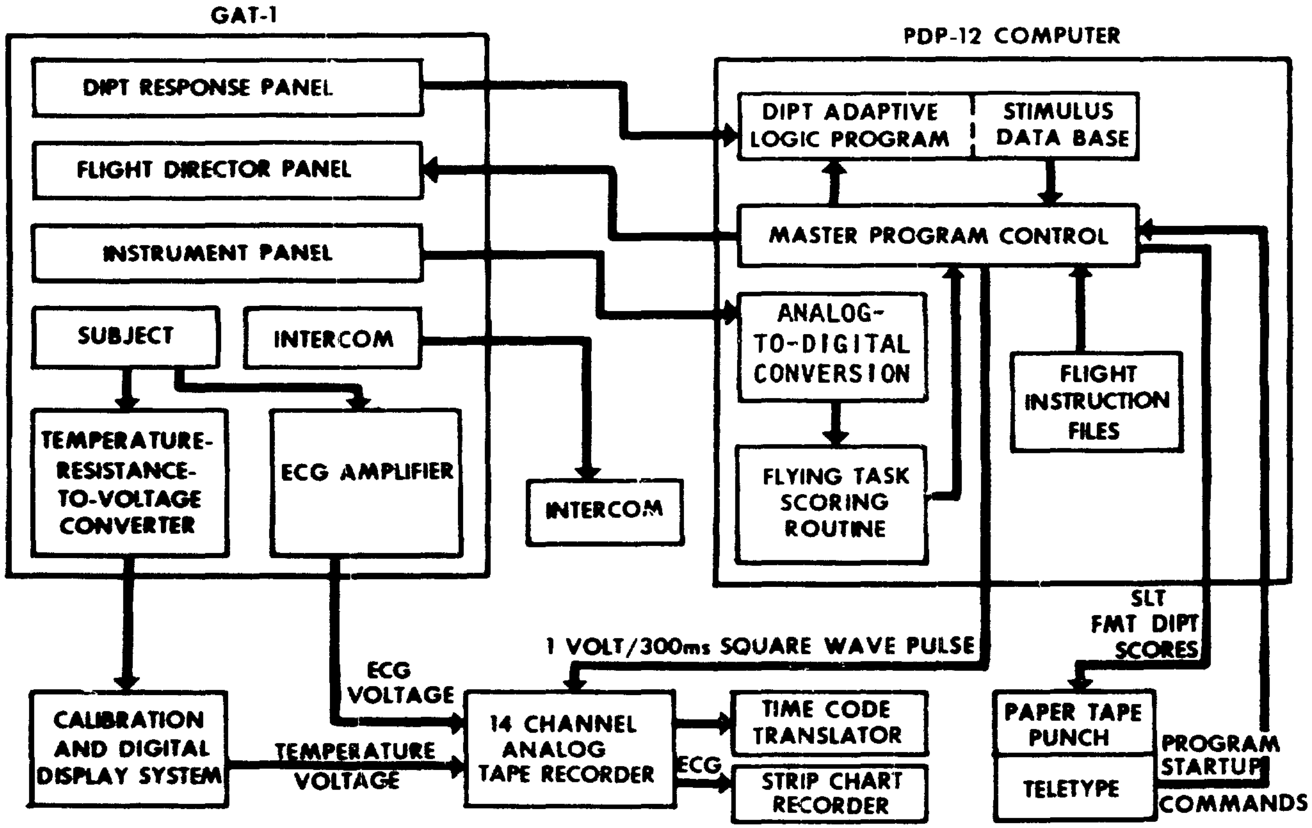


Figure 4. Linked components of the GAT-1, computer, and physiological measurement and recording systems. (All four GAT-1's were interfaced in the same manner.) [DIPT: Discrete Information Processing Test; GAT-1: moving base trainer, Model B63000; SLT: straight and level test.]

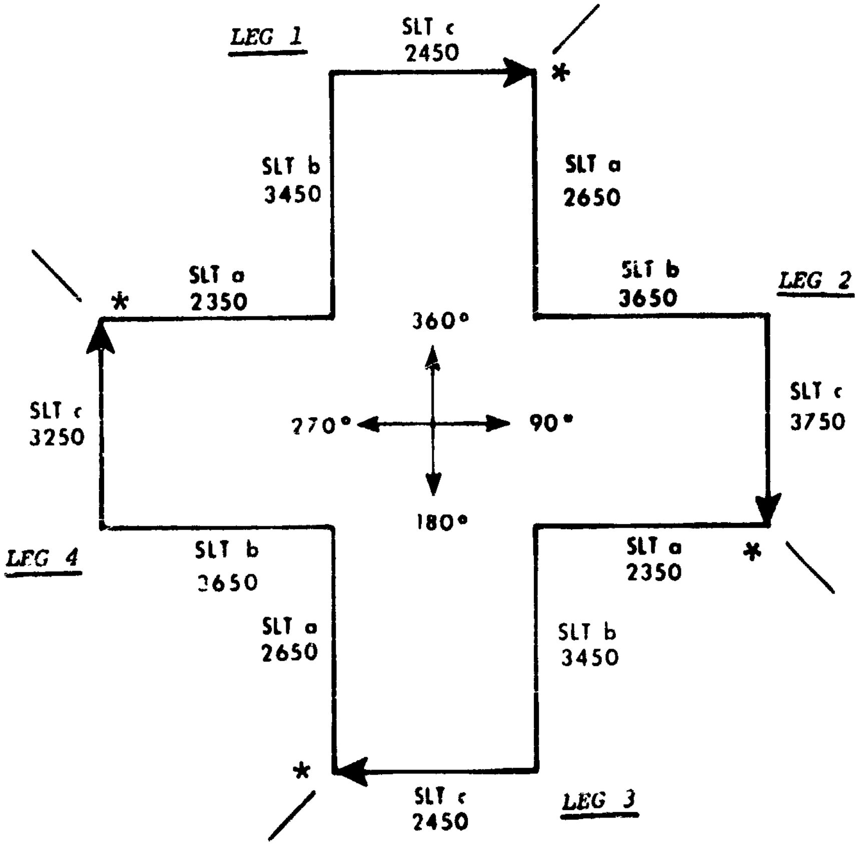


Figure 5. General flight plan formed from basic legs (1-4) of straight and level tests (SLT's: a - c), respectively.
[Note: Airspeed for each SLT = 90 mph; altitude, in feet, is given beneath each SLT; direction flow is in reference to center compass rose; and * is the starting point for each leg, regardless of the sequence in which the legs were flown.]

after SLT's 1, 4, 7, and 10 of the flight plan. The computer administered the DIPT after SLT's 2, 5, 8, and 11. The overall sequence of events for a 4.5-hr flight plan is presented in Figure 6.

For each SLT, the subject was given sufficient time to adjust his heading, altitude, and airspeed to the required values before the scoring began. Thus a perfect score was theoretically possible. For the FMT, however, even though the subject was given sufficient time to prepare for each maneuver, he needed about 3 sec at the start of the EXECUTE MANEUVER in order to adjust the vertical velocity and/or the turn rate indicators to the required values. Thus, the faster the subject responded to the maneuver command, the less his error score would be; but a perfect score was impossible, because the maneuver was scored from the start of the EXECUTE MANEUVER command.

Subjects who reached a new altitude or heading before the end of the allotted time for the maneuver (that is, before the panel changed to the next display) would be penalized if they then held that value, because the vertical velocity or turn-rate indicator would have to be placed in error to do so. The subject had the option of continuing the altitude change or of turning past the required value; but, then, he might not have time to regain the required altitude or heading in preparation for the next maneuver. In that case, poor performance on one maneuver could carry over to the next.

All turns were required to be standard-rate turns. During all maneuvers and SLT's, the ball of the turn coordination instrument was supposed to be kept centered between the "COORDINATED TURN" markers. During all descents at less than 90 mph, subjects were to use full flaps. Failure to do so was not scored directly; but the trainer is more difficult to control without use of flaps in these conditions, and an overall higher error score would result. In all other conditions, the flaps should have been up. The trainer could climb 90 mph at 500 fpm with full flaps. Thus, the error score might be higher if the subject failed to raise flaps at the conclusion of a descent. The three maneuvers involving descents which could require the use of flaps were randomly assigned airspeeds such that only 2 in each FMT had airspeeds less than 90 mph.

DIPT Stimulus Displays

Examples of the stimulus patterns used in DIPT displays are given in Figure 7. Numbers 1 to 5 appear randomly in the pattern on the left side of the display. One of 5 symbols appears on the right of the display: #, *, >, ~, or [. Moreover, 120 possible number patterns can be generated with 5 numerals. When paired with 1 of the 5 symbols, 600 unique stimulus patterns are created.

Two sets of randomized displays, containing subsets of 4 data files corresponding to the 4 DIPTs administered in each 4.5-hr flight, were constructed in the following manner. First 300 of the unique displays were randomly assigned to the first set; the remainder were assigned to the second. Next, each group of 300 randomly assigned to 1 of 4 subsets--75 to each subset. Then 10 from the total display set were randomly selected and assigned to each subset for a total of 85 displays per subset. The random selection process was arranged so that each symbol and each correct response location on the

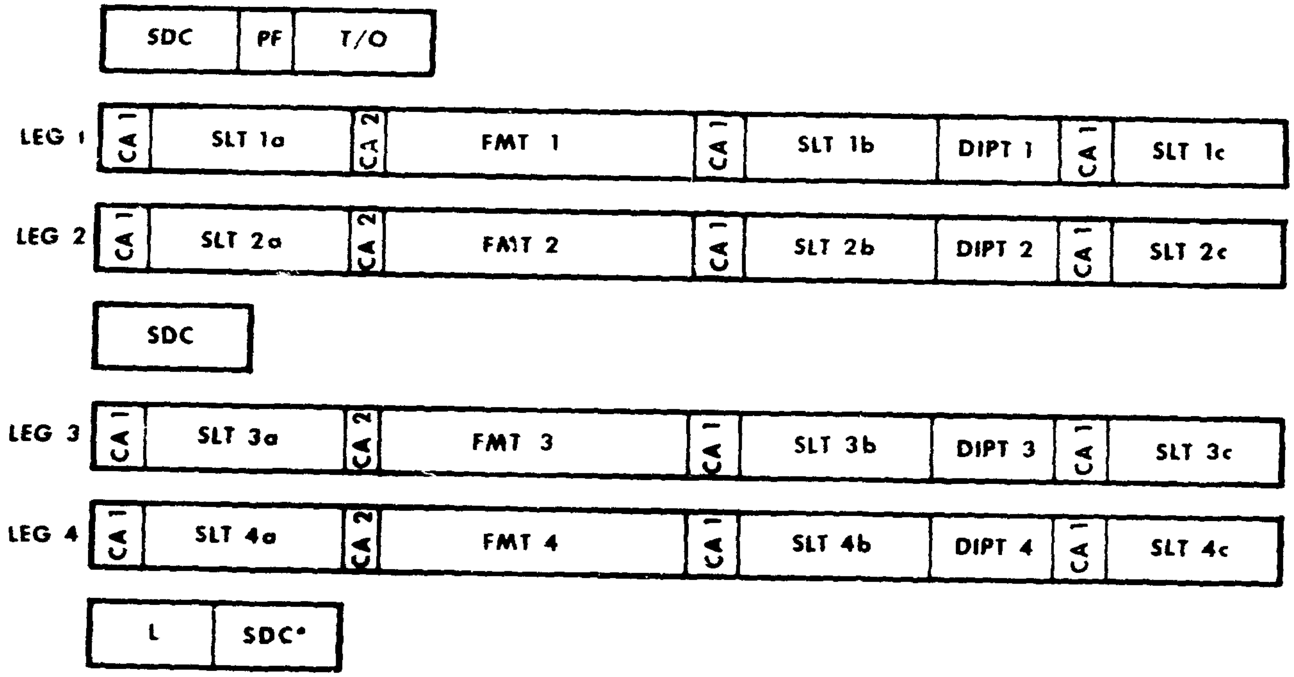


Figure 6. Sequence of events for the 4.5-hr flight plan.

CA 1 = course adjustment: 2 min
CA 2 = course adjustment: 1 min
DIPT = discrete information processing: 6 min (1.5 min/subject max)
FMT = flight maneuver test: 16 min
L = landing: 6 min
PF = preflight: 2 min

SDC = subjective fatigue and SSS data collection: 6 min
SLT = straight and level flying test: 10 min
T/O = takeoff: 6 min

*Collected only at end of second flight of flying crew duty day

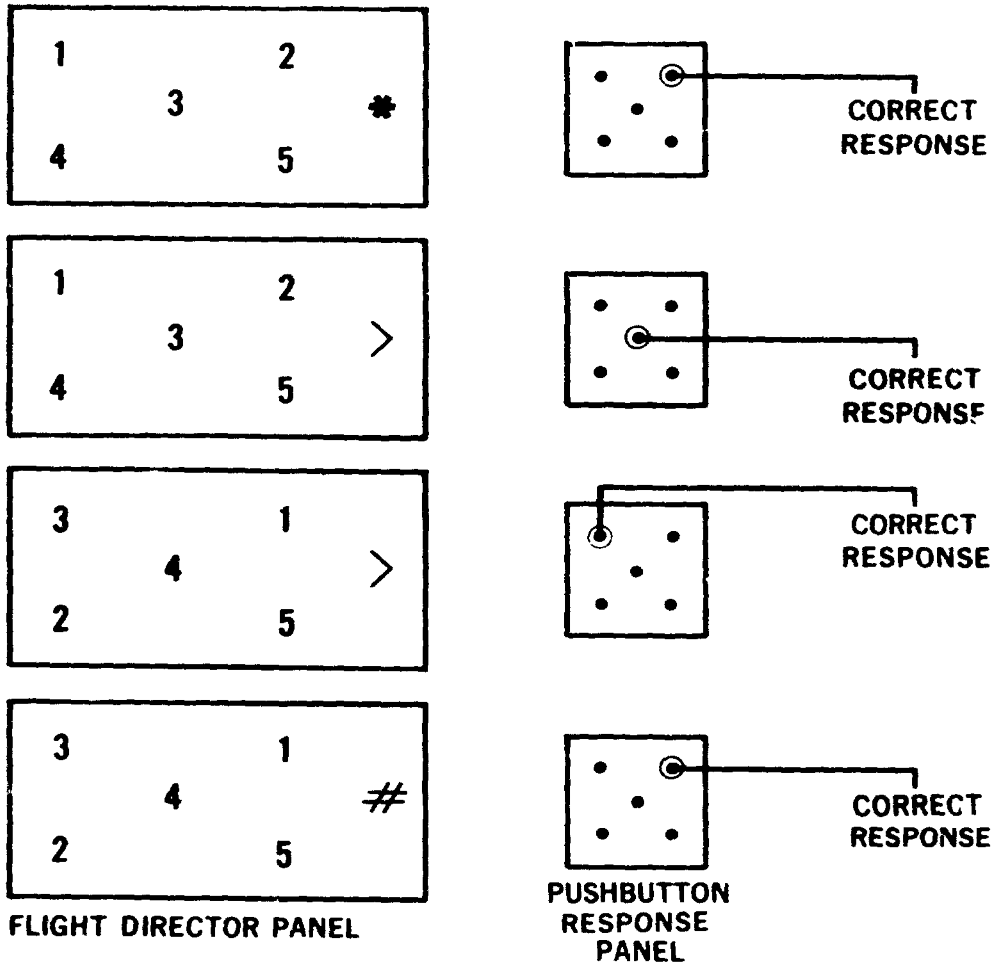


Figure 7. Examples of stimulus patterns used in DIPT displays. (Correct response is the pushbutton corresponding to the number position associated with displayed symbol.)

pushbutton panel occurred 17 times in each data file of 85 stimulus patterns. Additionally, the data file was inspected and, when two identical symbols appeared in a row, the second was randomly assigned to a new location in the list. This step was taken in order to help the subject recognize that a new trial was being presented.

Both sets of files were stored in the computer. The first set was always used with the first flight of the crew duty period, and the second was always used with the second flight. The file was repeated as necessary during the DIPT segment of each leg. Presentation started at the point in the file where the last subject ended.

The subjects were required to memorize the following number-symbol associations: 1-#, 2-*, 3->, 4-~, and 5-[. Five unlabeled pushbutton microswitches were mounted on the right center of the instrument panel in the same pattern as the numerals on the left side of the stimulus pattern. The subject's task was to press the pushbutton which corresponded to the position of the numeral associated with symbol displayed (Fig. 7). The subject was instructed to continue responding to the task as rapidly as possible, as long as new displays were presented.

In addition, the subjects were instructed to respond only when they knew the correct response. A subject learned that wrong answers slowed the presentation rate, and that the task would not end until he caused the rate to accelerate to the point at which he was unable to keep up with it when attempting to give correct answers.

In order to alert the subject, a ready panel containing the 5 numeral-symbol associations was displayed for 10 sec prior to the start of the DIPT. While performing the DIPT, the subject still had responsibility for flying the GAT-1; but no control responses were required and flying performance was not scored. The subject learned to ignore the flying task during this short period while engaged in the DIPT.

Training

All subjects, in both groups of 15, received an identical training program for operating the GAT-1 and performing the DIPT. The training session lasted 7 consecutive days, from 0800 to 2200 hr. Each student received 25 hr of classroom instruction and academic testing, and 15 hr of controlled practice performing standard flight maneuvers.

Practice time was equally divided over all four GAT-1's in order to familiarize the subjects with minor uncontrollable differences. This method permitted any subject to be randomly assigned, to any GAT-1 and to either schedule at the start of the experimental conditions, without the possibility of subjects having unequal experience on the assigned GAT-1's.

Days 1 and 2 served as a general introduction to the theory of flight and operation of the GAT-1. During this period, the subjects received two 1.5-hr practice sessions per day flying the GAT-1. The remainder of the training days were devoted to teaching the subjects to interpret and perform flight maneuvers being presented on the FDP, to understand the error tolerances for each instrument that was being scored, and to improve their overall skill in controlling the GAT-1 during precision maneuvers.

On days 3, 4, and 5, subjects flew two 1.5-hr training flights per day. A separate training-flight plan, increasing in difficulty and workload and containing segments resembling SLT's and FMT's, was used on each day. The flying performance scores generated were recorded but not analyzed statistically. Academic training also continued on these days.

On day 6, all subjects flew a 1-hr scored flight which was used to estimate their progress. Subjects whose scores were outside of acceptable limits, based on the group average, then received from 1 to 4 hr of additional practice that day. Subjects did not have to have equal flying scores or identical hours in the GAT-1, within limits, in order to participate in the study. An important requirement, however, was that their flying skill be close to reaching a plateau; otherwise, learning effects might overshadow fatigue-induced performance decrement. While individual differences in flying skill were inevitable, the author felt that individual differences in reaching a plateau could be reduced by providing additional training to those with poor scores, the assumption being that they were farther from plateau than others.

On day 7, all subjects flew a final 3-hr test flight with ECG leads in place. This score was used to determine the 12 most qualified subjects in each group for use in the study. Subjects also took a final written exam to demonstrate that they understod all aspects of the training program.

During days 1 and 2, all subjects received 1.5 hr of spaced practice per day on an apparatus which presented a task similar to the DIPT. The apparatus was located in the rear of the classroom. A slide projector displayed stimulus patterns identical to those used in the DIPT for periods of from 1 to 4 sec, with 1 sec between presentations. While not adaptive in nature, this practice served to teach the associations between numeral and symbol, and gave the subject experience in manipulating the response panel.

On day 3, all subjects performed the DIPT twice while sitting in the GAT-1 at the conclusion of the training flight. On days 4 and 5, the subjects performed the DIPT 4 times while flying the GAT-1. On day 6, the DIPT was not practiced. On day 7, subjects performed the DIPT 3 times during their final check flight.

While the training program covered four 3-hr blocks per day, not all subjects were in training at the same time. Subjects were assigned to one of four subgroups and the time of training for each subgroup was counterbalanced throughout all training days. Total amount of training time per day per subject was--day 1: 9 hr; days 2 to 5: 6 hr; day 6: 1 to 5 hr; and day 7: 7 hr.

Procedure

The sequence of events for the 25 days of the study for each group is presented in Figure 8. Subjects were assigned to the experimental conditions as follows. For each of the 3 blocks conducted per group, 4 subjects were randomly selected from the 12 most qualified. Next, 2 subjects were randomly assigned to each schedule condition, A or B. Finally, the four GAT-1's were randomly assigned to the 4 subjects for the duration of the block. Thus the total was six blocks, with four subjects (two in each condition) per block; and a total of 12 subjects per condition.

Subjects did not know if they were selected for the study until the day before their baseline data collection day. Furthermore, subjects did not know the length of the mission in which they would participate. They were told only that it would last from 5 to 8 days followed by 2 days of recovery, and that they might be required to stay awake continuously for up to 36 hr. They did know that crew rest periods would last 12 hr and that they would be permitted 8 hr of uninterrupted sleep.

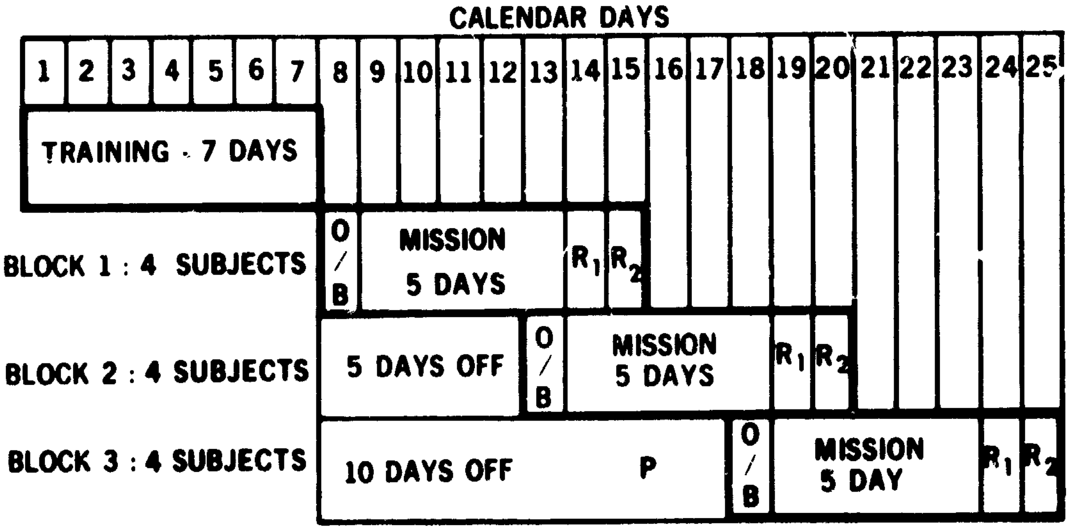


Figure 8. Sequence of events for training and data collection periods for three blocks of subjects. (The study adhered to this plan for both groups of subjects.) [O = time off; B = baseline; R = recovery; and P = 0.5-hr flying practice for skill level maintenance]

During the period when the subjects were not participating in the 8 days of data collection, they were permitted uncontrolled free time. They were required to remain in the local area, be back at the experimental facility by 2200 hr each day, and be at the breakfast formation each morning. During each block, subjects selected for that block were permitted uncontrolled free time on baseline and recovery days, with the stipulations that they refrain from vigorous physical activity and remain available for subjective data collection and urine samples at the required times.

The first block started 1 day after the training session. The second and third blocks started baseline on the last flying day of the previous block; so their blocks started 5 and 11 days after training, respectively. Two days prior to starting baseline, subjects in block 3 of each group received proficiency training for 0.5 hr in order to reduce loss of skill due to forgetfulness. All subjects were given a standard 10-min flight-refresher briefing by the author immediately before their first flight of the block.

The sequence of events for Schedules A and B is presented in Figure 9. Each crew received 12-hr crew-rest periods, with the middle 8 hr being designated for sleeping. The middle 5 days, designated the mission, contained 5 crew-rest periods in Schedule A--but only 4 in Schedule B. Crew-rest periods lasted from 2000 to 0800 hr, except for the third crew-rest period during the Schedule B mission, which lasted from 0800 to 2000 hr.

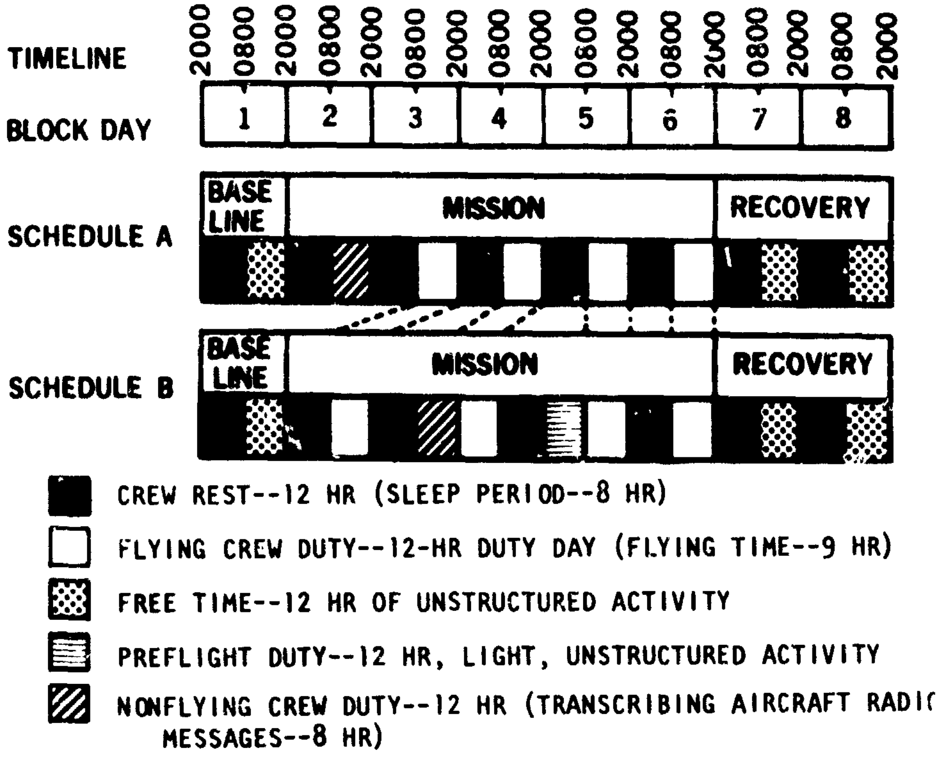


Figure 9. Sequence of events for subjects during one block for the two 8-day experimental schedules.

Each crew had 4 flying crew duty days lasting 12 hr, which included two 4.5-hr flights per day separated by a 1-hr meal break, making a total of 36 flying hours per subject per block. The remaining 2 hr of a flying crew duty day were taken up with preflight and postflight instrumentation procedures.

Schedule A contains four 12-hr crew duty days within the mission. Each crew duty day started at 0800 hr and ended at 2000 hr. Schedule B contains a 12-hr crew duty period, followed by two 24-hr crew duty periods, and concluded by a 12-hr crew duty period. Their flying crew duty day started at 0800 and lasted until 2000 hr, except for the second flying crew duty day which started at 2000 and lasted until 0800 hr.

Instrumentation of the subjects began at 0800 or 2000 hr, depending on the schedule. Subjects provided a urine sample immediately prior to entering the GAT-1's. For the first flight, takeoff was scheduled as close as possible to 0900 or 2100 hr; and landing occurred as close to 1330 or 0130 hr as possible. Takeoff for the second flight occurred as close as possible to 1430 or 0230 hr, and landing occurred as close as possible to 1900 or 0700 hr. Note that flying crew duty days for both schedules occurred at the same time for flying days 3 and 4 (block days 5 and 6), at which time all 4 subjects of a block flew together (Fig. 9).

The nonflying crew duty activity consisted of transcribing aircraft radio transmissions, each pair of subjects working as a team. No data were analyzed from this task. Subjects remained in the transcribing room (Fig. 3) except for short breaks and trips to the dining hall for breakfast, lunch, and dinner. The preflight duty for subjects in Schedule B consisted of light, unstructured activity: watching television, reading, or playing cards. Their activity was monitored by an experimenter.

In all cases, subjects were not permitted to smoke while in the GAT-1's. Alcohol consumption was not permitted from the day prior to the start of baseline measures until the completion of the second recovery day.

During baseline and recovery periods, subjects ate all meals at the base dining hall located approximately 0.5 miles from the experimental facility. During the 5 calendar days of the long-duration mission, subjects ate only breakfast at the base dining hall. All other meals (including lunch, dinner, and the night meal) consisted of a standard menu of frozen dinners prepared and consumed in a kitchen on the second floor of the experimental facility. Coffee and caffeine-containing drinks were permitted ad lib.

During training, the first group of 15 subjects were housed in the experimental facility; the second group of 15 was housed in a dormitory 0.5 miles from the experimental facility. During the data collection phase of the study, all subjects lived on the second floor of the experimental facility except for the 4 participating in the current block. They were housed, 2 to a room, on the first floor, adjacent to the trainer rooms (Fig. 3). Subjects on the same schedule roomed together. Rooms were essentially identical but were paired to the same schedule. One room, which was subject to less noise during the day than the other, was always assigned to Schedule B; for those subjects had one crew-rest period from 0800 to 2000 hr and would thus not be disturbed. Subjects were housed on the first floor from the first night of baseline until the completion of their last flight. That night they were moved back to the second floor for recovery.

Subjects not participating in the current block of data collection were not allowed in the first-floor experimental area.

Subjective fatigue and Stanford Sleepiness Scale (SSS) cards were filled out at 0845, 1330, 1900, and 2200 hr for the entire 8-day block. At these same periods, urine samples were collected, acidified with a 15-ml solution of 1.6 normal hydrochloric acid, and immediately frozen. Note, however, that these data were <u>not</u> collected from subjects in Schedule B on block day 4; for this was their assigned sleep period. Thus the number of urine samples from subjects: in Schedule A, totaled 32; and in Schedule B, 28.

Subjects awoke at 0600 hr and immediately voided. Thus, the 0845 urine sample was produced between 0600 and 0845 hr. This procedure was employed because overnight urine samples tend to vary greatly between individuals.

All subjects filled out a sleep survey form during the first data collection period after waking. This period was at either 0845 or 2045 hr.

During the flying duty period, subjective fatigue and SSS data were collected while the subject was in the GAT-1: during preflight and midflight of the first flight of the day; and during preflight, midflight, and postflight of the second. These five data-collection periods corresponded to 0845, 1115, 1330, 1545, and 1900 hr for the day flights, and to 2045, 2315, 0130, 0345, and 0700 hr for the night flight on day 2 of Schedule B.

The subjective reports of fatigue and sleepiness were transmitted from the GAT-1 to the experimenter via intercom. At the required time, the experimenter read the Subjective Fatigue Checkcard and SSS to each subject in turn, according to the order of DIPT testing, and recorded his responses.

A standard set of takeoff commands brought each subject to an altitude of 2400 ft, a heading of 90°, and an airspeed of 90 mph. The program was halted until all subjects reported that they had attained these values. The program was then continued, and was not halted again until the end of SLT 2c for a 6-min subjective data-collection period. The flight was then continued under computer direction until all subjects reported their landing.

Each subject was monitored during the entire flight via intercom, but no other planned communication took place between subject and experimenter. Subjects could not communicate between GAT-1's, and did not overhear communications between the experimenter and other subjects.

At least two experimenters were required at all times to instrument and monitor subjects, run the computerized flight performance and physiological measurement system, compile mission logs, and record data. In addition, an experimenter was present in the experimental facility at all times when subjects were in the building.

Reduction of Physiological Data

The rectal temperature voltage was filtered at 50 Hz, sampled every 5 sec, and digitized. The ECG signal was converted to an instantaneous beat-to-beat HR by means of a cardiotachometer circuit implemented on an EAI-680 Analog Computer (Electronic Associates, Inc., Long Branch, N.J. 07764). The output voltage of this circuit changed with each beat in proportion to the instantaneous beat-to-beat HR. Noise, appearing as HR less than 30 bpm or greater than 200 bpm, was filtered out of the signal. This voltage was then sampled every 5 sec and digitized.

A PDP-12 computer was used to adjust the digitized rectal temperature value, based on the calibration signal read from the magnetic tape, and to convert it to degrees Celsius. Rectal temperature and HR values were then grouped--according to the scored portion of the SLT, FMT, or DIPT epoch--when they were produced, by means of the edit pulses read from the magnetic tape. Mean HR and rectal temperature values and the standard deviation for HR (HRV) were then computed for each epoch.

Urine samples were analyzed by the USAFSAM Crew Technology Divison. Sodium and potassium values were determined by means of flame photometer (Model No. 143, Instrumentation Laboratory, Inc., 113 Hartwell Ave., Lexington, Mass. 02173). Creatinine and urea values were measured by means of an autoanalyzer (Technicon Instruments Corp., 511 Benedict Ave., Tarrytown, N.Y. 10591). Liquid column chromatography employing the Porter-Silber reaction (Silber and Porter: 189) was used to determine 17-OHCS concentrations by means of a spectrophotometer (Model B, Beckman Instruments, 2500 Harbor Blvd., Fullerton, Calif. 92634). Epinephrine and norepinephrine levels were determined by a trihydroxyindole autoanalyzer method developed by USAFSAM using a Technicon fluorometer. This method is based on the work of Euler (69).

RESULTS

Chemical analysis of the urine battery necessitated thawing and refreezing a large portion of the samples to create separate aliquots for each of the sodium, potassium, catecholamine, and steroid procedures. Visual inspection of portions of the resulting data showed marked differences between groups which were thawed multiple times and those thawed only once. In the past, thawing and refreezing were thought to have no significant effect on the final values of the substances being analyzed; but this procedure apparently has altered some of the values in a nonlinear fashion, so that drawing conclusions from the statistical analyses would be unwarranted. Thus, none of the results for the urine analysis have been reported. Hypothesis 7 could not be tested.

All statistical analyses were performed on the San Antonio Data Services Center computer using the "Statistical Analysis System" program (Barr and Goodnight: 12).

Where multiple analyses are presented on one table, the degrees of freedom (df) for all analyses are the same as in the left-hand df column, unless data were missing, in which case a specific df column is provided.

The following events account for most of the missing data. For Schedule A, equipment problems precluded data collection for two subjects in different blocks during Period 1 on Flight Day (FD) 4. Their data were also dropped for FD 4: Period 2 (even though these data had been collected).

A computer problem caused the loss of all four subjects' SLT 1 performance data for FD 3: Period 2, Leg 1, in Block 6.

For Schedule B, one subject became nauseated at 0430 hr, FD 2: Period 2, and was unable to continue that flight. He remained awake until the prescribed sleep period, had an uneventful day's sleep, and was able to complete the remaining flights with no ill effects. Two subjects failed to respond to the DIPT--one on FD2, Period 2, Leg 3; the other on FD 3, Period 1, Leg 2. The subjects reported that they momentarily nodded off to sleep and did not hear the alerting tone. Their HR, HRV, and rectal temperature values for DIPT epochs were also deleted. Recording problems necessitated deleting all HR, HRV, and rectal temperature data for one subject in Block 5.

In cases where the subject was unable to perform in the FAT-1 due to equipment malfunction, all appropriate data were deleted even though some were still collectable. For example, subjective fatigue data were discarded, because the reduced workload lowered the intensity of the fatigue stress the subject experienced.

In the following analyses of variance, where significant higher order interactions exist, simple main effects are ignored--and are interpreted only in the light of the interactions. Also, note that lower subjective fatigue scores indicate greater fatigue, but higher SSS scores indicate greater sleepiness.

Analyses of Fatigue Effects on Dependent Measures

To test the effects of the fatigue stressors on the SLT, FMT, DIPT, subjective fatigue, and SSS scores independently, a four-factor model was used: group (Schedule A or B); block (experimental units 1 to 6); period (first or second flight of the duty period); and leg (first, second, third, or fourth hour of the flight). There were repeated measures on the last two factors. Block effects were possible, due to the fact that the experimental sessions lasted over a 2-month period and subjects may have responded differently to uncontrollable differences between blocks (such as improvement of training techniques as the instructors became more familiar with the GAT-1 performance measurement system). However, Group x Block interaction was not considered likely, because block peculiarities would have affected subjects from both groups equally. Thus, this effect was pooled in the appropriate error term.

If blocking had an effect and it was not tested, then the true effect of the fatigue stressors could not be tested. Testing for block effects was more conservative, since the error term had fewer df's (17 vs. 23). But since the group x block interaction was pooled in the error term, the interaction error term yielded a less than maximally conservative test (17 vs. 12 df's).

Each block consisted of 8 days: block day (BD) 1 will be referred to as "premission baseline"; BD's 7 and 8 will be referred to as "Recovery Day 1 (R 1)," and "Recovery Day 2 (R 2)," respectively. Block days 3, 4, 5, and 6 correspond to FD's 1, 2, 3, and 4, for Schedule A; and BD's 2, 4, 5, and 6 correspond to FD's 1, 2, 3, and 4, for Schedule B (Figs. 8 and 9).

Group, block, period, and leg variables are considered fixed effects; subject is a random effect. Thus, a mixed model was used to determine error terms: appropriate interaction terms were used for main fixed effects; and residuals were used for main random effects and interaction tests.

To improve the efficiency of the analyses, FD's were not analyzed as a factor directly. Instead, four separate analyses were performed for each FD, using FD 1 as the reference. Although this technique increased the total number of analyses of variance performed, it had the effect of performing planned comparisons and identified the specific day on which decrement occurred, without requiring a post-hoc test. The FD 1 analysis was performed to insure that the groups were not significantly different, initially. The decrement scores, which were obtained by subtracting the subsequent FD from FD 1, means that each significant effect should be interpreted as an interaction between those two days and the variables in the source line. Except where noted, absolute values of scores were not of primary importance, but only a subject's change from FD 1. By using decrement scores, individual differences between subjects were eliminated, thus helping to control for skill differences between groups, if such occurred. Unless noted, in all following analyses, the SLT score used for each leg was the average of SLT: a, b, and c.

Hypotheses 1 and 2: Fatigue Effects on Flying Performance and Information Processing

The results of the analyses of variance for SLT, FMT, and DIPT performance scores for FD's 1 to 4 are presented in Tables 2 to 5.

TABLE 2. ANALYSES OF VARIANCE FOR SLT, FMT, AND DIPT SCORES FOR FLIGHT DAY 1

Source		SLT		FMT		DIPT	
	df	MS	F	MS	F	MS	F
Between Subjects (Ss)							
Group (G)	1	2520038.95	2.977	2428920.12	4.274	735075.000	4.051
Block (B)	5	1208268.36	1.427	1178259.88	2.073	259528.750	1.430
Ss within G,B	17	846523.14	—	568321.16	—	181440.074	—
Within Ss							
Period (P)	1	15205.95	0.162	89372.28	1.889	23852.083	1.188
G x P	1	7053.52	0.075	101807.34	2.152	101752.083	5.070*
B x P	5	117098.61	1.244	87375.08	0.636	38030.833	1.895
Ss x P within G,B	17	94138.37	—	47305.71	—	20070.098	—
Legs within P (L/P)	6	62434.83	2.617*	146328.65	5.377***	25728.819	0.877
G x L/P	6	27882.06	1.169	8419.81	0.309	7780.208	0.265
B x L/P	30	18188.41	0.762	17306.37	0.636	17819.236	0.608
Ss x L/P within B,G	102	23861.18	—	27216.22	—	29331.556	—

*p<.05
***p<.001
df = degrees of freedom
MS = mean square

TABLE 3. ANALYSES OF VARIANCE FOR SLT, FMT, AND DIPT DECREMENT SCORES FROM FLIGHT DAY (FD) 1 TO FD 2

Source	SLT			FMT			DIPT		
	df	MS	F	df	MS	F	df	MS	F
Between Subjects (Ss)									
Group (G)	1	1126207.864	2.294	1	1355755.064	3.392	1	2816.664	0.090
Block (B)	5	581306.066	1.184	5	225222.538	0.564	5	16542.042	0.527
Ss within G,B	16	491026.688	—	16	399669.730	—	15	31369.1924	—
Within Ss									
Period (P)	1	5225132.592	40.059***	1	1735955.004	9.360**	1	356484.376	10.063**
G x P	1	3758678.540	28.817***	1	1519913.100	8.191*	1	401709.376	11.339**
B x P	5	80555.446	0.618	5	149531.626	0.806	5	59186.417	1.671
Ss x P within G,B	16	130434.808	—	16	185565.148	—	15	35426.736	—
Legs within P (L/P)	6	208466.268	4.535***	6	91456.313	2.118	6	23839.377	0.505
G x L/P	6	233425.345	5.078***	6	105183.204	2.435*	6	20873.174	0.442
B x L/P	30	40909.349	0.890	30	28284.815	0.655	30	34170.523	0.723
Ss x L/P within B,G	100	45964.228	—	100	43187.909	—	99	47235.331	—

---

*p<.05
**p<.01
***p<.001

TABLE 4. ANALYSES OF VARIANCE FOR SLT, FMT, AND DIPT DECREMENT SCORES FROM FLIGHT DAY (FD) 1 TO FD 3

Source	SLT			FMT			DIPT		
	df	MS	F	df	MS	F	df	MS	F
Between Subjects (Ss)									
Group (G)	1	256106.00	0.995	1	1314364.080	6.177*	1	553.000	0.033
Block (B)	5	378342.69	1.470	5	144311.832	0.678	5	52386.888	3.124*
Ss within G,B	17	257410.19	—	16	212787.000	—	16	16766.936	—
Within Ss									
Period (P)	1	1146487.96	12.115**	1	218531.388	1.658	1	318529.412	4.222
G x P	1	770458.48	8.141*	1	54837.860	0.416	1	323297.120	4.285
B x P	5	102094.97	1.079	5	45774.544	0.347	5	40812.530	0.541
Ss x P within G,B	17	94634.77	—	16	131771.468	—	16	75451.676	—
Legs within P (L/P)	6	58227.42	1.566	6	70141.304	1.588	6	16304.019	0.253
G x L/P	6	44678.34	1.202	6	18219.127	0.413	6	20672.869	0.321
B x L/P	30	52330.64	1.408	30	62414.992	1.413	30	39460.544	0.613
Ss x L/P within B,G	102	37173.12	—	101	44162.948	—	101	64330.325	—

---

*p<.05
**p<.01

TABLE 5. ANALYSES OF VARIANCE FOR SLT, FMT, AND DIPT DECREMENT SCORES FROM FLIGHT DAY (FD) 1 TO FD 4

Source	df	SLT		FMT		DIPT	
		MS	F	MS	F	MS	F
Between Subjects (Ss)							
Group (G)	1	922886.441	3.386	42170.476	0.296	189840.076	4.055
Block (B)	5	638083.112	2.341	249506.906	1.751	80908.598	1.728
Ss within G,B	15	272546.696	—	142533.453	—	46817.189	—
Within Ss							
Period (P)	1	7410.888	0.063	123532.408	2.532	31645.455	0.498
G x P	1	7787.018	0.066	259221.443	5.314*	16663.712	0.262
B x P	5	176328.540	1.495	29263.185	0.600	63315.492	0.996
Ss x P within G,B	15	117970.939	—	48780.102	—	63554.225	—
Legs within P (L/P)	6	122533.139	3.341**	55838.524	1.238	32550.000	0.711
G x L/P	6	53441.436	1.457	16939.207	0.376	23425.833	0.512
B x L/P	30	33763.054	0.921	33208.855	0.736	47020.278	1.027
Ss x L/P within B,G	90	36676.194	—	45101.360	—	45777.630	—

---

*p<.05
**p<.01

Due to the volume of analyses generated by this experimental design, the complete partitioning of the between-treatment df's was not analyzed. The 2 periods and 4 legs yielded 7 df's: (P x L)-1. The period variable accounted for 1 df. The remaining 6 df's were pooled in a "leg within period" variable, which included: the leg effect (L-1) = 3 df's; and the leg x period interaction (L-1) (P-1) = 3 df's. In this study, the question of primary interest for between-treatment effects was: "Do the average leg values in period 1 differ significantly from the average leg values in period 2?" This question is answered by the period-effect test. The "leg within period" effect answers a question of secondary importance: "Did any significant differences exist between Legs 1 - 4 in Period 1, or between Legs 1 - 4 in Period 2?"

From Table 2, it appears that the two schedules did not differ significantly in SLT or FMT performance in terms of the minimal significance level required in this study, $p<.05$. However, the exact probability for the FMT approached this requirement ($p = .052$). Apparently, Schedule A's performance tends to be better initially than Schedule B's when viewed by periods (Fig. 10) or by legs (Fig. 11).

A possible explanation for these apparent differences is that subjects in Schedule B flew on the first day of the session, while Schedule A subjects had only to perform light nonflying duty. Subsequent analyses also found other slight differences on FD 1 for other measures. Possibly Schedule B subjects experienced a higher initial stress which affected their performance. By using decrement analysis based on each subject's FD 1 score, however, these differences were minimized. If, due to a failure of randomization, the two groups truly differed in their general ability to handle stress--and this factor interacted negatively with the fatigue stressors--then the values obtained on subsequent FD's might overestimate the decrement obtainable from the normal population. Both SLT and FMT scores show a strong "leg within period" effect. Thus the apparent drop in performance within periods is not significantly different between groups (Fig. 11).

The DIPT scores were also close to showing initial group differences ($p = .058$), and a significant Group x Period interaction ($p<.05$) occurred (Fig. 12). Schedule B's threshold of information processing was higher than that of Schedule A; but B's performance improved from Periods 1 to 2, thus indicating possible learning effects--while A's performance deteriorated slightly from Periods 1 to 2. A similar pattern of deterioration was noted within legs for both groups during Period 1, but not for Period 2 (Fig. 13); and no "leg within period" effects were found significant. On FD 2, continued improvement appeared to occur for Schedule B for the first three legs of Period 1, and their scores appeared comparable to those of Schedule A. However, Schedule B's performance deteriorated drastically thereafter, as was expected due to increasing fatigue during the night flight.

In Table 3, all three performance measures showed Period and Group x Period differences. SLT scores showed the most significant differences, $p<.001$, for the P, Group x Period, "Leg-within-Period," and Group x "Leg-within-Period" effects. Schedule B's SLT scores approached A's in Period 1; but, while A's error score showed a negligible drop in Period 2, B's more than doubled. SLT decrement scores are presented in Figure 14. FMT scores exhibited a similar pattern, but the Group x Period interaction was significant only at the 0.05

level. This finding may indicate that the FMT, due to its arousal value, is less susceptible to fatigue effects than the SLT as predicted by Hypothesis 9 (refer to the subsection on "The Statement of Hypothesis," in the "Introduction").

Magnitudes of FMT decrement scores (Figs. 10 and 14) are less than the SLT; but the absolute values are greater than the SLT scores, thus indicating the greater difficulty of the FMT. This pattern is true across all FD's.

As shown in Figure 11, a continual sharp drop occurs in SLT and FMT performance for Schedule B during Period 2, from Legs 1 to 3, followed by apparent plateauing. This halt in performance decrement could have been due to circadian performance effects: for Legs 3 and 4 occurred from approximately 0500 to 0700 hr, a time period when performance might be expected to improve.

The DIPT scores indicate strong period and Group x Period effects ($p = .006$ and .004, respectively); but, because of the dysordinal interaction between group, period, and flight day 1 to 2 (Fig. 12), Schedule B's high Period 1 (P1) reference score could have been the major source of the apparent fatigue effect. To investigate this possibility, a separate t-test was performed for both the Schedule A and B period means. The Schedule A mean differences were nonsignificant (P1 = 779.4 ms; P2 = 797.5 ms; $\underline{t}$ (11) = .752; and $p = .468$). The Schedule B means were still significantly different (P1 = 798.2 ms; P2 = 928.7 ms; $\underline{t}$ (10) = 2.987; and $p = .015$). As shown in Fig. 13, DIPT scores appear to exhibit the same apparent decrement pattern as the SLT and FMT scores to plateau in Period 2.

The analysis for the decrement from FD 1 to 3 is presented in Table 4. SLT scores still indicate significant period and Group x Period effects, but the effects may not be as strong ($p<.01$ and $p<.05$, respectively). In Figures 10, 11, and 14, the patterns of decrement appear similar between FD 2 and FD 3; but the magnitude of decrement appears less.

The FMT scores indicated only a group difference, $p<.05$. The apparent period or Group x Period effect in Figure 10 may not have shown up because the scores were referenced to FD 1. The DIPT period and Group x Period scores only approached significance ($p = .054$ and .052); however, all differences were in the expected directions, and also of less magnitude than FD 2 (Figs. 12 - 14). For no apparent reason, a block effect ($p<.05$) also appeared. Block effects were not expected to occur and least of all on this measure. With no other supporting evidence, this effect was assumed to be due to chance.

The analysis for the decrement from FD 1 to 4 has been presented in Table 5. The only significant effect found for SLT scores was for "leg within period" ($p<.01$). According to Figure 10, Schedule B had recovered almost completely from the previous flight's fatiguing effects; and some learning may have occurred, since the FD 4 scores were lower than the FD 1 scores for both groups.

The FMT scores demonstrated a Group x Period effect ($p<.05$) which seems (Fig. 14) to be primarily due to the fact that the Schedule B scores for FD 4, Period 2, were markedly improved over those for FD 1. Both groups have improved on FD 4 relative to FD 1, but Schedule B has improved more.

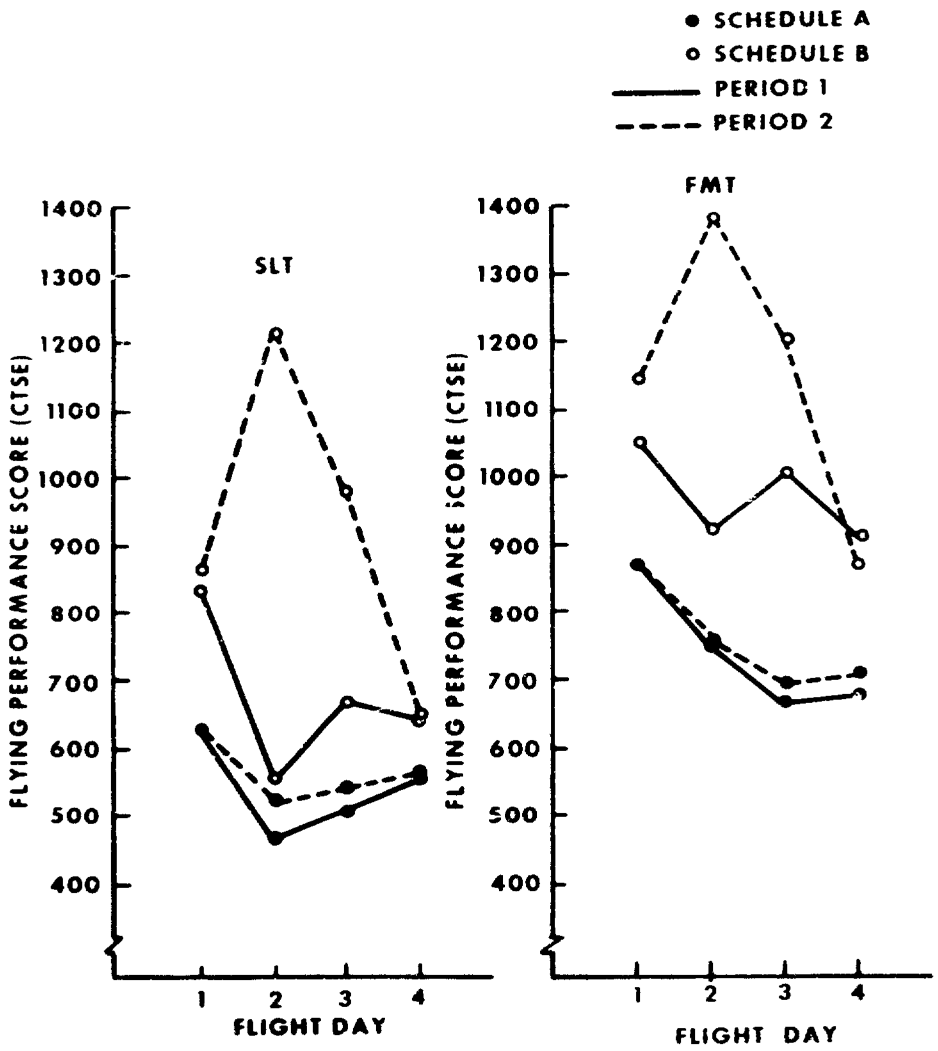


Figure 10. SLT and FMT performance scores for flight days 1 to 4, by period, for Schedules A and B.

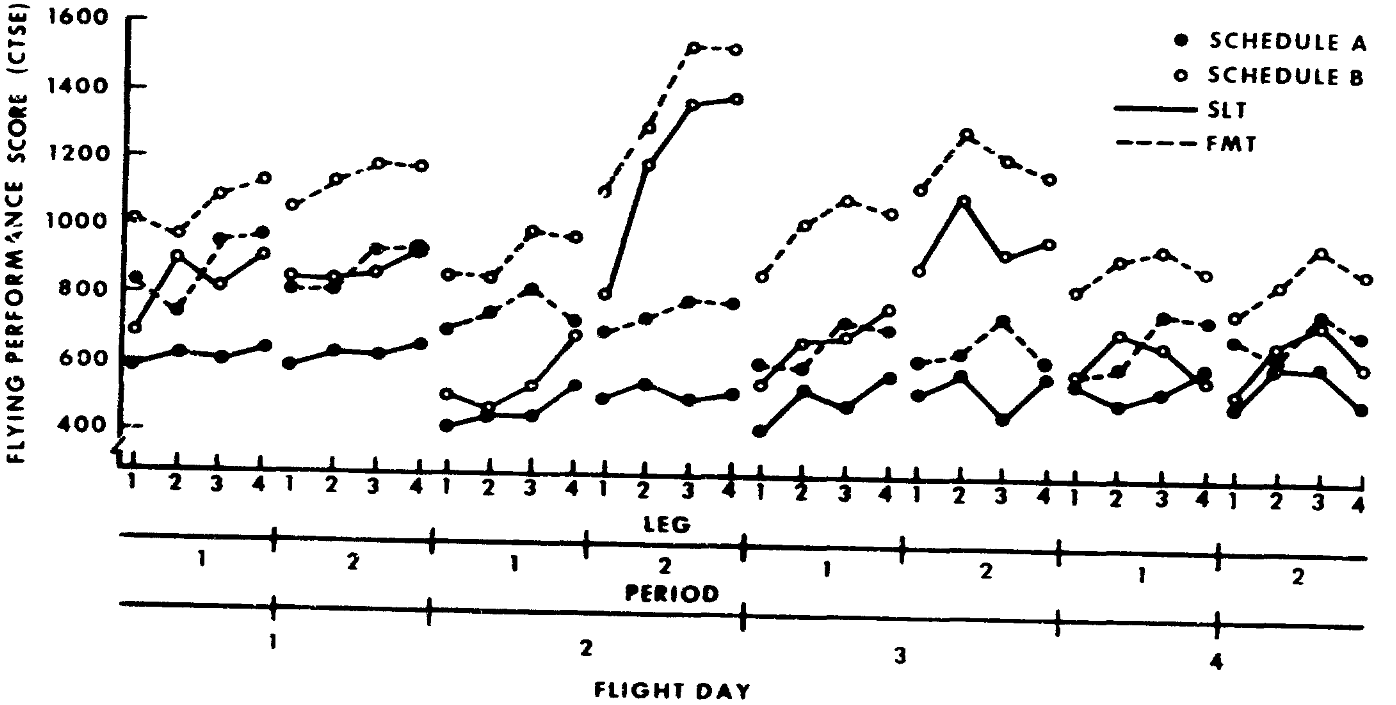


Figure 11. SLT and FMT performance scores for each leg of flight days 1 to 4. The three SLT scores obtained per leg have been averaged.

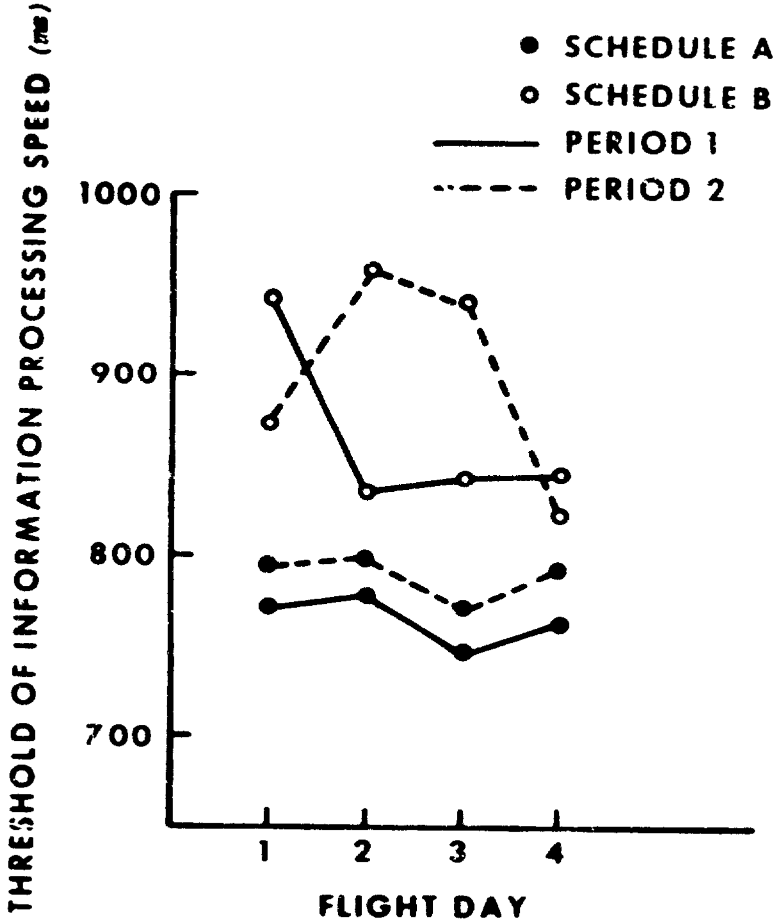


Figure 12. DIPT performance scores for flight days 1 to 4, by period, for Schedules A and B.

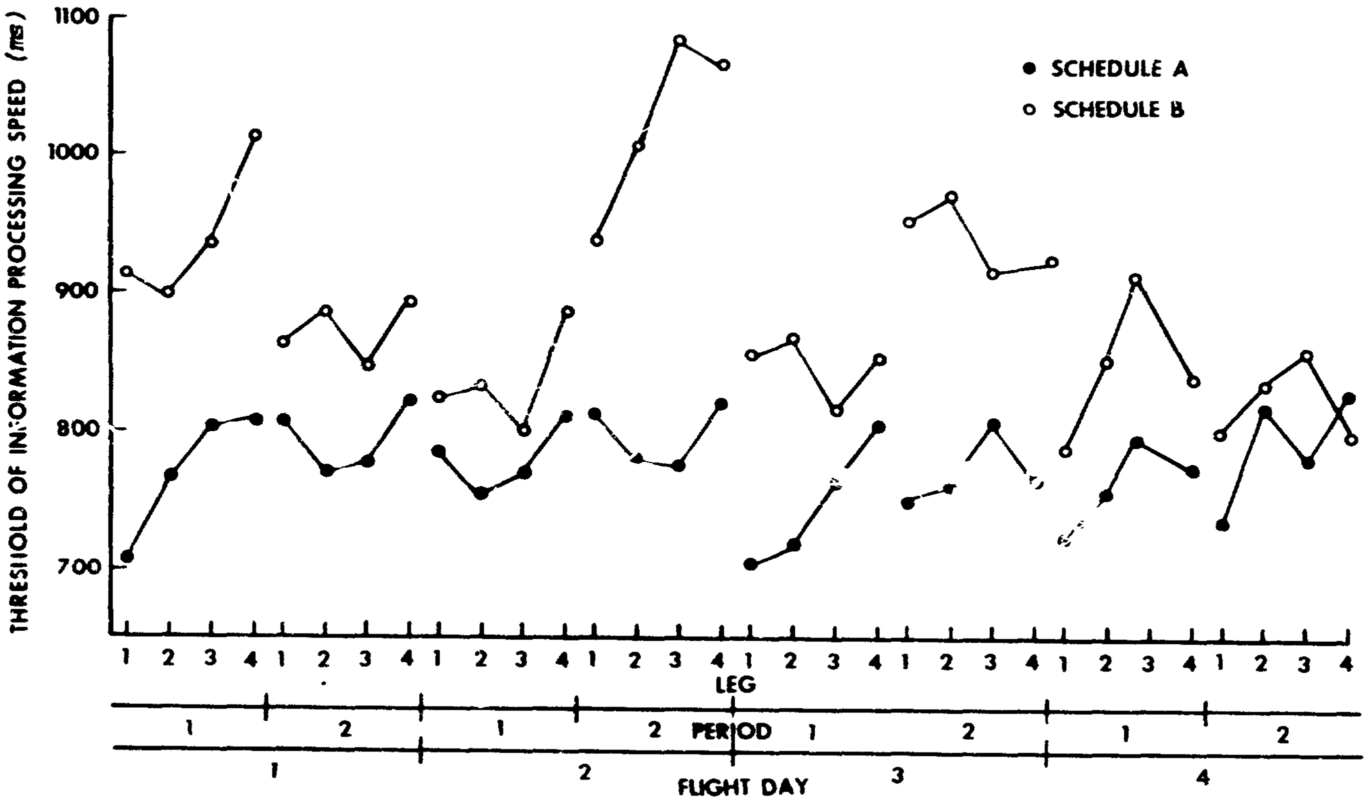


Figure 13. DIPT performance scores for each leg of flight days 1 to 4.

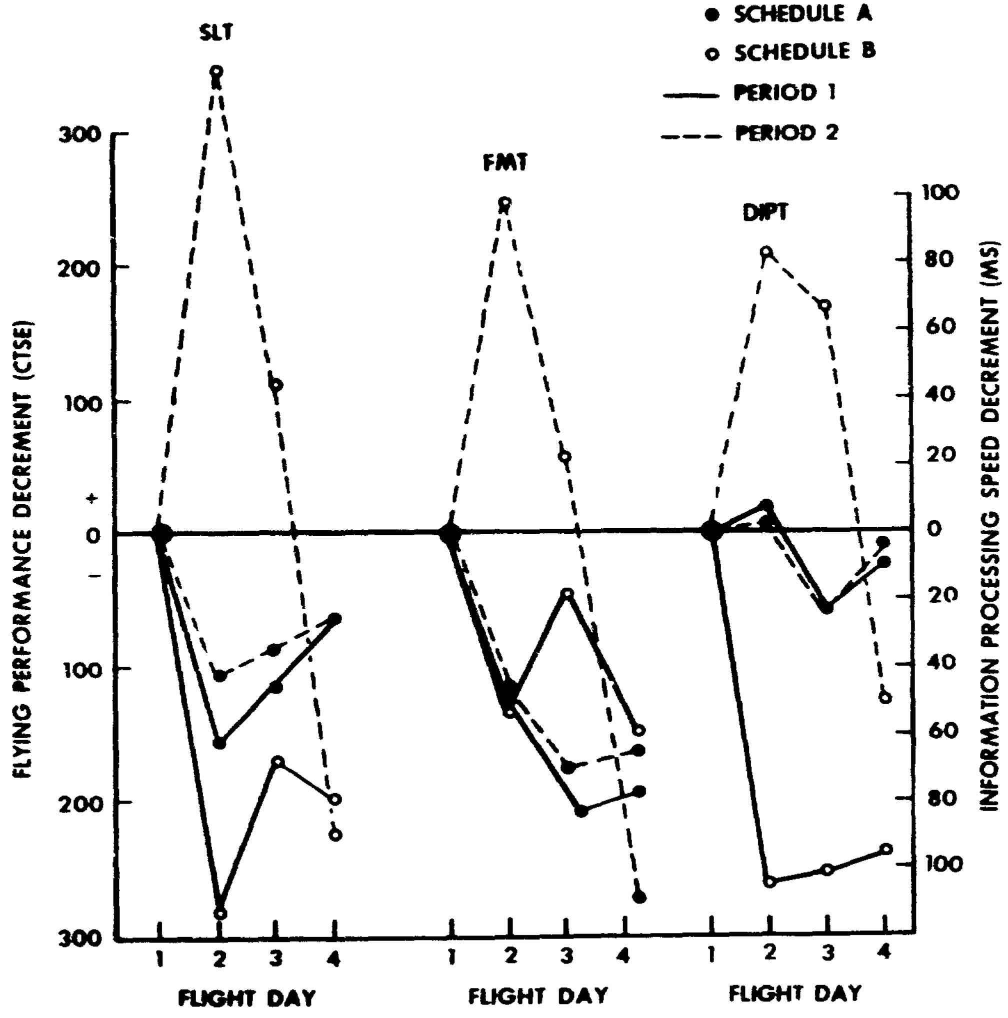


Figure 14. SLT, FMT, and DIPT decrement scores from flight day 1 to flight days 2, 3, and 4. (CTSE = combined total seconds of error)

The DIPT scores revealed no significant effects. From FD 2 on, Schedule B's Period 1 scores appear to have been stable (Fig. 12). If the subjects really had recovered from the effects of the fatiguing missions by FD 4, then this was probably their true baseline. The fact that their Period 2 scores on FD 4 were, if anything, lower, helped confirm the supposition that they had recovered. Hypotheses 1 and 2 are thus confirmed.

Learning Effects for Performance Measures

While performance was generally expected to decline with increasing fatigue, the possibility existed that overall performance would improve due to additional learning in the course of the four FD's. Thus, a three-way analysis of variance--for subjects, FD's, and Periods--was performed on Schedule A's SLT, FMT, and DIPT scores. Since Schedule A seemed to have no cumulative fatigue effects and, in the absence of any other fatigue stressors, the 9 hr of flight generated only mild fatigue, separate evaluation of their performance data was the most appropriate method for discovering any learning effects occurring across the 4 FD's. (The results of this analysis have been shown in Table 6.) A highly significant day effect exists for the SLT and FMT ($p<.01$ and $p<.001$, respectively), but not for the DIPT. As shown in Figure 10, the trend of the FMT FD means suggests learning, and the improvement is clear for the FMT. Since the FMT is a more complicated task than the SLT, greater continued improvement would be expected. Figure 12 shows the lack of improvement of Schedule A DIPT scores. Thus, unlike the SLT and FMT, the DIPT scores seemed to have plateaued prior to the start of the experimental session. Although the Period 2 scores were always in the predicted direction of performance decrement, these differences were not statistically significant.

Hypothesis 3: Fatigue Effects on Subjective Fatigue and Sleepiness

The analyses of variance for the Subjective Fatigue (SF) and the Stanford Sleepiness Scale (SSS) scores are presented in Tables 7 to 10. These data were not collected by legs, but at 5 specific times during the flight day: (a) just prior to takeoff of flight 1; (b) midflight; (c) just prior to takeoff of flight 2; (d) midflight; and (e) just after landing of flight 2.

As shown in Table 7, no statistically significant group or block differences occurred on FD 1; but both SF and SSS showed highly significant time-of-day effects ($p<.001$). Fatigue increased--from a preflight score of about 14 for both groups, to approximately 10 by midafternoon--and remained at this level until bedtime (Fig. 15). SSS scores ranged from approximately 2.25 at preflight to approximately 3.6, beginning at 1545 hr (Fig. 16).

The decrement in SF and SSS from FD 1 to FD 2 showed strong group effects (Table 8: $p<.001$). The SSS alone still showed a time effect ($p<.01$); but both SF and SSS showed a significant ($p<.05$) Group x Time interaction. Both patterns were similar and indicated that Schedule B reported significantly greater fatigue and sleepiness than on FD 1, depending on the point in time of the flight (Figs. 15 and 16).

TABLE 6. ANALYSES OF VARIANCE FOR SLT, FMT, AND DIPT SCORES FOR SCHEDULE A FOR FLIGHT DAYS 1 - 4

Source	df	Schedule A					
		SLT		FMT		DIPT	
		MS	F	MS	F	MS	F
Subjects (Ss)	9	2900324.00	88.520***	15506140.00	593.948***	265996.00	17.520***
Day (D)	3	249123.20	4.616**	690712.00	16.494***	12712.08	1.207
Ss x D	27	53972.40	—	41876.80	—	10530.12	—
Period (P)	1	139439.60	2.949	73283.60	3.073	72601.20	3.433
Ss x P	9	47285.00	—	23847.12	—	21147.08	—
D x P	3	26224.88	0.800	20205.48	0.774	6612.08	0.436
Ss x D x P	27	32764.60	—	26106.92	—	15182.00	—

**p<.01
***p<.001

TABLE 7. ANALYSES OF VARIANCE FOR SUBJECTIVE FATIGUE AND STANFORD SLEEPINESS SCALE SCORES FOR FLIGHT DAY 1

Source		Fatigue		SSS	
	df	MS	F	MS	F
Between Subjects (Ss)					
Group (G)	1	10.208	0.830	1.408	0.819
Block (B)	5	11.108	0.903	0.988	0.575
Ss within G,B	17	12.297	—	1.720	—
Within Ss					
Time (T)	4	62.929	21.916***	7.300	17.964***
G x T	4	2.188	0.762	0.492	1.210
B x T	20	2.454	0.855	0.580	1.427
Ss x T within G;B	68	2.871	—	0.406	—

***$p<.001$

TABLE 8. ANALYSES OF VARIANCE FOR SUBJECTIVE FATIGUE AND STANFORD SLEEPINESS SCALE DECREMENT SCORES FROM FLIGHT DAY (FD) 1 to FD 2

Source		Fatigue		SSS	
	df	MS	F	MS	F
Between Subjects (Ss)					
Group (G)	1	291.408	36.630***	38.533	43.191***
Block (B)	5	4.415	0.555	0.220	0.247
Ss within G,B	17	7.955	—	0.892	—
Within Ss					
Time (T)	4	3.675	1.264	1.742	3.701**
G x T	4	7.367	2.535*	1.575	3.347*
B x T	20	5.290	1.820*	1.037	2.203**
Ss x T within G,B	68	2.906	—	0.471	—

---

*p<.05
**p<.01
***p<.001

TABLE 9. ANALYSES OF VARIANCE FOR SUBJECTIVE FATIGUE AND STANFORD SLEEPINESS SCALE DECREMENT SCORES FROM FLIGHT DAY (FD) 1 to FD 3

Source		Fatigue		SSS	
	df	MS	F	MS	F
Between Subjects (Ss)					
Group (G)	1	175.208	14.303**	31.008	17.431***
Block (B)	5	12.108	0.988	0.668	0.376
Ss within G,B	17	12.250	—	1.779	—
Within Ss					
Time (T)	4	9.513	2.025	0.133	0.193
G x T	4	4.979	1.060	0.383	0.555
B x T	20	6.913	1.472	0.768	1.112
Ss x T within G,B	68	4.697	—	0.691	—

**p<.01
***p<.001

TABLE 10. ANALYSES OF VARIANCE FOR SUBJECTIVE FATIGUE AND STANFORD SLEEPINESS SCALE DECREMENT SCORES FROM FLIGHT DAY (FD) 1 to FD 4

Source		Fatigue		SSS	
	df	MS	F	MS	F
Between Subjects (Ss)					
Group (G)	1	1.900	0.095	0.000	0.000
Block (B)	5	2.080	0.105	0.860	0.434
Ss within G,B	17	3.979	—	0.396	—
Within Ss					
Time (T)	4	18.163	4.170**	1.170	2.584*
G x T	4	13.170	3.024*	2.456	5.424**
B x T	20	3.434	0.788	0.608	1.342
Ss x T within G,B	68	4.356	—	0.453	—

---

*p<.05
**p<.01

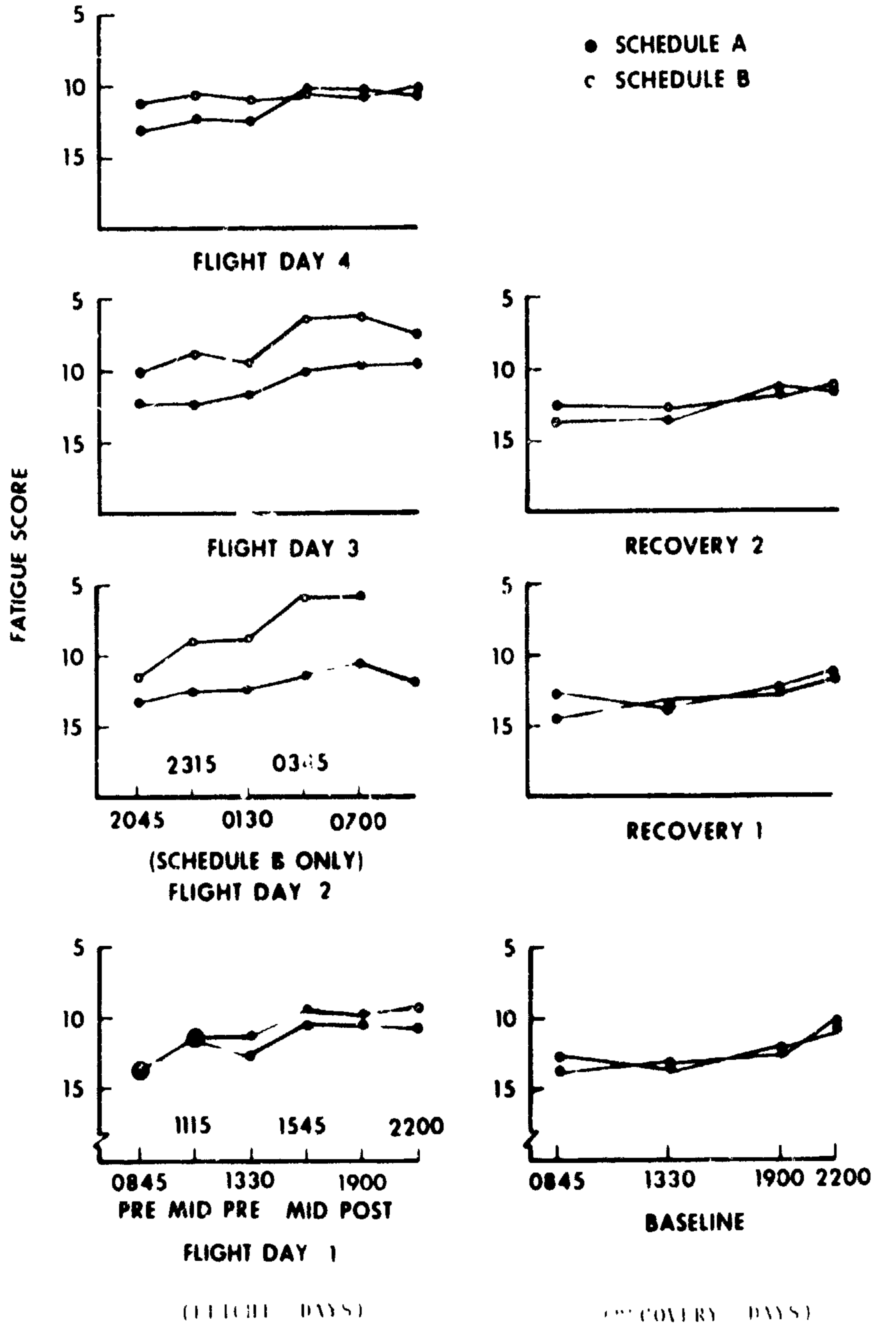


Figure 15. Subjective fatigue scores for baseline: flight days 1 to 4, and recovery days 1 and 2. (Times shown on flight day 2 apply to Schedule B only.)

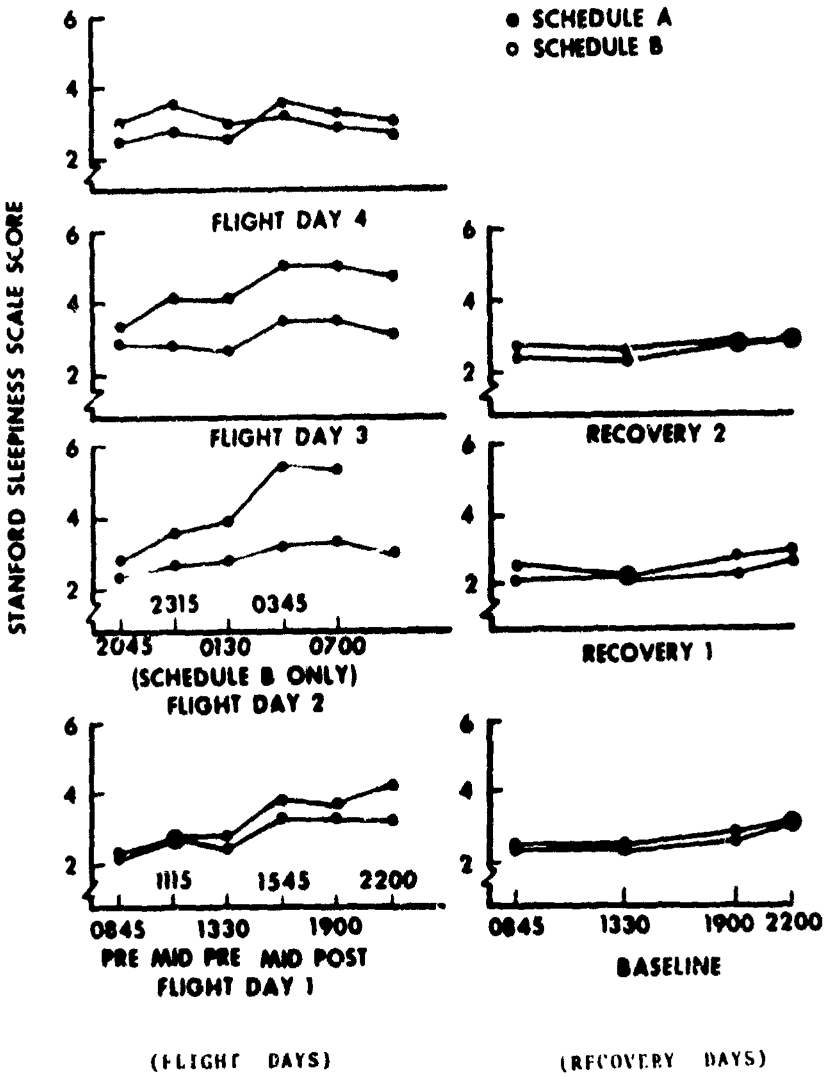


Figure 16. Stanford Sleepiness Scale scores for baseline: flight days 1 to 4, and recovery days 1 and 2. (Times shown on flight day 2 apply to Schedule B only.)

Scores ranged as follows: (a) SF preflight: Schedule A = 13.7, and Schedule B = 13.8; (b) SF postflight: Schedule A = 10.3, and Schedule B = 9.6; (c) SSS preflight: Schedule A = 2.3, and Schedule B = 2.2; and (d) SSS postflight: Schedule A = 3.3, and Schedule B = 3.7. Schedule B's greater preflight fatigue and sleepiness would be expected since they were reported at 2045 hr, after a full day of light nonflying duty. Both SF and SSS scores indicated a significant Block x Time interaction ($p<.05$ and .01, respectively). This finding might be interpreted as indicating that subjects fatigued at different rates, depending on their experimental schedule. Because of intersubject communication within a session, some correlation may have occurred between subjects for subjective reports during periods of intense fatigue. Because of computer resource limitations, the author predetermined that the Group x Block x Time interaction would not be analyzed because it was of secondary interest. However, since one subject's feelings might have influenced another's during high levels of fatigue, a reasonable assumption was that group differences might have existed; for subjects experiencing the same objective levels of fatigue roomed together and were in contact with each other more often.

On FD 3, SF and SSS reports showed only group differences (Table 9: $p<.01$ and .001, respectively). The time differences on FD 1 were interpreted as meaning that the fatigue and sleepiness levels differed between groups, but the patterns were not significantly different from the time effect reported for FD 1. As shown in Figures 15 and 16, patterns were similar to those for FD 2, but the ranges of scores reported were reduced in magnitude. Schedule B scores seem to indicate a higher Period 1 preflight fatigue and sleepiness than reported on FD 2--effects to be expected, since the subjects had been awake the prior night.

On FD 4 (Table 10), the SF and SSS scores indicated significant time effects ($p<.01$ and .05, respectively) and Group x Time interactions ($p<.05$ and .01, respectively). Comparison of FD 1 and FD 4 curves on Figure 15 indicated that the Schedule A shape and levels were approximately equivalent, but that Schedule B's FD 4 scores showed a flattening compared with those of FD 1, with a mean fatigue score of approximately 11. On FD 4, Schedule B started with higher fatigue than A, but A ended with about the same overall level as B. Both groups' SSS scores on FD 4 exhibited the same overall shape; but, in these scores as compared with those of A, Schedule B apparently started out feeling sleepier but ended feeling less sleepy (Fig. 16). Compared to FD 1, Schedule B appeared to have a lower overall FD 4 mean (3.06 vs. 3.30).

Additional analyses were performed to validate the conclusions drawn about recovery (Tables 11 and 12). No differences were found between groups during baseline, but strong time effects were still present for SF ($p<.01$) and SSS ($p<.001$). When baseline scores are compared to those of FD 1, the SF and SSS scores still show a strong time effect ($p<.001$). When subjects were not engaged in any demanding tasks, fatigue and sleepiness increased less and in a somewhat more regular fashion (Figs. 15 and 16). SF at the end of the day appears to be about 1 point less (11 vs. 10) than at the beginning.

When Recovery Days 1 and 2 were compared to baseline, no significant differences were found (Table 11). These data indicate essentially complete recovery from any fatigue, by both groups.

From Table 12, when R 1 and R 2 were compared to FD 1, strong time effects were noted (as with the baseline comparison to FD 1). For R 1 only a mild group effect (p<.05) was noted for the SSS scores. (The group effect for SF scores also approached significance: p = .053.) How the groups differ (Figs. 15 and 16) was not clear; but Schedule B's recovery scores may be slightly lower than their FD 1 scores, and Schedule A's scores slightly higher. Since on FD 1 the groups were not significantly different, apparently Schedule B subjects had recovered from the effects of their fatiguing session. Thus, Hypothesis 3 is confirmed. The conclusion is that the Schedule A group showed no significant increases in fatigue, except for a daily increase, over the 4-day mission. The Schedule B group experienced severe fatigue increases on FD's 2 and 3. On FD 4, while their mean overall fatigue level indicated recovery to FD 1 levels, the subjects seemed to have experienced moderate fatigue all day long.

Hypothesis 4: Fatigue Effects on Hours Slept

An analysis of variance was performed on the hours of sleep reported by each subject immediately after waking to determine if either group required more sleep or obtained more sleep than the other. Analyses were performed on sleep reported in the 24 hr prior to baseline, R 1 and R 2. As shown in Table 13, no significant group, block, or day effects were found. Thus, prior to the start of the experiment, all subjects apparently required and received equal amounts of sleep; and, by the conclusion of FD 4, all subjects apparently required and received approximately the same amount of sleep. For the sleep requirements of the fatigue generated by Schedule B, no carryover effects were evident. Sleep reported by both groups prior to each block day is presented in Figure 17. A Block x Day effect, significant at the 0.05 level, indicated that subjects slept different amounts depending on their respective block. This is an understandable result, because the four subjects in a block spent much of their nonsleeping crew rest periods together; and the unit as a whole exhibited some influence on the exact time the individuals would go to sleep.

Also in Table 13 are the results of the analysis of variance for sleep reported prior to BD's 2, 3, and 6. This analysis was performed to determine if sleep requirements before the flying missions increased during the height of the fatiguing period. This period was considered to be the conclusion of FD 3 (BD 6) for Schedule B. BD's 2 and 3 were used as chronological reference points. The activity prior to the sleep reported was the same on BD 2 for Schedule B and on BD 3 for Schedule A (nonflying duty). The activity prior to reporting for Schedule B on BD 3 (FD 1) might, however, have been more tiring than on BD 2 for Schedule A (free time during baseline). Thus, the comparison was conservative.

The block effect noted earlier was even more significant, possibly due to an increased cohesiveness among subjects during the actual experimental session. Strong differences between groups were also apparent, depending on which day they were reporting sleep (GxD, p<.001). Until the night mission was flown, the Schedule B subjects required slightly less sleep than the Schedule A (Fig. 17). After that, the Schedule B subjects appeared to have slept significantly longer than the Schedule A, until the first night after FD 4, at which time (from the previous analysis) they apparently did not sleep significantly more than Schedule A subjects. Thus, Hypothesis 4 is confirmed: Even though the amount of time available for sleep varied within narrow limits, subjects in the fatiguing schedule obtained more sleep than usual immediately after flights where they were awake 12 hr before flying.

TABLE 11. ANALYSES OF VARIANCE OF SUBJECTIVE FATIGUE AND STANFORD SLEEPINESS SCALE SCORES FOR BASELINE, BASELINE VS. FLIGHT DAY 1, AND BASELINE VS. RECOVERY DAYS 1 AND 2

Source		Baseline		FD 1 - Baseline		R 1 - Baseline		R 2 - Baseline	
	df	MS	F	MS	F	MS	F	MS	F
				Fatigue					
Between Subjects (Ss)									
Group (G)	1	5.510	0.520	4.167	0.64	26.042	3.010	1.760	0.25
Ss within G	22	10.579	—	6.553	—	8.650	—	6.973	—
Within Ss									
Time (T)	3	29.955	6.020**	33.694	5.73**	1.028	0.160	4.788	0.97
G x T	3	1.233	0.250	8.528	1.45	7.569	1.150	0.899	0.18
Ss x T/G	66	4.973	—	5.884	—	6.602	—	4.920	—
				Stanford Sleepiness Scale					
Between Subjects (Ss)									
Group (G)	1	0.260	0.24	3.010	3.58	0.667	1.000	1.760	1.62
Ss within G	22	1.079	—	0.840	—	0.667	—	1.086	—
Within Ss									
Time (T)	3	2.400	6.73***	4.288	4.82**	0.111	0.210	0.594	0.87
G x T	3	0.010	0.03	0.538	0.60	0.111	0.210	0.205	0.30
Ss x T/G	66	0.356	—	0.890	—	0.535	—	0.680	—

**$p<.01$

***$p<.001$

FD = flight day

R = recovery day

TABLE 12. ANALYSES OF VARIANCE OF SUBJECTIVE FATIGUE AND STANFORD SLEEPINESS SCALE SCORES FOR RECOVERY DAYS 1 AND 2 VS. FLIGHT DAY 1

Source		R 1 - FD 1		R 2 - FD 1	
	df	MS	F	MS	F
			Fatigue		
Between Subjects (Ss)					
Group (G)	1	51.042	4.323	0.510	0.074
Block (B)	5	6.517	0.552	4.885	0.713
Within Ss					
Ss/G,B	17	11.806	—	6.856	—
Time (T)	3	29.806	5.371**	42.399	8.746***
G x T	3	2.069	0.373	10.483	2.162
B x T	15	4.722	0.851	4.258	0.878
Ss x T/G,B	51	5.550	—	4.849	—
			Stanford Sleepiness Scale		
Between Subjects (Ss)					
Group (G)	1	6.510	4.724*	0.167	0.239
Block (B)	5	0.810	0.588	0.267	0.383
Within Ss					
Ss/G,B	17	1.378	—	0.696	—
Time (T)	3	5.454	7.067***	7.722	7.574***
G x T	3	0.316	0.409	1.333	1.308
B x T	15	0.872	1.129	0.856	0.839
Ss x T/G,B	51	0.772	—	1.020	—

---

*$p<.05$
**$p<.01$
***$p<.001$

TABLE 13. ANALYSES OF VARIANCE FOR HOURS OF PRIOR SLEEP REPORTED ON BLOCK DAYS 2, 3, AND 6, AND FOR BASELINE, AND FOR RECOVERY DAYS R 1 AND R 2

Source		BD 2 vs. BD 3 vs. BD 6		Baseline vs. R 1 vs. R 2	
	df	MS	F	MS	F
Between Subjects (Ss)					
Group (G)	1	0.222	0.919	0.031	0.049
Block (B)	5	1.631	6.740**	1.728	2.694
Ss within G,B	17	0.242	—	0.642	—
Within Ss					
Day (D)	2	2.837	8.841***	0.170	0.407
G x D	2	3.378	10.529***	0.823	1.966
B x D	10	0.433	1.348	0.912	2.179*
Ss x D within G,B	34	0.321	—	0.419	—

---

*p<.05

**p<.01

***p<.001

(**Note**: BD 6 corresponds to FD 4 for both schedules; see Fig. 17)

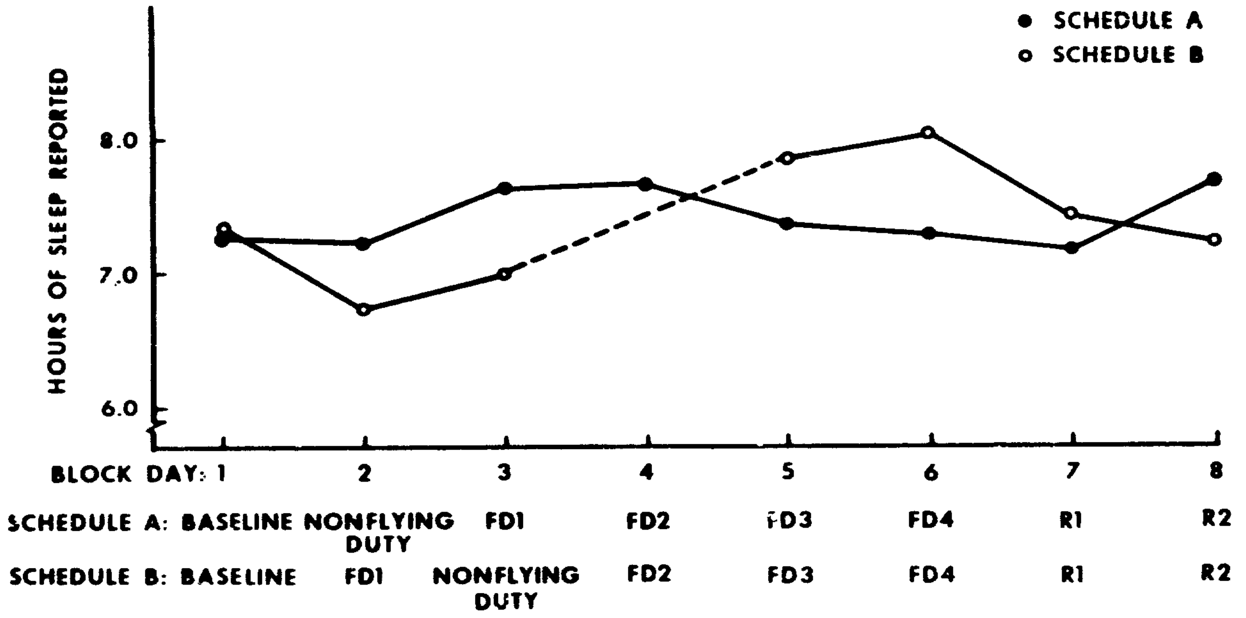


Figure 17. Hours of sleep in the 24 hr prior to each block day. Sleep reported by Schedule B subjects received prior to block day 5 occurred between 0800 and 1800 hr of block day 4. All other sleep reported occurred between 2000 and 0600 hr of the previous night. For Schedule B subjects, no sleep occurred between their nonflying duty day and flight day 2, so none was reported on block day 4.

One could assume that subjects in Schedule B would have slept even longer if time permitted, because they were always required to arise at a predetermined time. Given this specific work-rest cycle, however, their sleep requirements apparently returned to usual after one night of sleep in a normal duty cycle (i.e., 12 hr on, 12 hr off).

Hypothesis 5: Fatigue Effects on Heart Rate and Heart Rate Variability

Analyses for heart rate (HR), heart rate variability (HRV), and rectal temperature followed essentially the same model as performance analyses. Block was dropped as a factor, however; for minimal evidence of an effect was found in the previous analyses and, by pooling its df's in the appropriate error terms, a more sensitive test was obtained. Separate analyses were conducted for the HR, HRV, and rectal temperature data collected during performance of the SLT, FMT, and DIPT.

HR and HRV analyses of variance are presented in Tables 14 to 17. As shown in Table 14, significant period effects were found on FD 1 for HR: SLT $p<.001$), FMT ($p<.05$), and DIPT ($p<.05$). Although group differences appeared large (Fig. 18), they were not significant. Because subjects in Schedule B flew first, they may have been more anxious than the other subjects about what would happen to them and how long they would fly the first day. Schedule A subjects may have assumed that their first FD would be equal to that of Schedule B (although they were not so advised), and were thus less concerned.

Highly significant ($p<.001$) "leg within period" HR effects were found (Fig. 19). For both groups, a general decline in HR occurred throughout the first period, followed by recovery and decline in Period 2. (Note that Fig. 19 presents the average HR for all three tasks combined.) On FD 1, HR was always greater in Period 1 than in Period 2. Since HR actually fell within the flight, the period increase was probably of a circadian nature and not due to fatigue, even though fatigue increased as shown by the SF and SSS time effect. The decrease was probably also a function of loss of arousal as the subjects became bored with the flight duty, regardless of the fatigue.

The HRV on FD 1 showed only a "leg within period" effect (Table 14), for SLT ($p<.001$) and FMT ($p<.01$). No clear pattern was evident from Figures 20 or 21. (Note that the average HRV for all three tasks combined is plotted in Fig. 21.)

Significant for groups for all three tasks were the decrement in HR from FD 1 to FD 2 and a period effect for SLT (Table 15). In addition, the Group x Period interaction was significant for all three tasks (SLT, $p<.001$; FMT, $p<.001$; and DIPT, $p<.05$). Schedule A showed only a slight decrease from FD 1 to FD 2; but, for Schedule B: Period 2, FD 2 was much lower than Period 1 on both FD's 1 and 2 (Fig. 18). The large drop in HR was probably a combination of circadian effect and fatigue.

The SLT HR also showed a "leg within period" effect ($p<.01$) and a Group x "leg within period" effect ($p<.05$) for FD 2. The DIPT HR showed a "leg within period" effect ($p<.05$). In FD 2, Period 2, HR exhibited a much narrower range and seemed to "bottom out" (Fig. 19). Thus, with extreme fatigue and high task demand, as generated by flying, HR may be expected to begin increasing (as with physical exercise).

The only HRV FD 2 effect was noted in the SLT period difference (Table 15: $p<.01$). For the SLT alone, HRV clearly increased with the extreme levels of fatigue experienced in FD 2, Period 2 (Fig. 20).

The decrement scores for HR showed no significant effects on FD 3 (Table 16). Effects observed on FD 1 are considered to be present in FD 3.

For HRV, the FMT scores showed the only differences on FD 3 (Table 16). Group differences ($p<.05$) can be clearly seen in Figure 20. For the complex task, the FMT, both periods showed an increased HRV. In trying to maintain performance, Schedule B subjects may have had to increase their concentration for the SLT and thus lowered their HRV somewhat; but they were now unable to concentrate as fully on the FMT in either period as they had on FD 1.

The extremely high FD 3, Period 1, DIPT HRV mean for Schedule A cannot be explained, but is nonsignificant (Fig. 20). None of the other scores indicate what might have happened at this time. High HRV is taken to be an index of lack of concentration or boredom, yet the DIPT scores for that period were the lowest. No evidence of equipment malfunction could be uncovered. The subjects' HR scores, from which this measure is derived, seemed reasonable.

For HR on FD 4, only the "leg within period" DIPT scores were significantly different from those of FD 1 (Table 17: $p<.05$). For HRV on FD 4, only the "leg within period" effect for SLT scores was significantly different from that of FD 1 (Table 17: $p<.01$). As indicated earlier, "leg within period" effects indicate differences among legs within Period 1 or Period 2, or a combination of significant leg and/or leg x period effects. For HRV, the difference may have become evident because the patterns for Schedules A and B became markedly similar on FD 4 (Fig. 21). This finding was taken as an indication of the recovery of Schedule B, since Schedule A never demonstrated any HRV decrement. Schedule B's HR returned to a relatively high level on FD 4, but was still not significantly different from Schedule A's HR, thus indicating that the subjects (as a group) may simply have had a slightly higher metabolic rate than those in Schedule A (Fig. 19).

Also, the Schedule A HR indicated a tendency to fall across all 4 FD's (Fig. 18). This tendency was interpreted as a loss of arousal as the subjects became increasingly familiar with the task and, possibly, bored.

Hypothesis 5 is confirmed. HR and HRV are able to differentiate between levels of fatigue. With increasing fatigue, however, HR declined and then leveled off, but HRV increased.

Hypothesis 6: Fatigue Effects on Body Temperature

The results of the FD 1 rectal temperature analyses are presented in Table 18. Group effects were neither expected nor found, nor were period effects. In Figure 22, the expected circadian rise during Period 2 could be seen for Schedule A subjects; but the reverse was true for Schedule B subjects.

TABLE 14. ANALYSES OF VARIANCE FOR HEART RATE AND HEART RATE VARIABILITY FOR FLIGHT DAY 1

Source		SLT		FMT		DIPT	
	df	MS	F	MS	F	MS	F
				HR			
Between Subjects (Ss)							
Group (G)	1	5626.836	3.53	5686.540	2.800	8606.824	3.93
Ss within G	22	1594.308	—	2030.948	—	2189.460	—
Within Ss							
Period (P)	1	766.136	19.15***	320.076	7.030*	589.752	4.68*
G x P	1	2.884	0.07	0.576	0.010	56.660	0.45
Ss x P within G	22	40.000	—	45.500	—	125.984	—
Leg within P (L/P)	6	655.803	29.342***	895.716	34.711***	962.897	16.220***
G x L/P	6	6.301	0.282	10.513	0.407	78.714	1.326
Ss x L/P within G	130	22.351	—	25.805	—	59.364	—
				HRV			
Between Subjects (Ss)							
Group (G)	1	1.104	0.020	3459.504	0.010	15.912	0.430
Ss within G	22	53.312	—	66.632	—	37.404	—
Within Ss							
Period (P)	1	0.0016	0.000	13.700	3.450	0.100	0.050
G x P	1	0.852	0.240	0.884	0.220	8.652	0.470
Ss x P within G	22	3.560	—	3.972	—	18.544	—
Leg within P (L/P)	6	11.409	4.619***	10.660	3.800**	10.452	0.585
Group x L/P	6	0.734	0.297	1.240	0.442**	7.891	0.442
Ss x L/P within G	126	2.470	—	2.805	—	17.864	—

---

*p<.05
**p<.01
***p<.001

TABLE 15. ANALYSES OF VARIANCE FOR HEART RATE AND HEART RATE VARIABILITY DECREMENT SCORES FROM FLIGHT DAY (FD) 1 TO FD 2

Source	df	SLT MS	SLT F	FMT MS	FMT F	DIPT MS	DIPT F
				HR			
Between Subjects (Ss)							
Group (G)	1	2619.360	5.540*	3277.576	5.880*	6307.364	8.580**
Ss within G	19	473.056	—	557.816	—	735.324	—
Within Ss							
Period (P)	1	450.044	5.450*	336.128	3.670	291.888	1.640
G x P	1	1314.504	15.910***	1666.352	18.180***	1338.196	7.510*
Ss x P within G	19	82.644	—	91.648	—	178.120	—
Leg within P (L/P)	6	100.985	3.703**	78.819	1.860	295.517	2.839*
G x L/P	6	73.931	2.711*	65.584	1.548	180.564	1.735
Ss x L/P within G	121	27.272	—	42.380	—	104.091	—
				HRV			
Between Subjects (Ss)							
Group (G)	1	1.544	0.040	1.256	0.040	6.000	0.030
Ss within G	19	4.036	—	33.900	—	171.832	—
Within Ss							
Period (P)	1	88.668	8.170**	6.868	0.810	6.680	0.130
G x P	1	4.128	3.800	31.896	3.770	10.416	0.210
Ss x P within G	19	10.860	—	8.468	—	49.748	—
Leg within P (L/P)	6	6.337	1.674	10.300	2.072	7.077	1.325
G x L/P	6	7.183	1.898	6.631	1.334	114.201	1.440
Ss x L/P within G	118	3.784	—	4.971	—	79.302	—

---

*p<.05
**p<.01
***p<.001

TABLE 16. ANALYSES OF VARIANCE FOR HEART RATE AND HEART RATE VARIABILITY DECREMENT SCORES FROM FLIGHT DAY (FD) 1 TO FD 3

Source		SLT		FMT		DIPT	
	df	MS	F	MS	F	MS	F
				HR			
Between Subjects (Ss)							
Group (G)	1	304.684	0.890	484.192	1.060	1250.844	2.050
Ss within G	20	343.346	—	453.232	—	608.736	—
Within Ss							
Period (P)	1	195.888	3.320	15.848	0.280	27.840	0.330
G x P	1	75.084	1.270	56.824	1.010	191.112	2.260
Ss x P within G	20	58.932	—	56.128	—	84.484	—
Leg within P (L/P)	6	17.068	1.207	19.205	0.837	130.518	1.273
G x L/P	6	21.222	1.500	18.624	0.564	99.801	0.973
Ss x L/P within G	127	14.145	—	22.943	—	102.536	—
				HRV			
Between Subjects (Ss)							
Group (G)	1	21.836	0.840	96.176	4.490*	13.552	0.110
Ss within G	21	29.424	—	21.876	—	122.356	—
Within Ss							
Period (P)	1	0.144	0.010	20.672	1.990	59.180	0.900
G x P	1	6.372	0.590	5.024	0.480	12.520	1.900
Ss x P within G	21	10.796	—	10.364	—	65.928	—
Leg within P (L/P)	6	6.339	1.538	5.590	1.183	15.487	0.522
G x L/P	6	2.987	0.725	5.531	1.171	39.258	1.322
Ss x L/P within G	123	4.121	—	4.724	—	29.689	—

---

*$p < .05$

TABLE 17. ANALYSES OF VARIANCE FOR HEART RATE AND HEART RATE VARIABILITY DECREMENT SCORES FROM FLIGHT DAY (FD) 1 TO FD 4

Source	df	SLT MS	SLT F	FMT MS	FMT F	DIPT MS	DIPT F
				HR			
Between Subjects (Ss)							
Group (G)	1	615.320	0.690	770.052	0.680	732.308	0.770
Ss within G	20	896.928	—	1138.612	—	945.964	—
Within Ss							
Period (P)	1	17.188	0.120	76.456	0.410	22.356	0.090
G x P	1	68.988	0.480	143.484	0.770	8.976	0.030
Ss x P within G	20	142.720	—	186.400	—	256.564	—
Leg within P (L/P)	6	19.272	0.596	21.816	0.571	262.764	2.770*
G x L/P	6	9.948	0.308	17.313	0.453	107.835	1.137
Ss x L/P within G	113	32.346	—	38.182	—	94.875	—
				HRV			
Between Subjects (Ss)							
Group (G)	1	6.192	0.220	1.316	0.030	6.772	0.060
Ss within G	20	27.644	—	44.316	—	105.696	—
Within Ss							
Period (P)	1	0.084	0.000	23.376	0.690	4.164	0.140
G x P	1	0.104	0.010	14.536	0.430	20.400	0.670
Ss x P within G	20	19.088	—	33.760	—	30.424	—
Leg within P (L/P)	6	20.810	3.259**	12.072	1.227	38.819	0.635
G x L/P	6	3.650	0.572	7.133	0.725	8.049	0.132
Ss x L/P within G	111	6.385	—	9.838	—	61.163	—

---

*p<.05
**p<.01

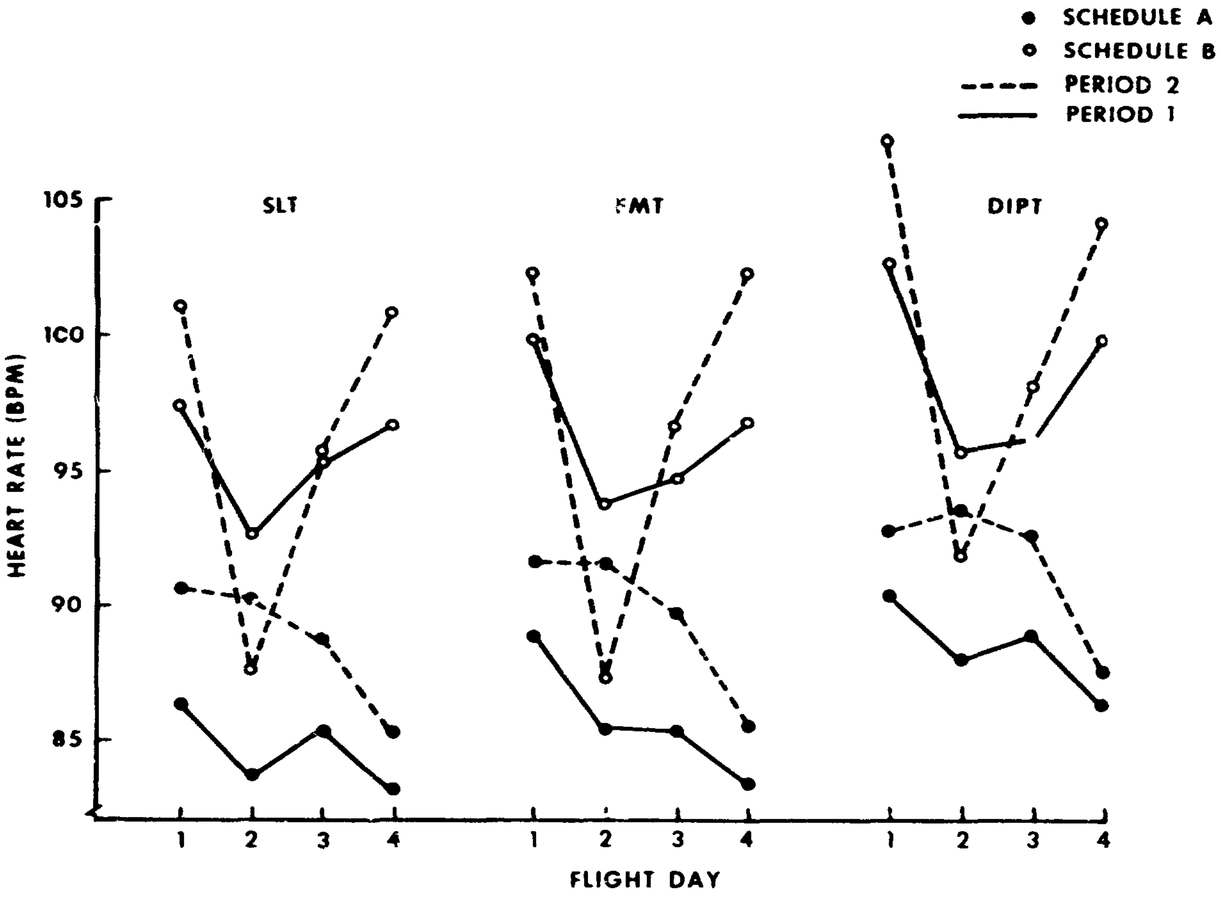


Figure 18. Heart rate in beats per minute averaged over SLT, FMT, and DIPT epochs for flight days 1 to 4.

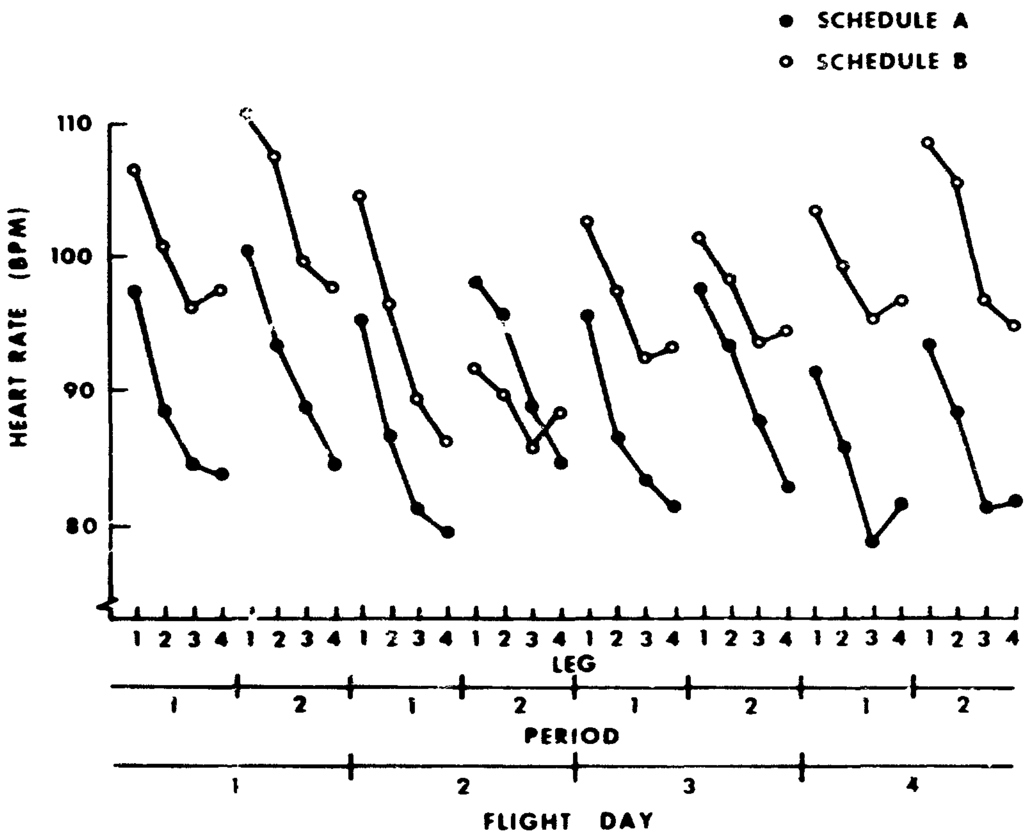


Figure 19. Heart rate for each leg of flight days 1 to 4. (Each data point is the average over corresponding SLT, FMT, and DIPT epochs.)

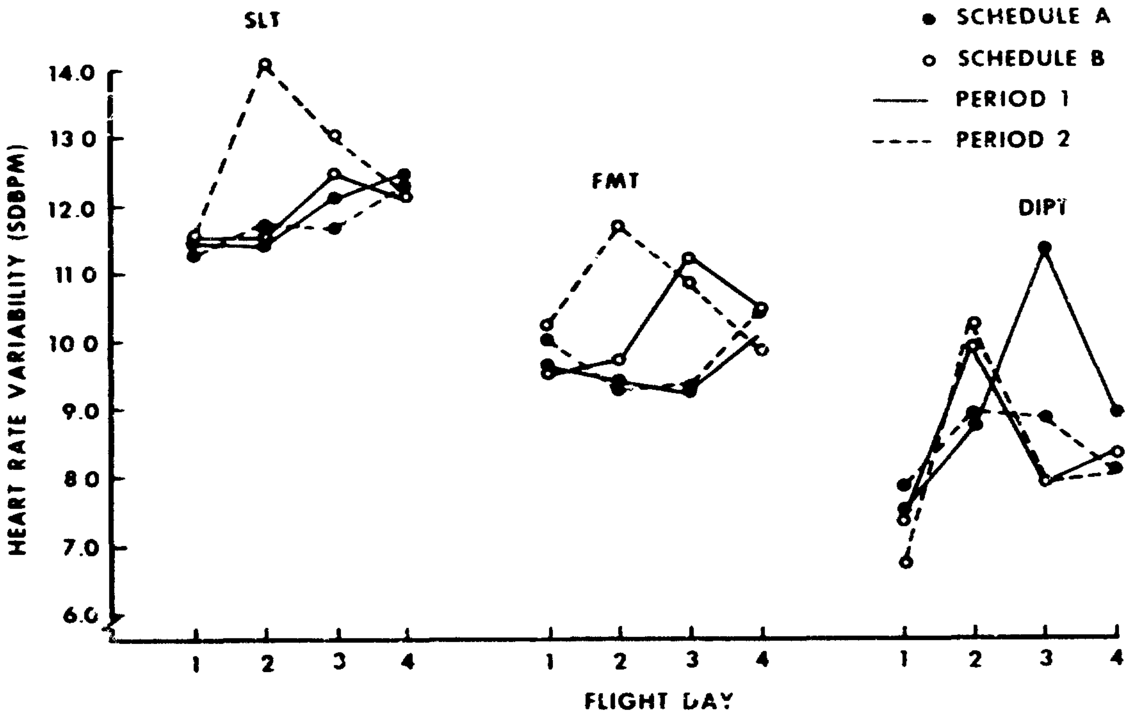


Figure 20. Heart rate variability in standard deviation units of beats per minute (SDBPM) averaged over SLT, FMT, and DIPT epochs for flight days 1 to 4.

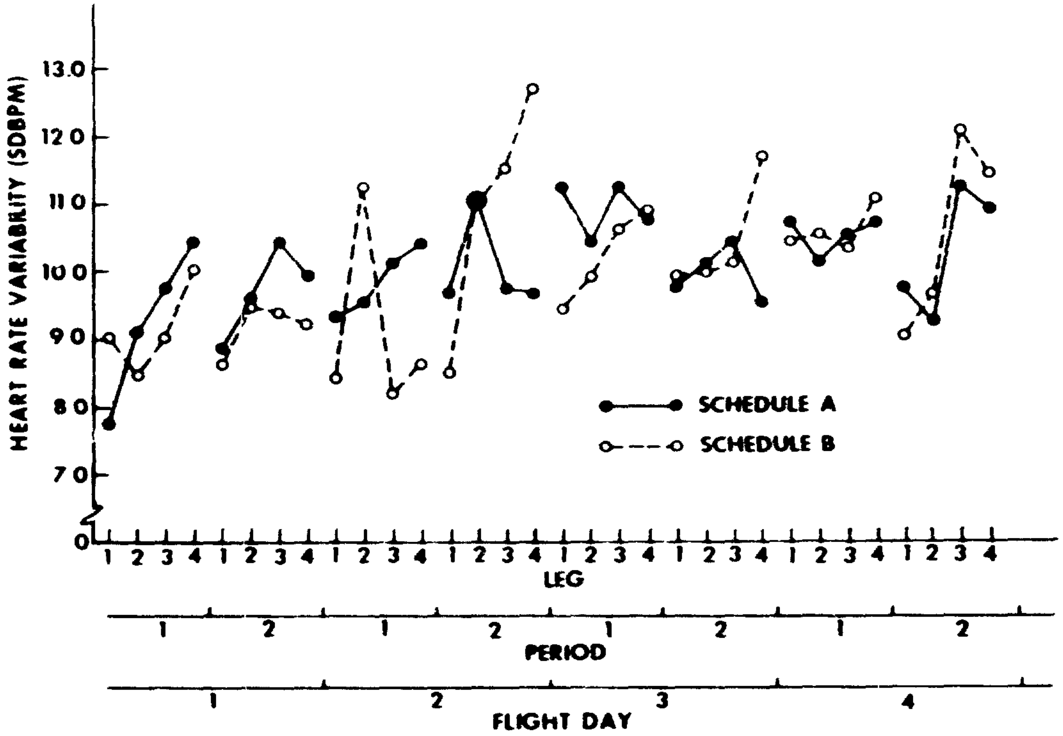


Figure 21. Heart rate variability for each leg of flight days 1 to 4. (Each data point is the average over corresponding SLT, FMT, and DIPT epochs. [SDBPM = standard deviation units of beats per minute]

TABLE 18. ANALYSES OF VARIANCE FOR RECTAL TEMPERATURE FOR FLIGHT DAY 1 AND DECREMENT SCORES FROM FLIGHT DAY (FD) 1 TO FD 2

Source		SLT		FMT		DIPT	
	df	MS	F	MS	F	MS	F
				FD 1			
Between Subjects (Ss)							
Group (G)	1	0.504	1.310	0.572	1.420	0.552	1.350
Ss within G	21	0.384	—	0.404	—	0.412	—
Within Ss							
Period (P)	1	0.004	0.020	0.001	0.000	0.002	0.010
G x P	1	0.332	1.110	0.288	0.980	0.238	0.950
Ss x P within G	21	0.300	—	0.296	—	0.304	—
Leg within P (L/P)	6	0.049	9.352***	0.073	8.295***	0.019	2.014
G x L/P	6	0.002	0.287	0.003	0.348	0.011	1.202
Ss x L/P within G	126	0.005	—	0.009	—	0.009	—
				FD 1 - FD 2			
Between Subjects (Ss)							
Group (G)	1	1.068	1.870	0.772	1.290	1.252	2.030
Ss within G	18	0.572	—	0.600	—	0.616	—
Within Ss							
Period (P)	1	1.736	3.990	1.908	4.290	1.440	3.040
G x P	1	2.260	5.190*	2.616	5.880*	2.144	4.520**
Ss x P within G	18	0.436	—	0.444	—	0.476	—
Leg within P (L/P)	6	0.058	9.008***	0.053	6.007***	0.073	6.103***
G x L/P	6	0.038	5.931***	0.036	4.071**	0.053	4.450***
Ss x L/P within G	116	0.006	—	0.009	—	0.012	—

---

*p<.05
**p<.01
***p<.001

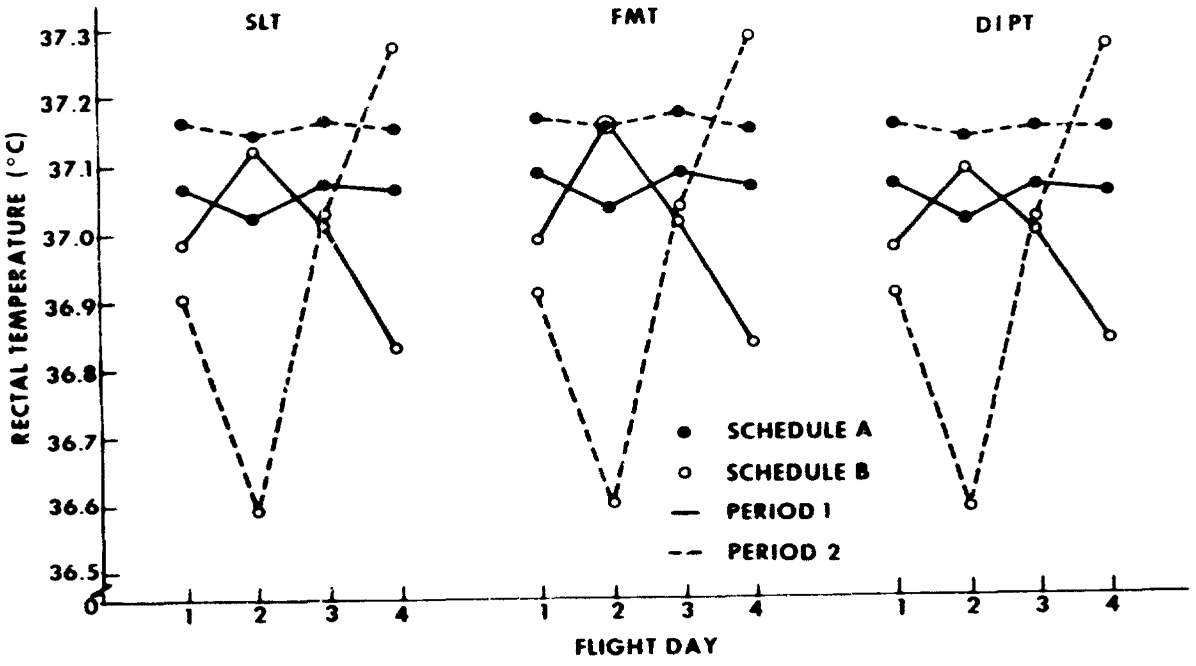


Figure 22. Rectal temperature averaged over SLT, FMT, and DIPT epochs for flight days 1 to 4.

As shown in Figure 23, both groups were similar during Period 1 but appeared to have level differences in Period 2. Moreover, no Group x Period interaction was found.

This information is further indication that Schedule B subjects may have differed somewhat from Schedule A subjects during FD 1. If, as performance scores tend to indicate (although not significantly), Schedule B subjects had a slightly lower skill level and were still learning the DIPT and flying tasks, then they might have been working harder on FD 1. In that case, rectal temperature might have been expected to be higher, not lower, than that of Schedule A subjects--especially if their metabolic rate was higher than Schedule A, as indicated by their HR.

SLT and FMT data both show significant "leg within period" effects ($p<.001$). Why the DIPT data did not show this effect is not clear.

The analyses of the decrement of rectal temperature from FD 1 to FD 2 now indicate a Group x Period interaction (Table 18: SLT, $p<.05$; FMT, $p<.05$; DIPT, $p<.01$). Schedule B's period values were higher than Schedule A's, probably due to the circadian effect of when they started. However, this effect then fell steadily, until leveling off during Period 2. Schedule A's temperature exhibited the same circadian rise as on FD 1.

Separate t-tests compared Period 1 (P1) to Period 2 (P2), for FD 2, for both groups. Schedule A showed a significant temperature increase (P1 = 37.01°C; P2 = 37.14°C; $\underline{t}$ (11) = 3.338; p = .007, two-tailed). Schedule B showed a significant decrease (P1 = 37.11°C; P2 = 36.55°C; $\underline{t}$ (10) = 4.042; p = .005). Schedule A data across all 4 FD's showed a consistent pattern of Period 2 circadian increase (Fig. 22). Thus, the significant FD 2 increase from Periods 1 to 2 was assumed indicative of a significant circadian rise across all FD's for Schedule A.

SLT, FMT, and DIPT temperature data also showed both a very strong "Leg Within Period" effect and Group (G) x "Leg within Period" (L/P) interaction (all $p<.001$--except FMT G x L/P, which was $p<.01$).

On FD 3 the only indication of a difference of rectal temperature from FD 1 was a "Leg Within Period" effect (Table 19: $p<.01$). Otherwise, the effects of the fatigue seem to have minimal effect on their FD 3 temperature (Figs. 22 and 23).

In examination of the FD 4 data in Figures 22 and 23, a striking shift was noted. Schedule B now exhibited the expected circadian rise in Period 2, but with an exaggerated range (P1 = 36.83°C; P2 = 37.27°C). As shown in Table 19, the SLT and DIPT temperature period effect was significant ($p<.05$), and the FMT temperature was significant for Period and Group x Period (both $p<.05$). For the SLT Group x Period interaction, p = .051. For the "Leg Within Period" effect, SLT and FMT temperature was significant at $p<.001$, and DIPT temperature was significant at $p<.01$. These data are the only strong evidence that the Schedule B subjects might still have been experiencing physiological effects from the two previous fatiguing duty periods.

Hypothesis 6 predicted that body core temperature would decline from stress effects when circadian effects were controlled. Body temperature was demonstrated to rise and fall in a predicted circadian fashion, being observed at maximum during Period 1, Leg 1, FD 2 for Schedule B subjects (Fig. 23). This leg occurred from 2100 to 2200 hr. Thereafter, given the workload of GAT-1 flying and the circadian effect, body temperature fell steadily during Period 2. The minimum temperature occurred during Period 2, at 0400 hr. The temperature then rose slightly until 0700 hr, the end of that flight (Fig. 23).

On FD 4, for Period 1, Schedule B, the fatigue effects did cause body core temperature depression below that caused by the circadian effects normally appearing in Schedule A's data, thus confirming Hypothesis 6. However, in Period 2, the temperature of Schedule B subjects increased above that which would be expected, due to the unforeseen rise in body temperature in the afternoon. Apparently the fatigue effects, rather than simply lowering overall body temperature, reduced the body's ability to control temperature, thus causing both overshoot and undershoot of the normal temperature range.

Hypothesis 7: Fatigue Effects on Urinary Constituents

As noted earlier, this hypothesis could not be tested, because freezing the urine samples altered their values in an unforeseen fashion.

Hypothesis 8: Arousal Values of Each Performance Measure

Hypothesis 8 predicted that HR and HRV would reflect the arousal generated by each performance task. The arousal was predicted to be a function of task demand--with the SLT creating the lowest demand; FMT, a higher demand; and DIPT, the highest. Some evidence for this order has been noted for HRV (Fig. 20). For HR, the effect cannot be clearly seen from Figure 18. A four-way analysis of variance was performed for subjects, FD's 1 to 4, periods, and tasks (SLT, FMT, and DIPT). The decision was made to eliminate groups as a factor and perform separate analyses for each; for direct comparison between groups was not of as much interest as task comparisons, and high variability in one or another of the groups' data might have obscured the task effect.

Rather than combining all three SLT's per leg, SLT b was chosen for analysis. The two reasons were: first, the data collection period for a single SLT (10 min) was more comparable to that for FMT (16 min); and, second, the middle SLT in each leg was not as likely to be affected by extraneous arousal factors (such as takeoff, landing, and subjective fatigue), and hence was a more pure sample of the arousal value of the task.

The results are presented in Table 20. For both groups, for HR and HRV, strong individual differences were found. For Schedule A, no other HR effects were found. For Schedule B, Day ($p<.01$) and Day x Period ($p<.05$) HR effects were found, as noted in earlier analyses. For HRV, Schedule B showed a significant period effect ($p<.05$). Schedule A showed a significant HRV effect for tasks ($p<.01$). The means can be ranked in the predicted order: SLT = 12.50, FMT = 9.81, and DIPT = 8.93 SDBPM.

Because the DIPT usually lasted no longer than 2 min, the sample time of DIPT HR was much shorter than that of the SLT or FMT, so further analysis with these data was eliminated. The results are presented in Table 21.

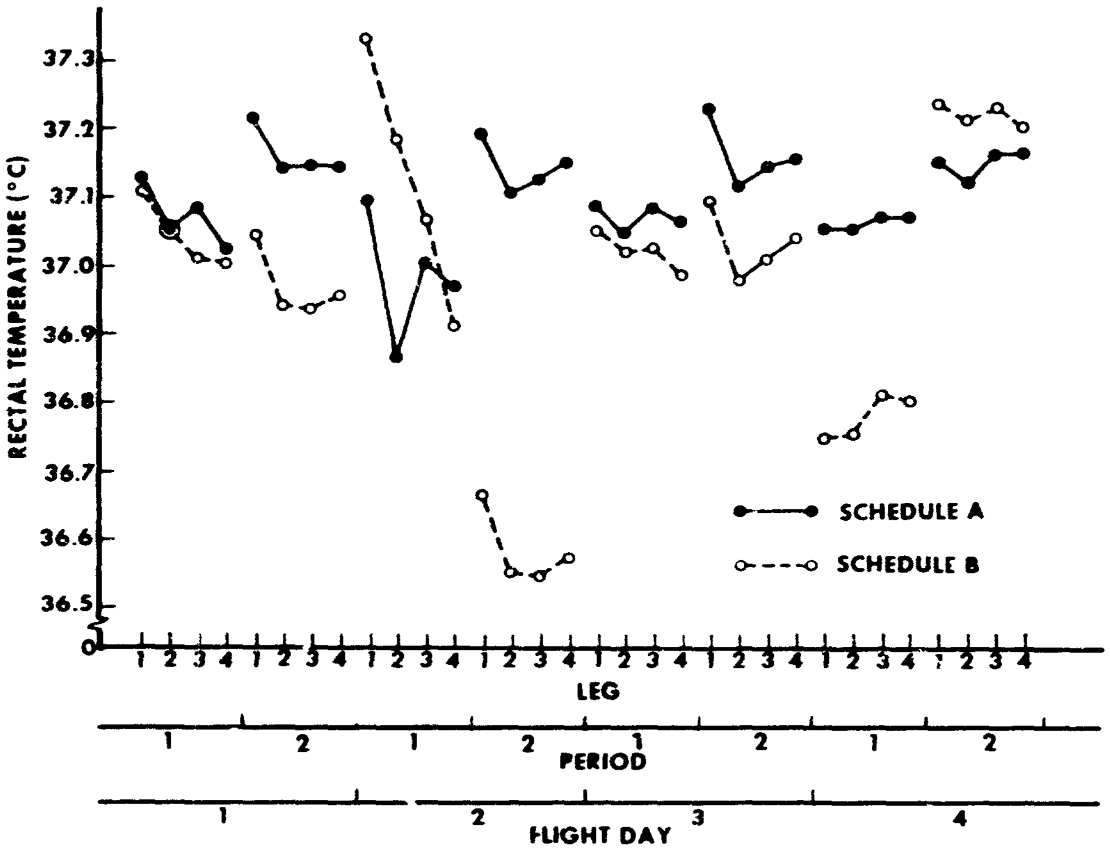


Figure 23. Rectal temperature for each leg of flight days 1 to 4.
(Each data point is the average over SLT, FMT, and DIPT epochs.)

TABLE 19. ANALYSES OF VARIANCE FOR RECTAL TEMPERATURE DECREMENT SCORES FROM FLIGHT DAY (FD) 1 TO FD 3 AND FD 1 TO FD 4

Source	df	SLT MS	SLT F	FMT MS	FMT F	DIPT MS	DIPT F
				FD 1 - FD 3			
Between Subjects (Ss)							
Group (G)	1	0.012	0.020	0.020	0.030	0.012	0.020
Ss within G	21	0.520	—	0.580	—	0.532	—
Within Ss							
Period (P)	1	0.068	0.280	0.116	0.480	0.100	0.410
G x P	1	0.084	0.360	0.080	0.330	0.072	0.300
Ss x P within G	21	0.240	—	0.244	—	0.240	—
Leg within P (L/P)	6	0.013	3.039**	0.016	2.120	0.009	0.999
G x L/P	6	0.001	0.197	0.004	0.502	0.009	0.999
Ss x L/P within G	125	0.004	—	0.008	—	0.009	—
				FD 1 - FD 4			
Between Subjects (Ss)							
Group (G)	1	0.001	0.000	0.004	1.010	0.002	0.000
Ss within G	19	0.380	—	0.396	—	0.404	—
Within Ss							
Period (P)	1	3.208	4.890*	3.436	5.230*	3.308	5.080*
G x P	1	2.840	4.330	2.916	4.440*	2.560	3.930
Ss x P within G	19	0.656	—	0.656	—	0.652	—
Leg within P (L/P)	6	0.059	7.824***	0.067	5.031***	0.047	3.343**
G x L/P	6	0.006	0.798	0.009	0.653	0.017	1.212
Ss x L/P within G	114	0.008	—	0.013	—	0.014	—

---

*p<.05
**p<.01
***p<.001

TABLE 20. ANALYSES OF VARIANCE COMPARING HEART RATE AND HEART RATE VARIABILITY AMONG SLT, FMT, AND DIPT SCORES ACROSS FLIGHT DAYS 1 TO 4, FOR SCHEDULES A AND B, RESPECTIVELY

Source	Schedule A					Schedule B				
	HR			HRV		HR			HRV	
	df	MS	F	MS	F	df	MS	F	MS	F
Subjects (Ss)	9	3139.49	11.40***	66.70	3.48**	8	4113.45	21.10***	351.92	8.51***
Day (D)	3	319.44	1.16	17.43	0.91	3	1028.30	5.27**	58.37	1.42
Ss x D	27	275.36	—	19.17	—	24	195.00	—	41.34	—
Period (P)	1	18.10	0.09	24.58	0.80	1	299.39	2.66	51.24	7.26*
Ss x P	9	192.10	—	30.88	—	8	112.44	—	7.06	—
D x P	3	163.51	2.65	5.93	0.37	3	311.15	4.36*	9.37	0.15
Ss x D x P	27	61.71	—	16.16	—	24	71.40	—	61.50	—
Task (T)	2	384.19	3.32	276.41	9.01**	2	235.52	1.42	231.02	1.59
Ss x T	18	115.67	—	30.67	—	16	166.16	—	145.50	—
D x T	6	16.36	0.75	4.03	0.38	6	18.87	0.75	30.25	0.83
P x T	2	17.46	1.06	2.72	0.09	2	56.00	2.95	4.20	1.24
Ss x P x T	18	16.51	—	29.69	—	16	19.00	—	3.39	—
Ss x D x T	54	21.76	—	10.72	—	48	25.17	—	36.27	—
D x P x T	6	16.47	0.61	7.91	0.59	6	30.61	1.05	72.73	1.39
Ss x D x P x T	54	27.23	—	13.40	—	48	29.08	—	52.31	—

---

*p<.05
**p<.01
***p<.001

TABLE 21. ANALYSES OF VARIANCE COMPARING HEART RATE AND HEART RATE VARIABILITY BETWEEN SLT AND FMT EPOCHS ACROSS FLIGHT DAYS 1 TO 4, FOR SCHEDULES A AND B, RESPECTIVELY

Source		Schedule A				Schedule B			
		HR		HRV		HR		HRV	
	df	MS	F	MS	F	MS	F	MS	F
Subjects (Ss)	9	1891.85	310.90***	74.94	34.82***	2300.65	195.36***	124.69	39.80***
Day (D)	3	241.53	1.37	9.38	1.24	784.98	6.31**	15.79	1.21
Ss x D	27	176.04	—	7.59	—	124.50	—	13.06	—
Period (P)	1	35.91	0.26	23.20	4.55	297.30	3.00	15.59	0.76
Ss x P	9	136.99	—	5.10	—	99.20	—	20.62	—
D x P	3	90.04	2.05	7.05	1.18	222.27	6.76**	31.86	3.53*
Ss x D x P	27	43.96	—	5.97	—	32.88	—	9.03	—
Task (T)	1	97.66	2.32	288.58	27.04***	210.91	6.11**	170.40	22.10**
Ss x T	9	42.06	—	10.67	—	34.53	—	7.71	—
D x T	3	20.39	2.26	5.79	3.07*	9.11	0.63	8.76	1.78
P x T	1	15.88	1.58	3.67	1.04	14.34	0.71	1.02	0.31
Ss x P x T	9	10.08	—	3.51	—	20.33	—	3.31	—
Ss x D x T	27	9.02	—	1.89	—	14.45	—	4.93	—
D x P x T	3	13.12	2.16	1.80	0.84	16.13	1.37	1.45	0.46
Ss x D x P x T	27	6.09	—	2.15	—	11.78	—	3.13	—

---

*p<.05
**p<.01
***p<.001

Strong individual differences still existed, but HR differences between tasks were uncovered for Schedule A. For Schedule B HR, the day effect was still strong ($p<.01$) and the Day x Period interaction was more obvious ($p<.01$). An HR task effect ($p<.01$, SLT = 95.21, FMT = 97.50 bpm) was also noted. This slight increase in HR may have been due to the slightly higher physical demand of the FMT as well as increased arousal. This increase may have shown up in Schedule B only because of the overall higher physical demands placed on the subjects by their fatiguing schedule.

For HRV with DIPT eliminated, Schedule B then showed a significant Day x Period interaction ($p<.05$). The HRV task effect for Schedule A has increased in significance ($p<.001$), and Schedule B also showed a significant task effect ($p<.01$, SLT = 12.47, FMT = 10.42 SDBPM; this appears comparable to Schedule A values). Schedule A also showed a Day x Task interaction (Fig. 24). Apparently the HRV for SLT b increased with each subsequent FD, but the FMT seemed more stable across days. Since a continued learning curve was apparent for Schedule A FMT scores (Fig. 10), the sustained concentration indicated by the FMT HRV could be interpreted as a function of learning in which sustained concentration across days was necessary for improving performance.

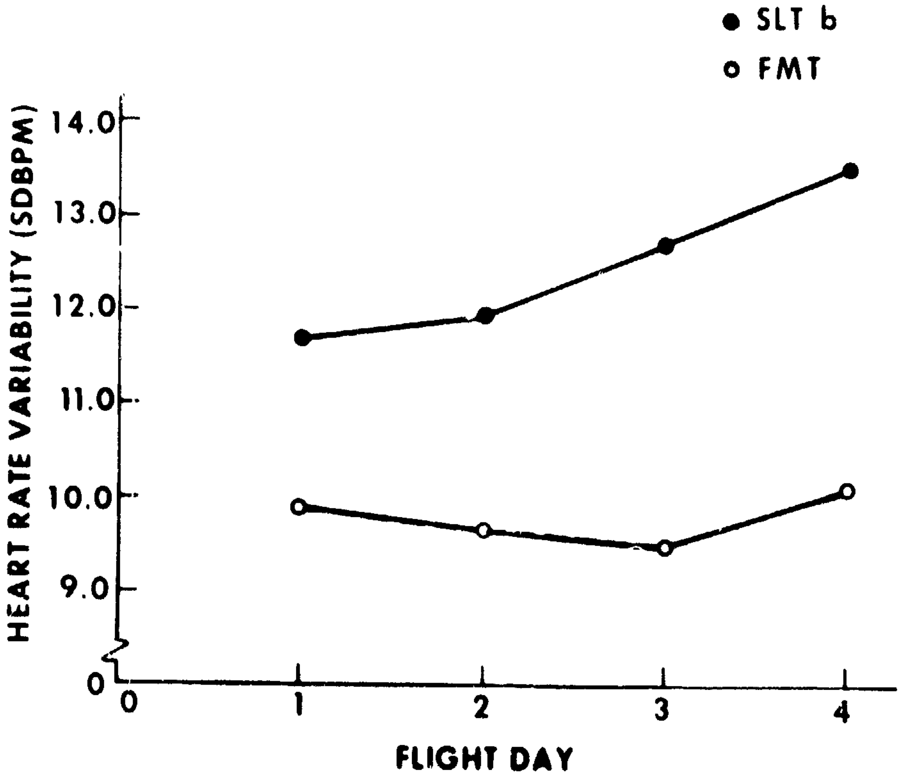


Figure 24. Day x Task interaction of heart rate variability during SLT b and FMT epochs for Schedule A. (Refer to Table 21.)

In summary, Hypothesis 8 was only partially confirmed. Only HRV for Schedule A properly ranked all three tasks. For SLT vs. FMT, HRV separated both tasks for both Schedules A and B. HR separated SLT from FMT for Schedule B alone. Only for subjects in a highly fatiguing situation does HR appear useful for separating task demands, and evidently the length of time of the HR sample must exceed some minimum to achieve a stable estimate. HRV, on the other hand, appears to be a more sensitive measure of task demands, especially in relatively unfatiguing situations.

Hypothesis 9: Interaction of Fatigue Level and Flying Performance Measures

Hypothesis 9 predicted an interaction between fatigue level and the respective arousal value of the FMT and the SLT. Because of the postulated difference in arousal levels between tasks, performance on each task was expected to be affected differently by fatigue--the more arousing task maintaining higher relative performance levels longer. The DIPT was excluded from testing in Hypothesis 9, because declines in DIPT performance were expected to be approximately equal to the FMT declines.

To test Hypothesis 9, an analysis similar to that for Hypothesis 8 was performed, with performance scores substituted for HR scores (Table 22). Again, strong individual differences were found (p<.001). For Schedule A, a task effect was noted (p<.05); but, more importantly, a Period x Task interaction (p<.01) was evident (Fig. 25). In contradiction to the arousal hypothesis, in the presence of only mild fatigue, the performance improved on the less arousing task (SLT) and degraded on the more difficult task (FMT) from periods 1 to 2. These results are more in line with typical fatigue effects: Performance degrades on difficult tasks first; and performance also degrades first on tasks which are newly learned or which have not reached plateau (Woodward and Nelson: 231).

In the more fatiguing situation, Schedule B demonstrated period (p<.01), Day x Period (p<.001), and task (p<.001) effects (Table 22). The effect of primary interest, however, was the Day x Period x Task interaction (p<.05) presented in Figure 26. On FD 1, an effect similar to that for Schedule A could be seen. On FD 2, however, during extreme fatigue, the SLT score degraded much more than the FMT score, thus supporting the arousal hypothesis. The SLT changed from 557 to 1215 CTSE, a 658-point decrement. The FMT changed from [illegible] to 1388 CTSE, a 468-point decrement. Similar decrements were apparent on FD 3, but of reduced magnitude. On FD 4, neither the SLT nor FMT degraded appreciably from Periods 1 to 2. Evidently, the arousal level of a task helps maintain performance only during periods of intense fatigue, after performance has already degraded significantly. Otherwise, in mild fatigue, performance on relatively less complex tasks is easier to maintain. Thus Hypothesis 9 is confirmed only for intense fatigue situations.

Hypothesis 10: Circadian Rhythm Effects

In order to determine if time awake prior to flight combined with the night circadian performance decrement caused greater performance decrement than that combined with cumulative fatigue effects, an analysis of variance

TABLE 22. ANALYSES OF VARIANCE COMPARING SLT AND FMT PERFORMANCE SCORES ACROSS FLIGHT DAYS 1 TO 4, FOR SCHEDULES A AND B, RESPECTIVELY

Source	Schedule A			Schedule B		
	df	MS	F	df	MS	F
Subjects (Ss)	9	1090810.00	40.454***	10	1135150.00	24.668***
Day (D)	3	170895.00	2.914	3	379904.00	1.674
Ss x D	27	58636.70	—	30	227003.00	—
Period (P)	1	116327.00	2.168	1	2980800.00	14.299**
Ss x P	9	53663.40	—	10	208457.00	—
D x P	3	2065.06	0.052	3	1376950.00	10.121***
Ss x D x P	27	39789.40	—	30	136052.00	—
Task (T)	1	1658570.00	7.779*	1	2100000.00	12.377**
Ss x T	9	213208.00	—	10	169676.00	—
D x T	3	70832.20	2.704	3	14420.10	0.322
Ss x D x T	27	26196.40	—	30	44786.70	—
P x T	1	387047.00	11.113**	1	25.51	0.001
Ss x P x T	9	34828.60	—	10	35587.20	—
D x P x T	3	25383.20	0.941	3	146676.00	3.187*
Ss x D x P x T	27	26964.00	—	30	46016.70	—

*p<.05
**p<.01
***p<.001

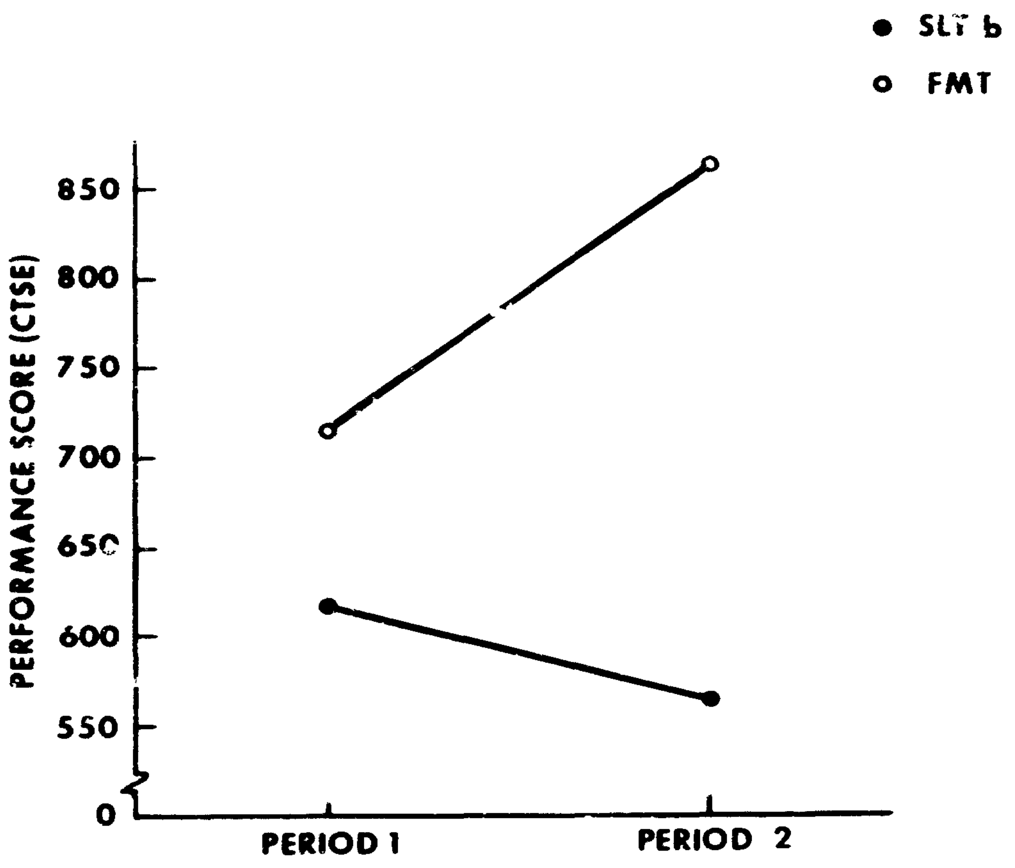


Figure 25. Period x Task interaction of SLT b and FMT scores for Schedule A. (Refer to Table 22.) [CTSE = combined total seconds of error.]

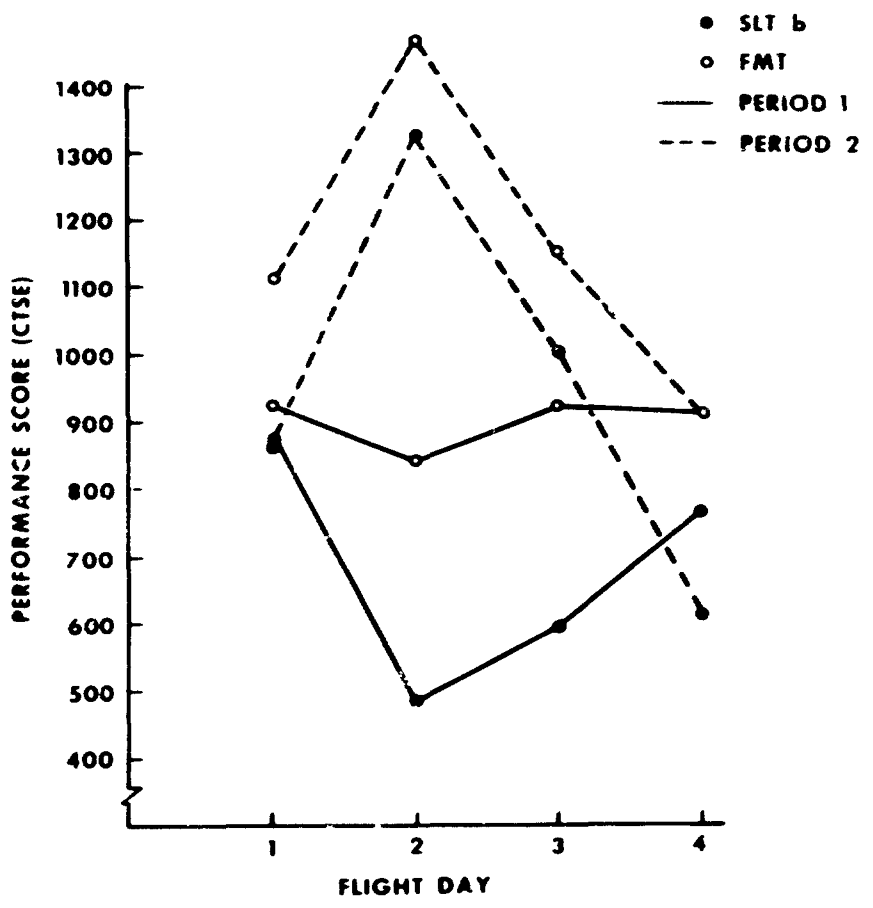


Figure 26. Day x Period x Task interaction of performance scores for SLT b and FMT for Schedule B. (Refer to Table 22.)

for SLT, FMT, and DIPT performance decrement scores for group and period was performed. The decrement scores were formed by subtracting FD 2 scores from FD 3 (Table 23). No difference was found between DIPT scores. This test could not distinguish which FD showed greater DIPT decrement. The slight improvement in FD 3, Period 2 scores--relative to that of FD 2--was not statistically significant (Fig. 12).

However, both SLT and FMT did show significant score differences between FD's 2 and 3 for both period and Group x Period effects. For the Group x Period effects ($p<.01$), see Figure 27.

For Schedule A, SLT performance was slightly better in Period 1 on either day; and FD 2 performance was, if different at all, slightly better than that of FD 3. For Schedule B, SLT performance for Period 1 was better on FD 2 than on FD 3; but performance for Period 2 was much worse on FD 2 than on FD 3.

For Schedule A, FMT performance was better on FD 3 than on FD 2, due to the learning effect discussed earlier. The performance change from Period 1 to 2 appeared negligible. The FMT performance for Schedule B followed the same pattern as the SLT performance, with only a reduced magnitude of decrement for the FMT. This information provided additional support for Hypothesis 9, showing that under conditions of extreme fatigue, performance degraded less on the arousing task during intense fatigue.

Thus, from a flying performance standpoint, Hypothesis 10 receives tentative confirmation: A long (12-hr) period of preflight awake time, coupled with flying during the normal sleep period, is more detrimental to flying performance than a similar preflight awake period coupled with the cumulative effects of a fatiguing night flight. This finding assumes that a 12-hr crew rest period with an 8-hr sleep period is provided immediately after the night flight to permit recovery from fatigue effects.

A similar analysis was performed for HR and HRV (Table 24). Again, the DIPT epoch showed no effects for either HR or HRV. For HR, both the SLT and FMT showed group effects ($p<.05$) and, again, Group x Period effects (SLT: $p<.05$; and FMT: $p<.01$), as presented in Figure 28. Now, for Schedule A during Period 1, SLT HR rose slightly from FD 2 to FD 3, but showed no change during FMT. For both SLT and FMT, HR was lower in Period 2 on FD 3 than on FD 2.

For Schedule B, both the SLT and the FMT HR were higher on FD 3 than FD 2 in Period 1. The increase is similar to the Period 1 SLT HR increase for Schedule A, but Schedule B's absolute levels are higher. This increase was due either to a circadian HR effect (HR was found to be higher in the afternoon) or to the possibility that Schedule B had a slightly higher normal HR than did Schedule A. At any rate, in Period 2, the SLT and the FMT HR are much higher on FD 3 than FD 2, contrary to the findings for Schedule A. The Period 2 HR would be expected to be low on FD 2 due to the circadian fall with sleepiness; but, in Period 2 on FD 3, HR rose with fatigue.

HRV showed only a period effect for the SLT ($p<.01$). During Period 1, both groups had lower HRV during the SLT on FD 2 than FD 3 and higher HRV during Period 2 on FD 2 (Fig. 29), probably because of a reduction of concentration due to increasing fatigue, boredom, or task mastery.

TABLE 23. ANALYSES OF VARIANCE FOR SLT, FMT, AND DIPT DECREMENT SCORES FOR SCHEDULES A AND B ON FLIGHT DAY 2 VS. 3

Source	SLT			FMT			DIPT		
	df	MS	F	df	MS	F	df	MS	F
Between Subjects (Ss)									
Group (G)	1	221028.552	0.85	1	19612.488	0.09	1	330.000	0.01
Ss within G	21	259320.120	—	21	223499.868	—	20	33252.252	—
Within Ss									
Period (P)	1	1427079.000	12.29**	1	675244.440	5.91*	1	67224.544	2.54
G x P	1	1069177.080	9.21**	1	942622.440	8.25**	1	21360.682	3.23
Ss x P/G	21	116099.732	—	21	114201.496	—	20	26431.500	—

---

*p<.05
**p<.01

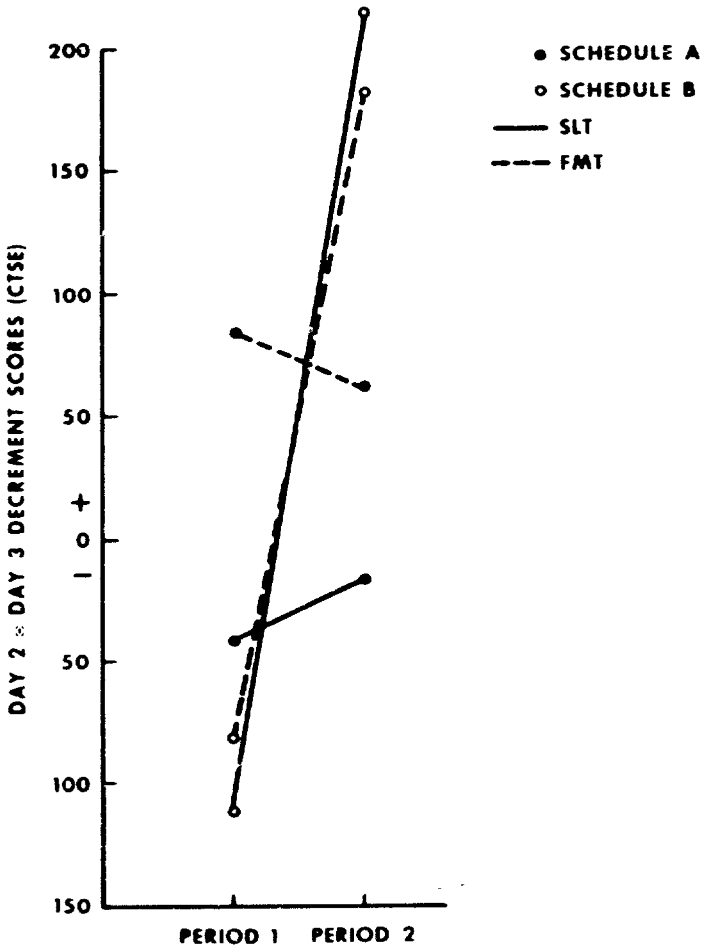


Figure 27. Group x Period interaction for separate analyses of SLT and FMT decrement scores from flight day 2 to 3. (See Table 23.) [CTSE = combined total seconds of error.]

TABLE 24. ANALYSES OF VARIANCE FOR HEART RATE AND HEART RATE VARIABILITY DECREMENT SCORES DURING SLT, FMT, AND DIPT EPOCHS FOR SCHEDULES A AND B ON FLIGHT DAY 2 VS. 3

Source	SLT			FMT		DIPT		
	df	MS	F	MS	F	df	MS	F
				HR				
Between Subjects (Ss)								
Group (G)	1	1342.596	4.38*	1912.336	4.68*	1	1518.924	2.61
Ss within G	20	306.360	—	408.680	—	19	582.888	—
Within Ss								
Period (P)	1	9.864	0.09	291.024	3.15	1	105.556	0.51
G x P	1	541.292	5.17*	801.840	8.67**	1	734.788	3.58
Ss x P/G	20	104.636	—	92.532	—	19	205.136	—
				HRV				
Between Subjects (Ss)								
Group (G)	1	24.096	1.48	79.912	3.68	1	160.752	2.23
Ss within G	20	16.332	—	21.712	—	19	71.932	—
Within Ss								
Period (P)	1	71.072	12.50**	27.624	3.73	1	4.224	0.04
G x P	1	11.512	2.02	36.420	4.91*	1	188.388	1.61
Ss x P/G	20	5.684	—	7.416	—	19	116.952	—

*p<.05
**p<.01

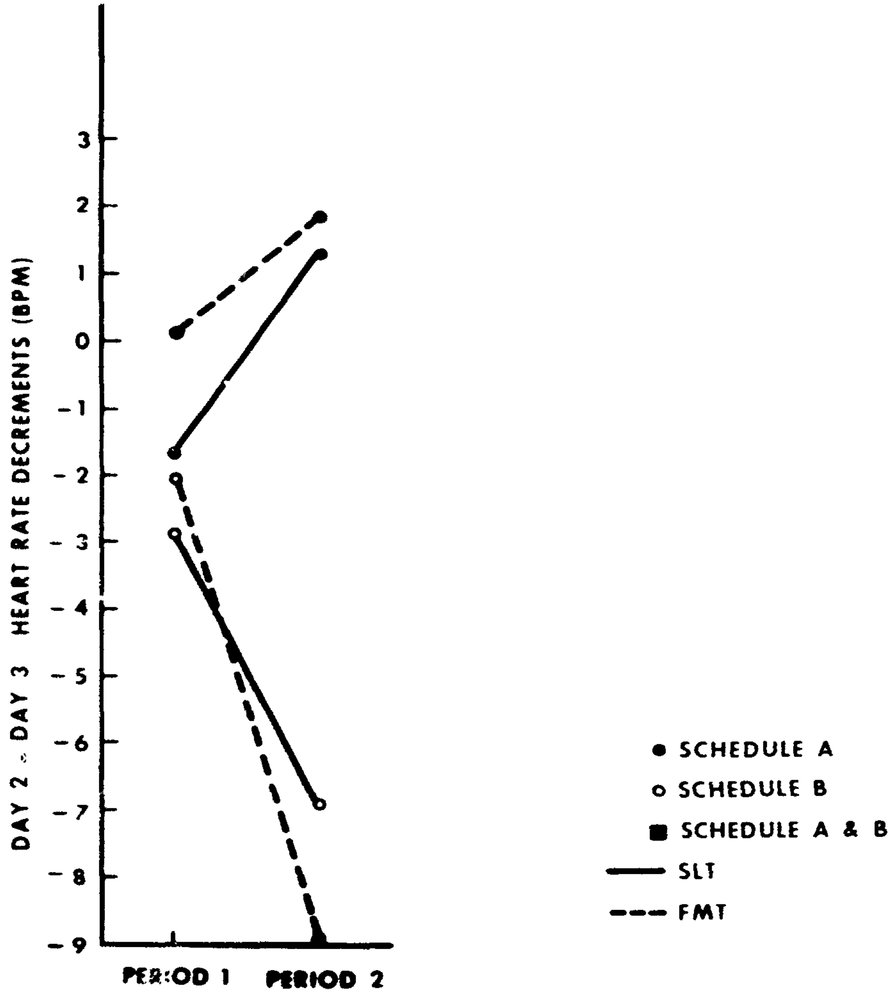


Figure 28. Group x Period interaction of heart rate decrements in beats per minute (BPM) from flight day 2 to 3. (See Table 24.)

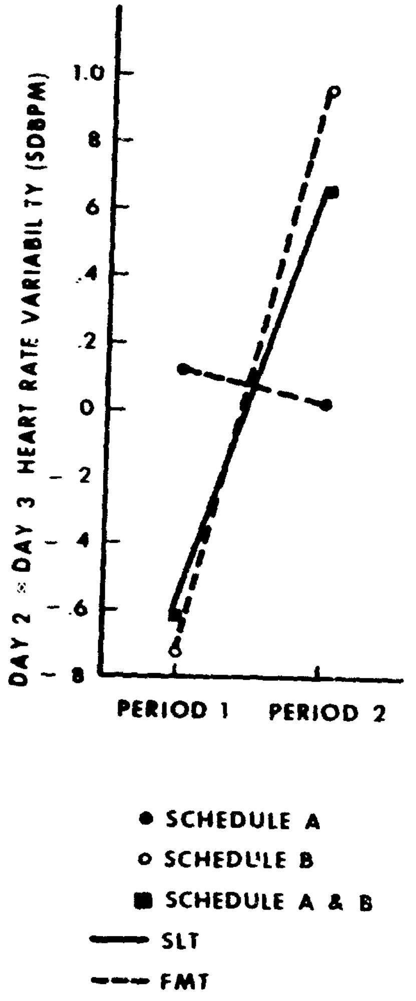


Figure 29. Group x Period interaction of heart rate variability decrements in standard deviation of beats per minute (SDBPM) from flight day 2 to 3. (Refer to Table 24.)

For the FMT, however, a Group x Period interaction ($p<.05$) occurred (Fig. 29). Little change took place in HRV either between periods, or from FD 2 to FD 3, for Schedule A. This finding indicated maintained concentration, associated with the learning effect. For Schedule B, HRV was lower in Period 1 on FD 2 than on FD 3, but higher on FD 2 in Period 2. Overall, on FD 2 a large increase from Period 1 to Period 2 was noted, and on FD 3, a slight decrease from Period 1 to Period 2; but the greatest HRV was on FD 3, Period 2, Leg 4 (Fig. 21).

A similar analysis was performed for rectal temperature values (Table 25). SLT, FMT, and DIPT epochs were highly significant for Period and Group x Period effects ($p<.001$), as shown in Figure 30. For both Schedules, the SLT, FMT, and DIPT values were almost identical. For Schedule A, Period 1 on FD 2 was only slightly, if at all, lower than on FD 3. For Period 2, no difference from FD 2 to FD 3 was apparent. Level differences did exist, due to circadian effects between the periods (as already noted in Fig. 22).

For Schedule B, Period 1 was much higher and Period 2 was much lower on FD 2 than on FD 3. This difference would appear to be due to the circadian temperature cycle, because the two periods began 12 hr apart.

A final analysis of variance was performed on subjective fatigue and SSS decrement scores from FD 3 to FD 2 (Table 25). Only for the SSS was a mild time-of-day effect found ($p<.05$). The mean SSS decrements were: preflight 1 = 0.54; midflight 1 = 0.33; preflight 2 = 0.08; midflight 2 = 0.13; and postflight 2 = 0.13. These decrements can be interpreted to mean that both groups reported slightly more sleepiness at the beginning of FD 3 than of FD 2, but slightly less sleepiness at the end of FD 3. Most of this effect would have been expected to be due to Schedule B reports; but, surprisingly, no group differences were found. The rate of sleepiness decrement was apparently not significantly different between groups.

Hypothesis 10 was confirmed in the predicted directions for HR (decrease), HRV (increase), and rectal temperature (decrease) for the night (FD 2) flight vs. the following day (FD 3) flight. For subjective report, only the SSS measure supported Hypothesis 10; and only in Period 2 of the night flight were SSS scores greater than those on the following day flight. SSS scores were higher in Period 1 of FD 3 than in Period 1 of FD 2.

Relationships Among Dependent Measures

Rather than calculate and report all possible correlations among the many types of variables which could be constructed, three sets have been selected as the more interesting and fixed for the following analyses: (a) values averaged over the entire FD; (b) decrement or change scores taken from two points in time on a given FD; and (c) for subjective fatigue and sleepiness data only, the postflight value obtained after the second flight on a given FD. This section of the analyses is exploratory, and not an exhaustive evaluation of the possible relationships among dependent variables.

TABLE 25. ANALYSES OF VARIANCE FOR RECTAL TEMPERATURE, SUBJECTIVE FATIGUE, AND STANFORD SLEEPINESS SCALE DECREMENT SCORES FOR SCHEDULES A AND B ON FLIGHT DAY 2 VS. 3

Source	SLT			FMT		DIPT		
	df	MS	F	MS	F	df	MS	F
			Rectal Temperature					
Between Subjects (Ss)								
Group (G)	1	1.118	2.22	0.972	1.73	1	1.388	2.52
Ss within G	19	0.536	—	0.564	—	18	0.552	—
Within Ss								
Period (P)	1	3.272	27.69***	3.484	30.70***	1	2.900	21.74***
G x P	1	4.120	34.87***	4.224	37.21***	1	2.700	27.74***
Ss x P/G	19	0.118	—	0.112	—	18	0.132	—
		Fatigue		SSS				
Between Subjects (Ss)								
Group (G)	1	73.500	1.54	2.040	0.34			
Ss within G	22	47.720	—	6.040	—			
Within Ss								
Time (T)	4	28.165	1.70	10.270	3.20*			
G x T	4	23.710	1.43	6.105	1.90			
Ss x T/G	88	16.575	—	3.210	—			

---

*p<.05

***p<.001

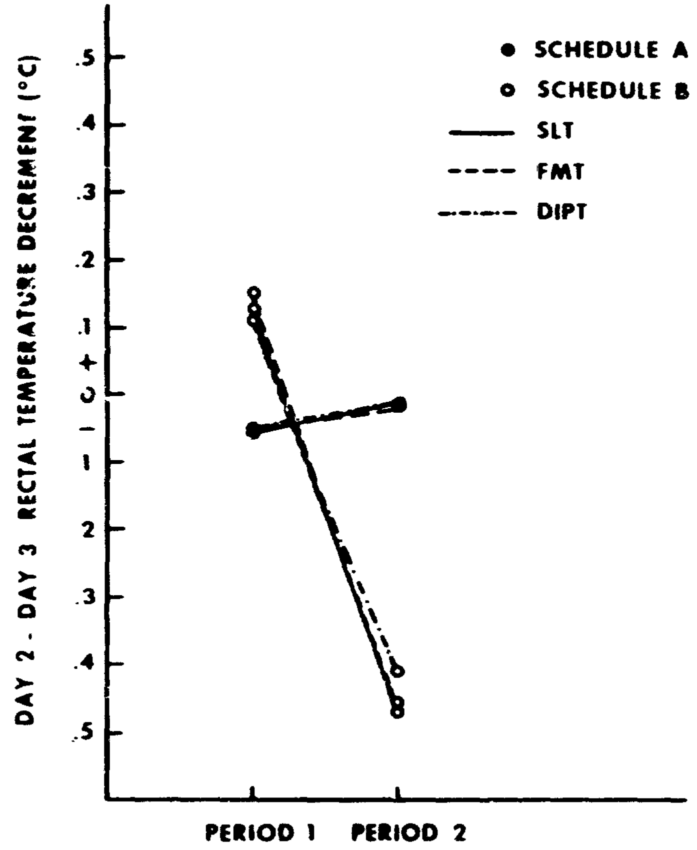


Figure 30. Group x Period interaction of rectal temperature decrement from flight day 2 to 3. (Refer to Table 24.)

Hypothesis 11: Correlations of Flying Performance and Information Processing Measures

In order to test Hypothesis 11, Pearson product-moment correlations were calculated for the averaged FD 4 SLT, FMT, and DIPT values for all subjects having no missing data. The FD 4 averaged values were chosen as being the most stable estimate of a subject's performance capability. If a meaningful correlation existed, it should probably have been seen when the subject had become most familiar with the tasks. (The results are presented in Table 26.)

Regardless of the difference in task complexity, the SLT and FMT scores appear strongly correlated ($r = .75$, $p<.001$). A moderate ($p<.05$) and similar correlation also seems to exist between the subject's DIPT score and his flying performance score. This relationship is important if DIPT performance is ever to be used for an analogue of his flying performance in an operational setting. Thus, Hypothesis 11 is confirmed.

Hypothesis 12: Correlations of Performance Decrement with Subjective Fatigue and Physiological Indices

Another way to use performance scores is to attempt to use changes in one measure to predict changes in the other. Accordingly, decrement performance scores were constructed for each subject having no missing data by subtracting Period 1, Leg 2 scores from those of Period 2, Leg 3. As in the arousal analysis, these legs were chosen to avoid extraneous takeoff and landing effects on performance. The next step was to look at the correlations across all 4 FD's; for FD's 2 and 3 should have induced more decrement from Period 1 to Period 2 than did FD's 1 or 4, and the increased decrement range was more likely to demonstrate the correlation. Moreover, such predictions would be most useful during the days of high fatigue. (These results are also presented in Table 26.) The only significant correlation between performance score decrements on FD 1 was the decrement relationship between the SLT and FMT ($p<.05$). The only other significant correlations for performance decrement were on FD 2 for SLT and FMT (now: $p<.001$), and between SLT and DIPT ($p<.01$). The lack of correlation between FMT and DIPT could indicate that the information processing and other task attributes were more similar between the DIPT and SLT, but the greater range of SLT decrement may possibly have provided greater opportunity for correlation than the FMT. Even though performance decrements were noted from FD 1 to FD 3 for the SLT and FMT, the decrements were not correlated on FD 3 between subjects. On FD 4, no significant correlations between performance decrements were found. True within-subject decrement was probably minimal or not occurring at all on FD's 4 or 1, thus accounting for the lack of correlation between measures.

To determine if performance declined when changes in subjective fatigue and sleepiness were reported, Pearson product-moment correlations were calculated between the previous SLT, FMT, and DIPT decrement scores and the other variables just mentioned. Changes in subjective fatigue and SSS scores were calculated by subtracting the Period 1 preflight score from the Period 2 postflight score.

TABLE 26. CORRELATIONS[a] OF AVERAGE PERFORMANCE SCORES AND DECREMENTS[b] IN PERFORMANCE WITH EACH OTHER, WITH CHANGES[c] IN SUBJECTIVE FATIGUE AND IN SSS SCORES, AND WITH PERIOD 2 POSTFLIGHT SUBJECTIVE FATIGUE AND SSS SCORES

		Flight Day (FD)			
		FD 1	FD 2	FD 3	FD 4
SLT	w. FMT	—	—	—	.75***
SLT	w. DIPT	—	—	—	.47*
FMT	w. DIPT	—	—	—	.53*
DSLT	w. DFMT	.45*	.65***	.16	.04
DSLT	w. DDIPT	.08	.63**	.01	.03
DFMT	w. DDIPT	.07	.30	.18	.33
DSLT	w. CSF	.31	-.50**	-.10	-.24
DFMT	w. CSF	-.18	-.37	.23	.14
DDIPT	w. CSF	.25	-.49*	.05	-.30
DSLT	w. CSSS	-.20	.60**	.14	.41
DFMT	w. CSSS	.39	.64***	.01	-.12
DDIPT	w. CSSS	-.27	.52**	-.08	.25
DSLT	w. PFSF	.23	-.76***	-.24	.18
DFMT	w. PFSF	-.31	-.54**	.11	.01
DDIPT	w. PFSF	.37	-.48*	.03	-.18
DSLT	w. PFSSS	-.18	.77***	.32	.09
DFMT	w. PFSSS	-.14	.62***	.05	-.04
DDIPT	w. PFSSS	-.14	.47*	-.07	-.05

---

*p<.05; **p<.01; ***p<.001

[a] Due to missing data, N's varied from 22 to 24; [b] Decrement derived by subtracting each subject's Period 2, Leg 3 performance score from that of Period 1, Leg 2; [c] Change score derived by subtracting each subject's Period 2 postflight score from his Period 1 preflight score.

[Note: Average performance scores = SLT, FMT, and DIPT; Decrements in performance = DSLT, DFMT, and DDIPT; C = change; D = decrement; DIPT = Discrete Information Processing Test; FMT = flight maneuver test; PFSF = postflight subjective fatigue; SLT - straight and level test; and PFSSS = postflight SSS.]

Correlations were also calculated between: the Period 2 postflight subjective fatigue and SSS score, and the performance decrement scores. The reason for using this score was that subjects who started a flight duty day in a fatigued state, and who had a higher initial fatigue level, may have shown a fatigue or sleepiness decrement score equal to initially nonfatigued subjects over the course of the day. However, the higher final fatigue score would discriminate between these two groups, and might show a stronger relationship with the performance decrement score than a fatigue decrement score.

Significant correlations between performance decrement, fatigue, and sleepiness were obtained only on FD 2, and all were in the predicted direction (Table 26). This finding indicates that only when severe fatigue is induced will subjective fatigue correlate with changes in performance. However, during periods of high fatigue the relationship will be, in some cases, substantial.

The change in subjective fatigue score was only moderately related to the change in SLT and DIPT performance, and did not correlate with changes in FMT performance. But the change in SSS score, while still correlated with SLT and DIPT, was apparently most strongly correlated with FMT scores.

The postflight scores seemed to be more highly correlated with performance decrement, particularly the SLT and FMT changes. This relationship appeared just as strong for either subjective fatigue or sleepiness scores.

Change scores were constructed for each subject having no missing data for HR and rectal temperature scores by subtracting Period 1, Leg 1 from Period 2, Leg 4 values for each SLT, FMT, and DIPT epoch, for each FD. For this initial analysis, takeoff and landing effects (if any) were to be included in these physiological measures, and values maximally separated in time were to be used. In addition, the average HRV for each subject was calculated for each SLT, FMT, and DIPT epoch, for each FD. The theory was that the average HRV, as an index of a subject's overall ability to concentrate, over the whole day, would be a better indicator of performance decrement than HRV decrement within an epoch.

These sets of physiological variables were then correlated with their corresponding SLT, FMT, or DIPT performance decrement score (Table 27). As can be seen, the average HRV never showed a relationship with decrements in performance. Decreases in rectal temperature were strongly associated with decreases only in SLT flying performance on FD 2. On FD 2, from Periods 1 to 2, Schedule A showed little decrement but significant increase in body core temperature. The strong correlation was probably due to the fact that Schedule B demonstrated both a severe performance decrement and a drop in rectal temperature (Figs. 11 and 23). On FD 3, even though Schedule B showed a performance decrement, the body core temperature was not significantly different from that on FD 1 (Table 19). Thus, as in Schedule A, any performance decrement would have been associated with an increase in rectal temperature due to its normal circadian rise throughout the day.

Changes in HR showed no relation to changes in performance except for DIPT scores on FD's 3 and 4 ($p<.05$). Because almost all subjects showed a decrease in HR from Period 1 to Period 2, the positive correlation on FD 3

TABLE 27. CORRELATIONS[a] OF DECREMENTS[b] IN PERFORMANCE WITH CHANGES[c] IN HR, WITH AVERAGE HRV, AND WITH CHANGES[c] IN RECTAL TEMPERATURE

		Flight Day (FD)			
		FD 1	FD 2	FD 3	FD 4
DSLT	w. CHR	-.16	-.20	.23	.28
DSLT	w. AHRV	.23	-.12	.02	.30
DSLT	w. CRT	-.03	-.74***	-.14	.05
DFMT	w. CHR	-.35	-.34	.05	-.34
DFMT	w. AHRV	.35	.32	.25	.19
DFMT	w. CRT	-.15	-.61**	-.03	-.01
DDIPT	w. CHR	.07	-.26	.41*	-.50*
DDIPT	w. AHRV	.29	-.37	.27	-.35
DDIPT	w. CRT	-.28	-.70***	-.20	-.02

*p<.05
**p<.01
***p<.001

[a] Due to missing data, the number of subjects varied from 21 to 24.

[b] Decrement derived by subtracting each subject's Period 2, Leg 3 performance score from that of Period 1, Leg 2.

[c] Change score for each physiological variable derived by subtracting each subject's Period 2, Leg 1 physiological score from that of Period 1, Leg 4, obtained during the corresponding performance task epoch.

[<u>Note</u>: Changes in HR - CHR; average HRV = AHRV; changes in rectal temperature = CRT; and decrements in performance = DSLT, DFMT, and DDIPT.]

meant that the greater the decrease in HR, the greater the change in information processing capability. On FD 4, however, the negative correlation indicated that the greater the decrease in HR, the less the change in information processing capability. On FD 3, decreases in arousal (as indicated by HR changes) indicated a reduced information processing capability. On FD 4, subjects who showed the greater HR change, indicating greater decrease in arousal, showed less decrease in information processing ability from morning to afternoon. Schedule B, as a whole, showed little change from Period 1, Leg 2 to Period 2, Leg 3 (Fig. 13). Within this group, subjects who maintained their arousal may not have dropped as much in information processing capability.

Thus, Hypothesis 12 received strong support only for the predicted relationships between changes in performance and in subjective fatigue and sleepiness, the postflight subjective fatigue and sleepiness scores, and changes in rectal temperature. Like the relationships between changes for performance measures, these relationships were only apparent during the severe fatigue condition on FD 2.

Hypothesis 13: Correlations Between Subjective Reports of Fatigue and Sleepiness

Correlations between change scores and postflight scores for subjective fatigue and sleepiness are presented in Table 28. Except for the lack of correlation on FD 1 between changes in subjective fatigue and the postflight SSS score, all other combinations are highly significant ($p<.001$) and in the predicted direction, up through FD 3. Note that, on each FD, the correlation between changes in subjective fatigue and the postflight SSS score is the lowest among the six. Except for FD 1, changes in SSS and postflight subjective fatigue scores are the second lowest among the six. The most highly correlated scores across all FD's seem to be the postflight subjective fatigue and sleepiness scores. The reason may be that subjects are more accurate at estimating fatigue or sleepiness when they are tired. Thus Hypothesis 13 is confirmed: In this study, fatigue and sleepiness can be interpreted to have a similar dimension.

Hypothesis 14: Correlations of Subjective Reports of Fatigue and Sleepiness with Body Temperature

Finally, to determine if changes in or postflight subjective reports of fatigue and sleepiness were related to changes in body core temperature, Pearson product-moment correlations were computed between each of these measures (Table 28). The change in body core temperature was computed by subtracting Period 1, SLT 1, Leg 1 values from Period 2, SLT 3, Leg 4 values; for these scores were most closely associated in time with the subjects' preflight and postflight subjective fatigue and sleepiness reports.

The change in subjective fatigue was never found to be significantly correlated with the subjects' change in rectal temperature. The other three measures were found to be correlated with temperature change, but only on FD 2. Postflight SSS may have been more significantly correlated ($p<.001$) than postflight subjective fatigue ($p<.01$) or change in SSS ($p<.05$); however, these

TABLE 28. CORRELATIONS[a] OF CHANGES[b] IN SUBJECTIVE FATIGUE AND SSS SCORES AND PERIOD 2 POSTFLIGHT SUBJECTIVE FATIGUE AND SSS SCORES WITH EACH OTHER AND WITH CHANGES[c] IN RECTAL TEMPERATURE DURING THE SLT

	Flight Day (FD)			
	FD 1	FD 2	FD 3	FD 4
CSF w. PFSF	.70***	.75***	.71***	.62***
CSF w. CSSS	-.66***	-.80***	-.75***	-.78***
CSF w. PFSSS	-.38	-.67***	-.65***	-.40*
CSSS w. PFSF	-.78***	-.75***	-.65***	-.43*
CSSS w. PFSSS	.82***	.83***	.83***	.59**
PFSF w. PFSSS	-.82***	-.91***	-.84***	-.71***
CSF w. CRT's	.07	.33	.28	.10
PFSF w. CRT's	.00	.57**	.10	.40
CSSS w. CRT's	-.08	-.51*	-.32	-.01
PFSSS w. CRT's	.00	-.64***	-.27	.07

*p<.05
**p<.01
***p<.001

[a] Due to missing data, the number of subjects varies from 22 to 24.

[b] Change score derived by subtracting each subject's Period 2 Postflight score from that of his Period 1 Preflight score.

[c] Change score derived by subtracting each subject's rectal temperature value for Period 2, Leg 4, from that of Period 1, Leg 1, obtained during the respective SLT epcoh.

[Note: Changes in subjective fatigue = CSF, in SSS = CSSS, and in rectal temperature during the SLT = CRTS; postflight SSS = PFSSS; and postflight subjective fatigue = PFSF.]

differences were not tested. The directions of the correlations were in accord with expectations for Schedule B: The greater the drop in rectal temperature, the greater the increase in subjective fatigue or sleepiness. Thus, during periods of high fatigue and falling body temperature, changes in sleepiness and postflight subjective fatigue were correlated. Note that on FD 3, when all subjects' body temperatures were assumed to be rising and fatigue and sleepiness increasing, no significant correlations were found. Hypothesis 14 was thus confirmed only for intense fatigue periods.

DISCUSSION

General

The results of this study refute Cameron's statement (42) that true performance decrement cannot be shown with the passage of time. The requirements of flying the GAT-1 pushed the subjects below their baseline performance capability during periods of intense fatigue. These periods are considered to be flights which started after the subjects had been awake for at least 12 hr. In contradiction to studies such as Pierson's (69), reports of subjective fatigue and sleepiness were found to be correlated with work performance decrement. Again, however, in order for this relationship to be apparent, the subjects had to have experienced intense fatigue. The Schedule B mean subjective fatigue scores of six or lower are taken as evidence that, in this study, subjects were experiencing intense fatigue. Thus, when subjects report such low scores, their performance on relatively complex tasks can be expected to be significantly below baseline and cannot be markedly improved without a substantial rest period.

From the present study, it is theorized that as a subject expends effort to maintain performance, subjective fatigue will normally increase. Not until the drop in performance occurs, however, can this correlation be obtained. Also adding to this effect is the supposition that subjects are more accurate in estimating feelings of high fatigue or sleepiness than of moderate fatigue or sleepiness. In other words, a subject is more able to distinguish between feeling "extremely tired" and "slightly pooped" than between "very refreshed" and "somewhat fresh." (Testing these assumptions is a task for future research.) The results of this study support the statement (Welford, Brown, and Gabb: 213) that, when valid subjective and physiological manifestations of fatigue occur, an underlying performance decrement is always present if one knows where to look for it.

Even though the significance levels were not as high in most cases as the SLT and FMT, the DIPT was shown to be sensitive to fatigue effects of less than 24 hr of sleep loss. Hitherto, it had been thought that a test had to last at least 30 min to show this effect (Wilkinson: 224). In the present study, the workload was much higher than that in most previous sleep loss experiments, but the subject testing time was much shorter than 30 min. The SLT score was generated from 30 min of testing per hour of flight time, and the FMT from 13 min of testing per hour; but the DIPT required on the average, for all 4 days of Schedule A, only 63.2 sec per hour (Fig. 31). For Schedule B, the DIPT average was somewhat longer--81.3 sec--for two possible reasons. First, on FD 1, where their scores might not yet have plateaued, the mean was 93.1 sec. On Schedule B, FD 2, however, the Period 1 mean was 65.8 sec--very similar to the Schedule A, FD 2 mean of 65.3 sec. This information provided additional evidence that, by the beginning of FD 2, Schedule B DIPT scores were initially not significantly different from those of Schedule A.

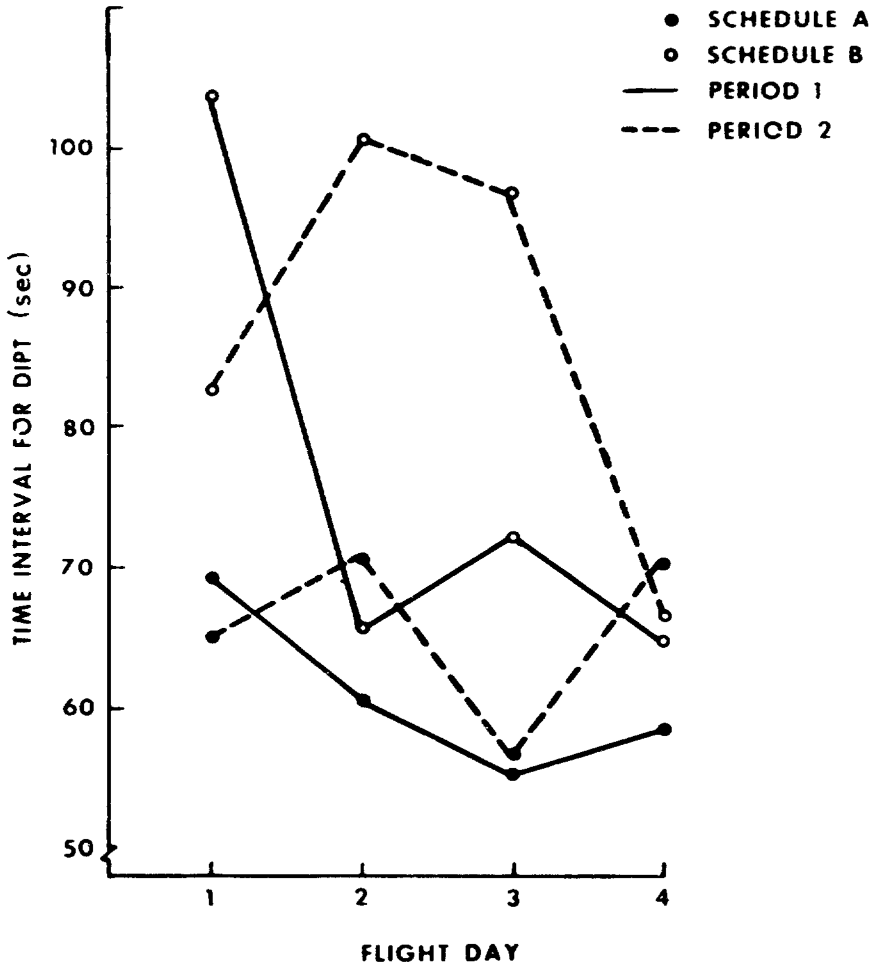


Figure 31. Average time interval (in seconds) required by each group to complete the DIPT. (The DIPT was administered to each subject four times per period.)

The Period 2 mean for Schedule B was 94.6 sec. This increase was not interpreted as a learning problem, but as an effect of fatigue. As the subjects became truly fatigued, their responses became more variable and they were required to produce additional blocking scores before they met the adaptive algorithm criterion of two consecutive blocking scores of less than 200 ms difference. In addition, a necessary increase occurred in the time required to perform the task; for, when fatigued, the subjects had blocking scores which were higher than baseline. This increase would be negligible as compared with that increase, in task time, which was due to producing additional block scores. In fact, it is possible to produce equal DIPT scores on two separate occasions which differ in time required to complete the DIPT.

Because of these two effects, future research should attempt to use, as modifier variables for the final DIPT score: the time interval required to perform the task; and the number and variability of the blocking scores produced while performing the task. This method may further improve the sensitivity of the test to fatigue and other stressor effects, and may indicate when a subject has plateaued during training.

On FD 3, the same time interval increase was noted in Period 2 for Schedule B; but, on FD 4, the daily mean time intervals had returned to approximately the same level as those for Schedule A (64.3 vs. 65.6 sec). This information supports the notion that the Schedule B subjects had minimal carryover effects from their two fatiguing FD's, and that one 12-hr crew-rest period with an 8-hr sleep period provided sufficient recovery to perform a 9-hr flight, provided the crew duty day started 1 hr after waking.

The DIPT achieved many of the requirements of a field-portable test device. Test admininstration time was short. Apparently, the DIPT could be quickly learned to a stable baseline. Schedule A subjects showed no learning effects, even though they received less than 2 hr total practice time on the DIPT; 1.5 hr of that time was on a device which was not adaptive in nature. Their SLT and FMT scores continued to improve, however, even after 15 hr of practice flying the GAT-1.

Schedule B subjects had evidently not received sufficient DIPT training to reach baseline until the start of FD 2. In the future, increased practice time may have to be provided. However, if the entire practice time is spent on an adaptive device, 2 hr may still be sufficient. This possibility should be investigated in subsequent research.

Both the SLT and FMT evidently did not have to last as long as they did in order to demonstrate significant decrement. Future research into the minimal time necessary for these flight tasks may show that very short psychomotor tracking tasks could also be adapted for field-research performance assessment. In addition, a "blocking" score might be obtained from the tracking tasks to further enhance their sensitivity to fatigue or their stressors. Tracking movements may become more erratic, and the periods of no control input would be predicted to increase with fatigue.

For several measures, Schedule B appeared to be slightly different (although not significantly so) from Schedule A on FD 1. Performance scores indicated possibly lower ability; HR indicated possibly higher arousal; and rectal temperature indicated possibly higher physical exertion. BD 2 could have been made the first FD for both groups; and BD 3, the nonflying crew duty day for both. FD 2 would still have begun on the evening of BD 3 for Schedule B, and of BD 4 for Schedule A. Thereafter, the design would have been the same. This approach would have made initial treatment for the two groups even more similar, and possibly would have minimized any group differences. The approach was not used in the present study, because data on regular crew-duty cycles for 4 consecutive FD's was considered more important than any minor differences which might have been induced between groups. However, any differences obtained did not affect the results or conclusions, especially since most analyses used subjects as their own control.

The rectal temperature measure provided the best indication that the subjects had residual effects from the intense fatigue flights. Future research should determine, however, if the body temperature control system would be upset with extremely high workloads which generate high fatigue but which do not interfere with the subject's normal sleep period. Operationally, this period might correspond to 10 days of high workload flying duty occurring from 0700 to 2200 hr. In most cases, high fatigue is generated by long work hours which impinge on a subject's normal sleep period, even though the duration of his sleep period may be 8 hr; circadian disruption is thus confounded with intense fatigue. Also, falling body temperature was found to be most strongly related to performance decrement. When performance was declining but body temperature was rising due to circadian effects, minimal correlation was obtained.

Future analyses of these data should determine if absolute body temperature is related to flying performance scores, as was found by Kleitman and Jackson (123). In addition, analysis of the psychological test data collected on each subject might be useful in providing hypotheses for future research to account for individual differences in performance and susceptibility to fatigue effects.

Generally, HR and HRV supported the arousal theory. Subjects in less arousing situations had lower HR and higher HRV. In contrast to Corcoran's findings (52), HR did not fall constantly for these subjects, who were in a more demanding situation than his. Schedule B showed an HR increase from FD 2 to FD 3, probably due in part to a circadian effect (Fig. 19). This increase might have been due, however, either to cumulative fatigue effects or to partial recovery from the FD 2 fatigue effects.

In addition, the HR for Schedule B tended to rise slightly from Period 2, Leg 3 until the end of the flight (Fig. 19). At this point the increasing HR was probably reflecting the increasing effort on the part of the subjects to stay awake.

On the other hand, HRV appeared to have a slightly higher overall FD 3 average than FD 2, 11.4 vs. 10.0 SDBPM (Fig. 20); but the Schedule B, Period 2 average was higher on FD 2, 12.02 vs. 10.61 SDBPM (Fig. 20). Even though HR seemed to fall and rise with fatigue, HRV appeared to increase continually with fatigue regardless of HR. HR may fall and then rise with fatigue, somewhat independently of HRV, thus adding to the difficulty of finding a relationship between HR and performance decrement.

The change in HR and average HRV variables chosen for correlation with performance decrement were not significant on FD 2. This lack of correlation is in contrast with almost all other factors chosen for correlation during this intense fatigue period. The finding cannot be taken as proof that HR and HRV values are unrelated to performance decrement, but only as an indication that: different points in time should be chosen to construct the HR and HRV variables for future attempts at correlation; and a more precise association in time between the physiological variables and the performance epoch may be necessary. Possibly a change score for HRV would be more useful in predicting performance decrement, since HRV was found to increase more or less continually with fatigue.

By showing that complex tasks have greater arousal value (Hypothesis 8) and are somewhat less susceptible to fatigue effects than simple ones (Hypothesis 9), the requirement for optimizing workload levels has become more apparent both for aircraft cockpit design and for aircrew scheduling. Additional research is needed to provide information on relative workloads among duties to convince designers and managers to take these factors into account.

In contradiction to the premise that the night circadian decrement caused greater fatigue than the cumulative fatigue, one could argue that Schedule B's flight performance was better on FD 3 than on FD 2 simply because of a learning improvement, as occurred with Schedule A. This possibility cannot be refuted, given the present experimental design in which the subject's flying performance had not plateaued. The DIPT was supposed to have plateaued, and those scores were not significantly different from FD 2 to FD 3. The FD 3 scores were not significantly different from those of FD 1, even though FD 1 was significantly different from FD 2.

The primary reason for maintaining that the FD 2 fatigue induced the greater flying performance decrement is that the FD 3 scores appeared to be lower than those of FD 2--that is, lower than the improvement which would have occurred due to learning. However, this reason is speculative; and Hypothesis 10 will have to await further research before being fully confirmed. The possibility that Schedule B subjects had learned to cope more adequately with the extreme fatigue stressors on FD 3 cannot presently be dismissed. Additional support for Hypothesis 10 was provided by the HRV analysis, since HRV was greater on FD 2 in Period 2 than on FD 3 and would indicate less concentration ability on FD 2. According to some evidence, SSS scores indicated greater sleepiness at the end of FD 2 than of FD 3. HR and rectal temperature scores were in the predicted direction (lower) on FD 2 than FD 3, following a circadian pattern. However, this information is not evidence that the fatigue stressor was worse on FD 2 than on FD 3. Possibly the learning ability was reduced due to fatigue (Welford, Brown, and Gabb: 213); but, in terms of absolute improvement in flying performance, both groups appeared to have improved equally by FD 4. Therefore one can safely conclude that Schedule B's performance had recovered from the intense fatigue flights.

Implications of Findings

The most important remaining work is to determine the operational significance and relationship of flying performance decrements, observed in the laboratory, as applied to actual flying conditions. Attempts should be made to ascertain how large a drop in CTSE must occur in the laboratory before a pilot, with an equivalent loss in aircraft control or with similar limits for information processing decrement, could not be expected to complete a real-world mission safely. In addition, performance assessment technology must be developed to permit evaluation of performance decrement during actual flying operations. If further testing validates the DIPT, it would seem to be a prime candidate for such a device.

The continued use of Airmen trainees matched to pilot sample as subjects seems justified; however, because of differences in the skill levels of pilots and their greater experience in handling stress and fatigue, actual pilots should be tested in the GAT-1 performance measurement system periodically to insure that nonpilot decrements being obtained are reasonable estimates for the target populations.

Future research using the GAT-1 performance measurement system should also examine the effects of: shorter, less favorable crew-rest conditions, longer flight duration; increased numbers of flight days; higher inflight workloads, as generated by special mission requirements and more complex aircraft; and other stressors (such as hypoxia and drugs).

Of the four fatigue stressors in this [illegible]y, the daily flying duty duration and the total mission duration appeared to have had minimal impact on performance. Time awake prior to flying and circadian effects caused all of the significant performance decrement. A daytime flight of 9 hr over a 4-day period caused no apparent performance decrement. However, the 9 hr of flying time, starting 12 hr after the pilot awakes, creates a performance decrement which may result in dangerously poor flying performance.

Night flights seemed to create a more severe performance decrement even though the following flights had greater cumulative fatigue. Because of the ever-increasing capability for nighttime flying operations, pilots and aircrew schedulers should be especially concerned about overextending flying resources during disadvantageous circadian cycles. During the present study, this disadvantageous period appeared to be between the hours of 0200 and 0700 (Figs. 11 and 13). Additional research, aimed specifically at determining night-flight performance decrement over extended missions, is required.

Since the SAM Subjective Fatigue Checkcard and the Stanford Sleepiness Scale (SSS) measure the same underlying dimension, the SSS would appear to be the instrument of choice in field studies. The SSS can be answered in a shorter period of time than the SAM Checkcard, and for almost every statistical test yielded a more significant probability value. While the significance levels themselves cannot be directly compared, the SSS would appear to have a better chance of detecting any significant increase in sleepiness or fatigue. The fact that the subject can better develop a frame of reference for the SSS (since he can remember his previous sleepiness reports) may account for the better performance of this measure. Investigation into the development and validation of a 7-point fatigue-oriented scale for field use may be warranted. The large existing data base collected with the SAM Checkcard must also be considered in this decision. However, it should be possible to calibrate a new scale against the Checkcard to permit comparison of data across studies.

The temporal order of change in the three classes of measures was expected to be: first, subjective fatigue and sleepiness reports; second, physiological changes; and third, the performance decrement. On FD 1 significant changes occurred only in subjective fatigue and sleepiness, showing sensitivity to relatively mild, objective fatigue stressors; this time-of-day effect was significant even for baseline. Although a significant change occurred in HR and rectal temperature on days of mild fatigue, this effect was probably the normal circadian response and not a function of fatigue. Whether performance declined before physiological changes is difficult to determine, especially since these changes were highly correlated during intense fatigue.

For recovery, performance was expected to return first, and to be followed by reduction of subjective fatigue and return of physiological indices to pre-mission values. Essentially all Schedule B measures had returned to pre-mission values by FD 4, except for rectal temperature which was indicating overshoot of normal circadian values. It is difficult to determine if any of the other measures returned to baseline before another. Still, these data imply that physiological variables, especially body temperature, should be monitored to ascertain when a subject has fully recovered from the effects of exhausting work-rest cycles. Because the urine data (which could not be reported) were expected to be the best indicator of recovery, it was collected during both R 1 and R 2. Future studies may find urinalysis a useful measure of fatigue and recovery which can be related to performance decrement.

REVIEW AND CONCLUSIONS

The purpose of this study was to assist in the development of flight duration parameters for scheduling aircrew work-rest cycles. This aim was achieved by using objective measures of flying performance, subjective reports of fatigue and sleepiness, and three physiologic indicators of stress. Urine samples were also collected; but, because these data were invalidated during chemical analysis, none of the results were reported.

An additional objective was to demonstrate the relationship of these measures to each other and to a Discrete Information Processing Test (DIPT) which incorporated the special requirements of a performance assessment device usable in field research. This is a computer-controlled 5-choice reaction-time task which adapts the presentation rate of the stimuli to the subject's response accuracy and continuity. Through an iterative process is determined the rate at which the subject becomes unable to keep pace and stops responding for two presentations in a row. This rate is considered his threshold of information processing speed.

It was hypothesized that four fatigue stressors would cause significant changes in the following measures: (a) simple flight performance test (SLT); (b) complex flight performance test (FMT); (c) threshold of information processing speed (DIPT); (d) heart rate (HR); (e) heart rate variability (HRV); (f) rectal temperature; and (g) subjective reports of sleep duration, fatigue, and sleepiness. The four fatigue stressors were: (a) time awake prior to start of flight (1 vs. 12 hr); (b) daily flight duration (two 4.5-hr flights); (c) total mission duration (4 days); and (d) flight during a normal sleep period (night vs. day flight). Two tests of the arousal hypothesis were made to attempt to: (a) separate tasks based on their inherent arousal value; and (b) relate task arousal values to fatigue-induced decrements. The circadian rhythm hypothesis (that performance rises and falls with time of day) was tested to establish whether a night flight or the following day flight (which included greater cumulative fatigue) caused either: (a) greater performance decline; (b) greater subjective estimates of fatigue and sleepiness; or (c) greater changes in HR, HRV, or rectal temperature. Finally, an attempt was made to determine if the dependent measures were correlated with each other.

To represent the USAF pilot population, 24 Airmen were selected on the basis of flight aptitude scores, class II flight physicals, and personal interviews. Each received an intensive, 7-day flight training program in Link GAT-1 Trainers. Then, 12 subjects were randomly assigned a mission which regularly alternated 12-hr duty days with 12-hr crew-rest periods for 4 days; the remaining subjects had duty days of 12, 24, 24, 12 hr, all duty days being separated by 12-hr crew-rest periods. During each duty day, all subjects flew two 4.5-hr flights, separated by a 1-hr rest. Flying performance was evaluated by two mulitidimensional time-on-target tracking scores derived from heading, altitude, airspeed, turn rate, turn coordination, and vertical velocity errors. Each flight hour, a PDP-12 computer administered and scored both tests of flight performance (the SLT and the FMT) and the DIPT. Continous HR and rectal temperature, subjective fatigue and sleepiness reports, and sleep logs were collected throughout each mission. HRV was computed from the HR data collected.

From the foregoing analyses, the following 16 conclusions are warranted in relation to the work-rest schedules in this study:

1. Flying performance, as measured by both the SLT and FMT, was significantly degraded during two 4.5-hr flights when subjects had been awake 12 hr prior to the flight.

2. Subjects could perform at least 4 consecutive days of two 4.5-hr flights, involving 12-hr duty days, without showing any significant flying performance or threshold of information processing degradation, provided that 12 hr of crew rest was obtained between duty days.

3. A subject's threshold of information processing speed was significantly increased during two 4.5-hr night flights that started 12 hr after he awoke.

4. One 12-hr crew-rest period appeared adequate for recovery from fatigue generated by two consecutive 24-hr flying duty days, each having two 4.5-hr of flight.

5. A subject's threshold of information processing speed was correlated with his flying performance ability during periods of intense fatigue (24-hr duty days). Thus the DIPT, which was designed to meet the requirements for a device to assess performance in the flight environment, was shown to be a valid index of fatigue-induced flight performance decrement.

6. Subjective fatigue as measured by the USAFSAM Subjective Fatigue Checkcard was highly correlated with reports of sleepiness as measured by the Stanford Sleepiness Scale (SSS). Thus both measures appeared to tap the same underlying dimensions. According to some indications, however, the SSS was more likely to demonstrate significant changes than the SAM Checkcard. The greater sensitivity of the SSS was possibly due to its 7-point-scale response format.

7. Both subjective fatigue and sleepiness measures were highly correlated with flying performance and threshold of information processing speed during periods of intense fatigue generated by disrupted circadian rhythms and a 24-hr duty day.

8. Subjects attempted to obtain significantly more sleep only immediately following the 24-hr duty days.

9. In the absense of intense fatigue, HR tended to fall as subjects became accustomed to the flying task, both within flights and across flight days. During intense fatigue, however, HR rose throughout the day and fell at night.

10. HRV increased not only as subjects became accustomed to the flying task, but also during periods of intense fatigue.

11. Rectal temperature showed a normal circadian rise from morning to evening during 12-hr duty days involving two 4.5-hr flights. Rectal temperature dropped sharply during night flights, and was then significantly correlated with decrements in flying performance and reduction of information processing speed. Periods of intense fatigue generated by disrupted circadian rhythms and 24-hr duty days caused exaggerated swings in the circadian rhythm temperature cycle.

12. During night flights which started 12-hr after waking, decreasing rectal temperature was significantly correlated with postflight subjective fatigue and sleepiness scores.

13. HR significantly increased for the more complex flight task (FMT), as compared with the less complex flight task (SLT), only for subjects on the more fatiguing schedule. This is an indication that complex tasks increase arousal.

14. HRV was a better indicator of task arousal and complexity than HR, and could be used to rank-order the SLT, FMT, and DIPT in both schedules.

15. Even though the absolute scores were better on the SLT than the FMT, the less arousing task (SLT) declined more from baseline (than the FMT) during periods of intense fatigue. Thus, more complex flight tasks are more resistant to decline during intense fatigue periods. During actual flight, however, the performance of pilots experiencing the maximum fatigue levels generated in this study would probably have degraded to an unacceptable and unsafe level on either a simple or complex flight task.

16. According to some indications, night flights which started 12 hr after subjects awoke caused more severe flight performance decline than day flights which started 12 hr after subjects awoke, even when cumulative fatigue effects from the preceding 24-hr duty day were included. Higher HRV during the night flight, relative to the following day flight, indicated a loss of concentration and tended to support this conclusion. Sleepiness scores also appeared to be higher at the end of the night flight than the day flight. However, due to probable continued improvement of both flight tasks across all flight days, the possibility exists that the better performance on the following day flight was a learning effect.

One must remember that these conclusions were drawn from simulated-flight data, using inexperienced subjects as pilots. The amount of performance decrement in actual flying operations following similar schedules will probably vary with type of aircraft, mission, and pilot experience levels. Nevertheless, the rate of performance decrement in actual flight should approximate that in laboratory-controlled, simulated-flight conditions. The foregoing generalizations to flight operations are thus warranted.

REFERENCES

1. AF Regulation 60-1, *Flight management*. Washington, D.C.: Department of the Air Force, Headquarters U.S. Air Force, January, 1975.

2. AF Regulation 80-33, *Use of volunteers in aerospace research*. Washington, D.C.: Department of the Air Force, Headquarters U.S. Air Force, August, 1969.

3. AF Regulation 160-43, *Medical examinations and medical standards*. Washington, D.C.: Department of the Air Force, Headquarters U.S. Air Force, June, 1976.

4. Ainsworth, L. L., and H. P. Bishop. *The effects of a 48-hour period of sustained field activity on tank crew performance* (Humrro Tech. Rep. No. 71-16). Alexandria, Va: Human Resources Research Organization, July, 1971. (NTIS No. AD-731 219)

5. Alluisi, E. A. Sustained performance. In E. A. Bilodeau and I. Mc D. Bilodeau (Eds.), *Principles of skill acquisition*. New York: Academic Press, 1969.

6. Aschoff, J., M. Fatranska, U. Gerecke, and H. Giedke. Twenty-four hour rhythms of rectal temperature in humans: effects of sleep-interruptions and of test-sessions. *Pflügers Archives*, 1974, *346*, 215-222.

7. Aschoff, J., H. Giedke, E. Popple, and R. Wever. The influence of sleep-interruption and of sleep-deprivation on circadian rhythms in human performance. In W. P. Colquhoun (Ed.), *Aspects of human efficiency*. London: The English Universities Press Ltd, 1972.

8. Ax, A. F., and E. D. Luby. Autonomic responses to sleep deprivation. *Archives of General Psychiatry*, 1961, *4*, 55-59.

9. Banks, J. H., J. J. Sternberg, J. P. Farrell, C. H. DeBow, and W. A. Delhamer. *Effects of continuous military operations on selected military tasks* (BESRL TRR No. TR-1166). Arlington, Va: U.S. Army Behavioral and Systems Research Laboratory, December, 1970. (NTIS No. AD-178 253)

10. Barlay, S. *The search for air safety*. New York: William Morrow & Co., 1970.

11. Barmack, J. E. Boredom and other factors in the physiology of mental effort: an exploratory study. *Archives of Psychology*, 1973, No. 218, 2-83.

12. Barr, A. J., and J. H. Goodnight. A users guide to the statistical analysis system. Raleigh, N.C.: North Carolina State University, August, 1972.

13. Bartlett, F. Fatigue in the air pilot (Report No. 448). Cambridge: Flying Personnel Research Committee, Air Ministry, August, 1942.

14. Bartlett, F. Psychological criteria of fatigue. In W. F. Floyd and A. T. Welford (Eds.), Symposium on fatigue. London: H. K. Lewis & Co., 1953.

15. Bartley, S. H., and E. Chute. Fatigue and impairment in man. New York: McGraw-Hill, 1947.

16. Bateman, S. C., R. Goldsmith, K. F. Jackson, H. P. Ruffell Smith, and V. S. Mattocks. Heart rate of training captains engaged in different activities. Aerospace Medicine, 1970, 41, 425-429.

17. Bergström, B. The effect of sleep loss and threat induced stress upon tracking. Scandinavian Journal of Psychology, 1972, 13, 54-60.

18. Bergum, B. O. A taxonomic analysis of continuous performance. Perceptual and Motor Skills, 1966, 23, 47-54.

19. Berkun, M. M., H. M. Bialek, R. P. Kern, and K. Yagi. Experimental studies of psychological stress in man. Psychological Monographs, 1962, 76 (15, Whole No. 534).

20. Bertelson, P., and R. Joffe. Blocking in prolonged serial responding. Ergonomics, 1963, 6, 109-116.

21. Billings, C. E., R. J. Gerke, and R. L. Wick, Jr. Comparisons of pilot performance in simulated and actual flight. Aviation, Space, and Environmental Medicine, 1975, 46, 304-308.

22. Bills, A. G. Blocking: a new principle in mental fatigue. American Journal of Psychology, 1931, 43, 230-245.

23. Bills, A. G. Fatigue, oscillations, and blocks. Journal of Experimental Psychology, 1935, 18, 562-573.

24. Bills, A. G. General experimental psychology. New York: Longmans, Green and Co., 1934.

25. Bills, A. G. Mental fatigue and anoxemia compared. Journal of Experimental Psychology, 1937, 20, 437-452.

26. Bills, A. G. The psychology of efficiency. New York: Harper & Bros., 1943.

27. Bills, A. G. Some causal factors in mental blocking. Journal of Experimental Psychology, 1935, 18, 172-185.

28. Bills, A. G., and M. J. Shapin. Mental fatigue under automatically controlled rates of work. Journal of General Psychology, 1936, 15, 335-347.

29. Bitterman, M. E. Fatigue defined as reduced efficiency. American Journal of Psychology, 1944, 57, 569-573.

30. Bjerner, B. Alpha depression and lowered pulse rate during delayed actions in a serial reaction test: a study in sleep deprivation. Acta Physiologica Scandinavica, 1949, 19 (Supplementum 65), 1-93.

31. Blake, M. J. F. Temperment and time of day. In W. P. Colquhoun (Ed.), Biological rhythms and human performance. New York: Academic Press. 1971.

32. Bliss, E. L., C. J. Migeon, C. H. Hardin-Branch, and L. T. Samuels. Reaction of adrenal cortex to emotional stress. Psychosomatic Medicine, 1956, 18, 56-76.

33. Blix, A. S., S. B. Stromme, and H. Ursin. Additional heart rate - an indicator of psychological activation. Aerospace Medicine, 1974, 45, 1219-1222.

34. Boyce, P. R. Sinus arrhythmia as a measure of mental load. Ergonomics, 1974, 17, 177-183.

35. Broadbent, D. E. Noise, paced performance, and vigilance tasks. British Journal of Psychology, 1953, 44, 295-303.

36. Broadbent, D. E. Perception and communication. London: Pergamon Press, 1958.

37. Broverman, D. M., E. L. Klaiber, W. Vogel, and Y. Kobayashi. Short-term vs. long-term effects of adrenal hormones on behaviors. Psychological Bulletin, 1974, 81, 672-694.

38. Brown, I. D. The measurement of perceptual load and reserve capacity. Transactions of the Association of Industrial Medical Officers, 1964, 14, 44-49.

39. Buck, L. Circadian rhythms in step-input pursuit tracking. Ergonomics, 1977, 20, 19-31.

40. Buck, L. Psychomotor test performance and sleep patterns of aircrew flying transmeridional routes. Aviation, Space, and Environmental Medicine, 1976, 47, 979-986.

41. Buck, L. Sleep loss effects on movement time. Ergonomics, 1975, 18, 415-425.

42. Cameron, C. A theory of fatigue. Ergonomics, 1973, 16, 633-648.

43. Cannon, W. B., and D. de la Paz. Emotional stimulation of adrenal secretion. *American Journal of Physiology*, 1911, *27*, 64-70.

44. Cantrell, G. K., R. W. Trimble, and B. O. Hartman. *Long-term aircrew effectiveness (A literature survey)* (SAM-TR-71-4). Brooks AFB, Tx: USAF School of Aerospace Medicine, Aerospace Medical Division, April, 1971. (NTIS No. AD-722 417)

45. Carmichael, L., and J. L. Kennedy. Some recent approaches to the experimental study of human fatigue. *Science*, 1949, *110*, 445.

46. Chambers, E. A. Industrial fatigue. *Occupational Psychology*, 1961, *35*, 44-57.

47. Chiles, W. D. *Experimental studies of prolonged wakefulness* (WADC TR-55-395). Wright-Patterson AFB, Oh: Wright Air Development Center, Air Research and Development Command, December, 1955.

48. Chiles, W. D., E. A. Alluisi, and O. S. Adams. Work schedules and performance during confinement. *Human Factors*, 1968, *10*, 143-195.

49. Collins, W. E. *Some effects of sleep deprivation on tracking performance in static and dynamic environments* (FAA-AM-76-12). Oklahoma City, Oklahoma: Federal Aviation Administration Office of Aviation Medicine, Civil Aeromedical Institute, October, 1976. (NTIS No. AD-A033331/0)

50. Colquhoun, W. P. (Ed). *Biological rhythms and human performance*. New York: Academic Press, 1971.

51. Conroy, R. T. W. L., and J. N. Mills. *Human circadian rhythms*. London: J & A Churchill, 1970.

52. Corcoran, D. W. J. Changes in heart rate and performance as a result of loss of sleep. *British Journal of Psychology*, 1964, *55*, 307-314.

53. Craik, K. J. W. Theory of the human operator in control system II. Man as an element in a control system. *British Journal of Psychology*, 1948, *38*, 142-148.

54. Dahl, H., and D. P. Spence. Mean heart rate predicted by task demand characteristics. *Psychophysiology*, 1971, *7*, 369-376.

55. Davis, D. R. *Pilot error, some laboratory experiments*. London: His Majesty's Stationery Office, 1948.

56. Davy, J. On the temperature of man. *Philosophical Transactions*, 1845, *2*, 319-333.

57. Dodge, R. The laws of relative fatigue. *The Psychological Review*, 1917, *24*, 89-113.

58. Donnell, J. M. Performance decrement of total sleep loss and task duration. Perceptual Motor Skills, 1969, 29, 711-714.

59. Drew, G. C. An experimental study of mental fatigue (Paper 227). Cambridge: Flying Research Personnel Committee, Air Ministry, December, 1940.

60. Drucker, E. H., L. D. Cannon, and J. R. Ware. The effects of sleep deprivation on performance over a 48-hour period (TR 69-8). Washington, D.C.: Human Resources Research Office, George Washington University, May, 1969. (NTIS No. AD-688-950)

61. Duffy, E. Activation and behavior. New York: John Wiley & Sons, 1962.

62. Dukes-Dobos, F. N. Fatigue from the point of urinary metabolites. In K. Hashimoto, K. Kogi, & E. Grandjean (Eds.), Methodology in human fatigue assessment. London: Taylor & Francis Ltd., 1971.

63. Eason, R. C., and L. M. Dudley. Physiological and behavioral indicants of activation. Psychophysiology, 1971, 7, 223-232.

64. Edwards, A. S. Effects of the loss of one hundred hours of sleep. American Journal of Psychology, 1941, 54, 80-91.

65. Ellis, N. C., A. L. Lowes, W. G. Matheny, and D. A Norman. The feasibility of using an adaptive technique in flight simulator training. Ergonomics, 1971, 14, 381-389.

66. Entwistle, W. H. Oscillation. British Journal of Psychology, 1937, 27, 313-328.

67. Ettema, J. H., and R. L. Zielhuis. Physiological parameters of mental load. Ergonomics, 1971, 14, 137-144.

68. Euler, U. S. v. Quantitation of stress by catecholamine analysis. Clinical Pharmacology and Therapeutics, 1964, 5, 398-404.

69. Euler, U. S. v., and F. Lishajko. The estimation of catecholamines in urine. Acta Physiologica Scandinavia, 1959, 45, 22-132.

70. Euler, U. S. v., and U. Lundberg. Effect of flying on the epinephrine excretion in Air Force personnel. Journal of Applied Physiology, 1953, 6, 551-555.

71. Fenz, W. D., and J. G. Graig. Autonomic arousal and performance during 60 hours of sleep deprivation. Perceptual Motor Skills, 1972, 34, 543-553.

72. Fine, P. M., and B. O. Hartman. Psychiatric strengths and weaknesses of typical Air Force pilots (SAM-TR-68-121). Brooks AFB, Tx: USAF School of Aerospace Medicine, Aerospace Medical Division, November, 1968. (NTIS No. 681495)

73. Floyd, W. F., and A. T. Welford. Symposium on fatigue. London: H. K. Lewis & Co. Ltd., 1953.

74. Flügel, J. C. Practice, fatigue and oscillation: a study of work at high pressure. British Journal of Psychology, Monograph Supplement 13, 1928.

75. Foley, P. J., and M. Humphries. Blocking in serial simple reaction tasks. Canadian Journal of Psychology, 1962, 16, 128-237.

76. Frankenhaeuser, M., I. Mellis, A. Rissler, C. Björkvall, and P. Pátkai. Catecholamine excretion as related to cognitive and emotional reaction patterns. Psychosomatic Medicine, 1968, 30, 109-120.

77. Freeman, G. L., and C. I. Hovland. Diurnal variations in performance and related physiological processes. Psychological Bulletin, 1934, 31, 777-799.

78. Fröberg, J., C.-G. Karlson, L. Levi, and L. Lidberg. Circadian variations in performance, psychological ratings, catecholamine excretion, and diuresis during prolonged sleep deprivation. International Journal of Psychobiology, 1972, 2, 23-36.

79. Gaillard, A. W. K., and D. A. Trumbo. Drug effects on heart rate and heart rate variability during a prolonged reaction task. Ergonomics, 1976, 19, 611-622.

80. Gartner, W. B., and M. R. Murphy. Pilot workload and fatigue: a critical survey of concepts and assessment techniques (TN-D-8365). Moffett Field, Ca: NASA Ames Research Center, November, 1976. (NTIS No. A-6630)

81. Gold, R. E., and L. L. Kulak. Effect of hypoxia on aircraft pilot performance. Aerospace Medicine, 1972, 43, 180-183.

82. Grandjean, E. P., G. Wotzka, R. Schaad, and A. Gilgen. Fatigue and stress in air traffic controllers. Ergonomics, 1971, 14, 159-165.

83. Guyton, A. C. Textbook of medical physiology (4th Ed.). Philadelphia: W. B. Saunders, 1971.

84. Hale, H. B., J. P. Ellis, Jr., and E. W. Williams. Urinary creatine-based ratios in relation to season. Aerospace Medicine, 1968, 39, 1048-1051.

85. Hale, H. B., B. O. Hartman, D. A. Harris, R. E. Miranda, and E. W. Williams. Physiologic cost of prolonged double crew flights in C-5 aircraft. Aerospace Medicine, 1973, 44, 99-1008.

86. Harris, W., and J. F. O'Hanlon. A study of recovery functions in man (U.S. Army Tech Memorandum 10-72). Aberdeen Proving Ground, Md: Aberdeen Research and Development Center, Human Engineering Laboratory, April, 1972. (NTIS No. AD-741 828)

87. Harris, D. A., H. B. Hale, B. O. Hartman, and J. A. Martinez. Oral temperature in relation to inflight work/rest schedules. Aerospace Medicine, 1970, 41, 723-727.

88. Hartman, B. O. Fatigue effects in 24-hour simulated transport flight (SAM-TR-65-16). Brooks AFB, Tx: USAF School of Aerospace Medicine, Aerospace Medical Division, April, 1965. (NTIS No. AD-456 174)

89. Hartman, B. O. Psychological factors in flying fatigue. In C. Perry (Ed.), Psychiatry in Aerospace Medicine (Vol 4.). Boston: Little, Brown, & Co., 1967.

90. Hartman, B. O. Time and load factors in astronaut proficiency. In B. E. Flaherty (Ed.), Aspects of space flight. New York: Columbia University Press, 1961.

91. Hartman, B. O., and G. K. Cantrell. MOL: Crew performance on demanding work/rest schedules compounded by sleep deprivation (SAM-TR-67-99). Brooks AFB, Tx: USAF School of Aerospace Medicine, Aerospace Medical Division, November, 1967. (NTIS No. AD-665845)

92. Hartman, B. O., and G. K. Cantrell. Psychomotor monitoring for general efficiency. In Medical education for national defense symposium on biomedical monitoring. Brooks AFB, Tx: USAF School of Aerospace Medicine, Aerospace Medical Division, November, 1964. (NTIS No. AD-467 036)

93. Hartman, B. O., and P. M. Fitts. The development of techniques and procedures for the study of alertness in aviation personnel. Washington, D.C.: National Research Council Committee on Aviation Psychology, Civil Aeronautic Administration, October, 1950.

94. Hartman, B. O., D. E. Flinn, A. B. Edmunds, F. D. Brown, and J. E. Shubert. Human factors aspects of a 30-day extended survivability test of the Minuteman missile (SAM-TR-64-62). Brooks AFB, Tx: USAF School of Aerospace Medicine, Aerospace Medical Division, October, 1964. (NTIS No. A-456174)

95. Hartman, B. O., W. F. Storm, J. E. Vanderveen, H. B. Hale, and R. R. Bollinger. Operational aspects of variations in alertness (AGARDograph No. 189). Neuilly Sur Seine, France: Advisory Group for Aerospace Research & Development, North Atlantic Treaty Organization, August, 1974. (NTIS No. AD-786 123/OGI)

96. Hasbrook, A. H., and P. G. Rasmussen. Pilot heart rate during in-flight simulated instrument approaches in a general aviation aircraft. Aerospace Medicine, 1970, 41, 1148-1152.

97. Hauty, G. T., and T. Adams. Phase shifts of the human circadian system and performance deficit during the periods of transition: I East-West flight, II West-East flight, III North-South flight. Aerospace Medicine, 1966, 37, 668-674, 1027-1033, 1257-1262.

98. Hebb, D. O. Drives and the C.N.S. (Conceptual Nervous System). Psychological Review, 1955, 62, 243-254.

99. Henry, P. H., T. Q. Davis, E. J. Engelken, J. H. Triebwasser, and M. C. Lancaster. Alcohol-induced performance decrement assessed by two Link trainer tasks using experienced pilots. Aerospace Medicine, 1974, 45, 1180-1189.

100. Henry, P. H., J. A. Flueck, J. F. Sanford, H. N. Keiser, R. C. McNee, W. H. Walter, III, K. H. Webster, B. O. Hartman, and M. C. Lancaster. Assessment of performance in a Link GAT-1 flight simulator at three alcohol dose levels. Aerospace Medicine, 1974, 45, 33-44.

101. Hick, W. E. The discontinuous functioning of the human operator in pursuit tasks. Quarterly Journal of Experimental Psychology, 1948, 1, 36-51.

102. Hockey, G. R. J., and W. P. Colquhoun. Diurnal variation in human performance: a review. In W. P. Colquhoun (Ed.), Aspects of human efficiency. London: English Universities Press Ltd, 1972.

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

104. Hodge, D. C. Military requirements for research on continuous operations (U.S. Army Tech. Memorandum 1272). Aberdeen Proving Ground, Md: Human Engineering Laboratory, Aberdeen Research & Development Center. (NTIS No. AD-744782)

105. Hudson, E. M. Adaptive techniques in multiparameter problems. Human Factors, 1969, 11, 561-567.

106. Iampietro, P. F., C. E. Melton, Jr., E. A. Higgins, J. A. Vaughn, S. M. Hoffman, G. E. Funkhouser, and J. T. Saldivar. High temperature and performance in a flight task simulator. Aerospace Medicine, 1972, 43, 1215-1218.

107. Jackson, K. F. Time relationships in pilot performance. Occupational Psychology, 1959, 33, 80-97.

108. Johnson, L. C. Psychological and physiological changes following total sleep deprivation. In A. Kales (Ed.), Sleep: physiology and pathology. Philadelphia: J. B. Lippincott, 1969.

109. Johnson, L. C., and P. Naitoh. The operational consequences of sleep deprivation and sleep deficit (AGARDograph No. 193). Neuilly Sur Seine, France: Advisory Group for Aerospace Research & Development, North Atlantic Treaty Organization, June, 1974. (NTIS No. AD-783 199/3GI)

110. Kahneman, D. Attention & effort. Englewood Cliffs, N. J.: Prentice-Hall, 1973.

111. Kahneman, D., B. Tursky, D. Shapiro, and A. Crider. Pupillary, heart rate, and skin resistance changes during a mental task. Journal of Experimental Psychology, 1969, 79, 164-167.

112. Kalsbeek, J. W. H. Sinus arrhythmia and the dual task method in measuring mental load. In W. T. Singleton, J. G. Fox, and D. Whitfield (Eds.), Measurement of man at work. London: Taylor & Francis, 1971.

113. Kappauf, W. E. Use of an on-line computer for psychophysical testing with the up-and-down method. American Psychologist, 1969, 24, 207-211.

114. Kelley, C. R. The measurement of tracking proficiency. Human Factors 1969, 11, 43-64. (a)

115. Kelley, C. R. What is adaptive training? Human Factors, 1969, 11, 547-556. (b)

116. Kelley, C. R., and M. J. Wargo. Cross-adaptive operator loading tasks. Human Factors, 1967, 9, 395-404.

117. Kennedy, J. L. Some practical problems of the alertness indicator. In W. F. Floyd and A. T. Welford (Eds.), Symposium on fatigue. London: H. K. Lewis & Co., 1953.

118. Klein, K. E., H. Bruner, H. Holtman, H. Rehme, J. Stolze, W. D. Steinhoff, and H. M. Wegmann. Circadian rhythm of pilots' efficiency and effects of multiple time-zone travel. Aerospace Medicine, 1970, 41, 125-132.

119. Klein, K. E., H. M. Wegmann, and B. I. Hunt. Desynchronization of body temperature and performance circadian rhythms as a result of outgoing and homegoing transmeridian flights. Aerospace Medicine, 1972, 43, 119-132.

120. Klein, K. E., H. M. Wegmann, G. Athanassenas, H. Hohlweck, and P. Kuklinski. Air operations and circadian performance rhythms. Aviation, Space, and Environmental Medicine, 1976, 47, 221-230.

121. Kleitman, N. Sleep and wakefulness (2nd Ed.). Chicago: University of Chicago Press, 1963.

122. Kleitman, N. Studies on the physiology of sleep: I. The effects of prolonged sleeplessness in man. American Journal of Physiology, 1923, 66, 67-92.

123. Kleitman, N., and D. P. Jackson. Body temperature and performance under different routines. Applied Physiology, 1950, 3, 309-328.

124. Kleitman, N., S. Titelbaum, and P. Feiveson. The effect of body temperature on reaction time. American Journal of Physiology, 1938, 121, 495-501.

125. Kollar, E. J., G. R. Slater, J. O. Palmer, R. F. Docter, and A. T. Mandell. Stress in subjects undergoing sleep deprivation. Psychosomatic Medicine, 1966, 28, 101-113.

126. Krahenbuhl, G. S., J. R. Marett, and N. W. King. Catecholamine excretion in T-37 flight training. Aviation, Space, and Environmental Medicine, 1977, 48, 405-408.

127. Kreider, M. B. Effects of sleep deprivation on body temperatures. Federation Proceedings, 1961, 20, 214. (Abstract)

128. Lacey, B. C., and J. I. Lacey. Studies of heart rate and other bodily processes in sensorimotor behavior. In P. A. Obrist, A. H. Black, J. Brener, and L. V. Di Cara (Eds.), Cardiovascular psychophysiology. Chicago: Aldine Publishing Co., 1974.

129. Lacey, J. I. Somatic response patterning and stress: some revisions of activation theory. In M. H. Appley and R. Trumbull (Eds.), Psychological stress. New York: Appleton-Century-Crofts, 1967.

130. Lacey, J. I., J. Kagan, B. C. Lacey, and H. A. Moss. The visceral level: situational determinants and behavioral correlates of autonomic response patterns. In P. H. Knapp (Ed.), Expressions of emotion in man. New York: International Universities Press, Inc., 1963.

131. Laslett, H. R. Experiments on the effects of the loss of sleep. Journal of Experimental Psychology, 1928, 11, 370-396.

132. Lee, M. A. M., and N. Kleitman. Studies on the physiology of sleep: attempts to demonstrate functional changes in the nervous system during experimental insomnia. American Journal of Physiology, 1923, 67, 141-152.

133. Leonard, J. A. Five-choice serial reaction apparatus (Report No. 326-59). Cambridge: Applied Psychology Research Unit, Medical Research Council, 1959.

134. Lindsley, D. B. Psychophysiology and motivation. In M. R. Jones (Ed.), Nebraska symposium on motivation. Lincoln: University of Nebraska Press, 1957.

135. Lisper, H.-O., and A. Kjellberg. Effects of 24-hour sleep deprivation on rate of decrement in a 10-minute auditory reaction time task. Journal of Expermental Psychology, 1972, 96, 287-290.

136. Lubin, A. Performance under sleep loss and fatigue. In S. S. Kety, E. V. Evarts, and H. L. Williams (Eds.), Sleep and altered states of consciousness. Baltimore: Williams & Wilkins, 1967.

137. Mackworth, N. H. Researches on the measurement of human performance. London: His Majesty's Stationery Office, 1950.

138. Malmo, R. B. Activation: a neurophysiological dimension. Psychological Review, 1959, 66, 367-386.

139. Malmo, R. B., and W. W. Surwillo. Sleep deprivation: changes in performance and physiological indicants of activation. Psychological Monographs, 1960, 74 (15, Whole No. 502).

140. Marchbanks, V. H., Jr. Flying stress and urinary 17-OHCS levels during twenty-hour missions. Aerospace Medicine, 1960, 31, 639-643.

141. McIntosh, B. B., J. L. Milton, and E. L. Cole. Pilot performance during extended periods of instrument flight (AFTR 6725). Wright-Patterson Air Force Base, Oh: Aero Medical Laboratory, Air Research & Development Command, May 1952. (NTIS No. AD-168 862)

142. Melton, C. E., and V. Fiorica. Physiological responses of low-time private pilots to cross-country flying (FAA-AM-71-23). Oklahoma City, Ok: FAA Civil Aeromedical Institute, April, 1971. (NTIS No. AD-728-317)

143. Migeon, C. J., F. H. Tyler, J. P. Mahoney, A. A. Florentin, H. Castle, E. L. Bliss, and L. T. Samuels. The diurnal variation of plasma levels and urinary excretion of 17-OHCS in normal subjects, night workers, and blind subjects. Journal of Clinical Endocrinology, 1956, 16, 622-633.

144. Miller, R. G. Secretion of 17-hydroxycorticosteroids (17-OHCS) in military aviators as an index of response to stress: a review. Aerospace Medicine, 1968, 39, 498-501.

145. Mohler, S. R. Fatigue in aviation activities. Aerospace Medicine, 1966, 37, 722-732.

146. Morgan, B. B., B. R. Brown, and E. A. Alluisi. Effects on sustained performance of 48 hours of continuous work and sleep loss. Human Factors, 1974, 16, 406-414.

147. Morris, G. O., H. L. Williams, and A. Lubin. Misperception and disorientation during sleep deprivation. AMA Archives of General Psychiatry, 1960, 2, 247-254.

148. Mosso, A. Fatigue. London: Allen & Unwin, 1915.

149. Murray, E. J. Sleep deprivation and personality adjustment. In L. E. Abt and B. F. Riess (Eds.), Progress in clinical psychology (Vol. 8). New York: Grune & Stratton, 1968.

150. Murray, E. J., H. L. Williams, and A. Lubin. Body temperature and psychological ratings during sleep deprivation. Journal of Experimental Psychology, 1958, 56, 271-273.

151. Muscio, B. Is a fatigue test possible? British Journal of Psychology, 1921, 12, 31-46.

152. Naitoh, P. Sleep loss and its effects on performance (Report No. 68-3). San Diego, Ca: Navy Medical Neuropsychiatric Research Unit, August, 1969. (NTIS No. AD-699505)

153. Naitoh, P., and R. E. Townsend. The role of sleep deprivation research in human factors. Human Factors, 1970, 12, 575-585.

154. North, R. A., and D. Gopher. Measures of attention as predictors of flight performance. Human Factors, 1976, 18, 1-14.

155. Obrist, P. A., A. H. Black, J. Brenner, and L. V. Di Cara (Eds.), Cardiovascular physiology. Chicago: Aldine, 1974.

156. Obrist, P. A., S. I. Hallman, and D. M. Wood. Autonomic levels and lability, and performance time on a perceptual task and a sensory-motor task. Perceptual and Motor Skills, 1964, 18 753-762.

157. Opmeer, C. H. J. M., and J. P. Krol. Towards an objective assessment of cockpit workload: 1. physiological variables during different flight phases. Aerospace Medicine, 1973, 44 (5), 527-532.

158. Orr, D. B. Research of behavioral impairment due to stress: an experiment in long-term performance. Journal of Experimental Psychology, 1964, 68, 94-102.

159. Orr, W. C., and H. J. Hoffman. A 90-minute cardiac biorhythm: methodology and data analysis using modified periodograms and complex demodulation. IEEE Transactions on Biomedical Engineering, 1974, BME-21 (2), 130-143.

160. Orr, W. C., H. J. Hoffman, and F. W. Hegge. Ultradian rhythms in extended performance. Aerospace Medicine, 1974, 45, 995-1000.

161. Page, R. M. Measuring human energy cost in industry: a general guide to the literature. Genetic Psychological Monographs, 1932, 11, 321-537.

162. Page, R. M. On supplanting the industrial fatigue concept. Journal of Business, 1929, 2, 137-150.

163. Pátkai, P. Catecholamine excretion in pleasant and unpleasant situations. Acta Psychologica, 1971, 35, 352-363.

164. Patrick, G. T. W., and J. A. Gilbert. On the effects of loss of sleep. The Psychological Review, 1896, 3, 469-483.

165. Pearson, R. G. Task proficiency and feelings of fatigue (SAM-TR-57-77). Randolph AFB, Tx: School of Aviation Medicine, USAF, April, 1957.

166. Pearson, R. G., and G. E. Byars. The development and validation of a checklist for measuring subjective fatigue (SAM-TR-56-115). Randolph AFB, Tx: School of Aviation Medicine, USAF, December, 1956.

167. Pepler, R. D. Warmth and lack of sleep: accuracy or activity reduced? Journal of Comparative and Physiological Psychology, 1959, 52, 446-450.

168. Philpott, S. J. F. A theoretical curve of fluctuations of attention. British Journal of Psychology, 1934, 25, 221-255.

169. Pierson, W. R. Fatigue, work decrement, and endurance in a simple repetitive task. British Journal of Medical Psychology, 1963, 36, 279-282.

170. Pincus, G., and H. Hoagland. Steroid excretion in the stress of flying. Journal of Aviation Medicine, 1943, 14, 173-193.

171. Poffenberger, A. T. The effects of continuous mental work upon output and feelings. Journal of Applied Psychology, 1928, 12, 459-467.

172. Polzella, D. J. Effects of sleep deprivation on short-term recognition memory. Journal of Experimental Psychology: Human Learning and Memory, 1975, 104 (2), 194-200.

173. Porges, S. W. Heart rate variability and deceleration as indexes of reaction time. Journal of Experimental Psychology, 1972, 92, 103-110.

174. Poulton, E. C. Range effects in experiments on people. American Journal of Psychology, 1975, 88, 3-32.

175. Preston, F. S., and S. Bateman. Effect of time zone change on the sleep pattern of BOAC B 707 crews on world-wide schedules. Aerospace Medicine, 1970, 41, 1409-1415.

176. Robinson, E. S., and S. O. Herrmann. Effects of loss of sleep. Journal of Experimental Psychology, 1922, 5, 19-32.

177. Roman, J., J. J. Perry, L. R. Carpenter, and S. A. Awni. Flight research program: VI. heart rate and landing error in restricted field of view landings. Aerospace Medicine, 1967, 38, 128-132.

178. Roscoe, A. H. Use of pilot heart rate measurement in flight evaluation. Clinical Medicine, 1976, 47 (1), 86-90.

179. Rothballer, A. B. The effects of catecholamines on the central nervous system. Pharmacological Review, 1959, 11, 494-547.

180. Routtenberg, A. The two-arousal hypothesis: reticular formation and limbic system. Psychological Review, 1968, 75, 51-76.

181. Rubin, R. T., E. J. Kollar, G. Slater, and B. R. Clark. Excretion of 17-hydroxycorticosteroids and vanillylmandecic acid during 205 hours of sleep deprivation in man. Psychosomatic Medicine, 1969, 31, 68-79.

182. Ruffell Smith, H. P. Heart-rate of pilots flying scheduled airline routes. Aerospace Medicine, 1976, 38, 1117-1119.

183. Rutenfranz, J., J. Aschoff, and H. Mann. The effects of a cumulative sleep deficit, duration of preceding sleep period and body-temperature on multiple choice reaction time. In W. P. Colquhoun (Ed.), Aspects of human efficiency. London: English Universities Press Ltd., 1972.

184. Ryan, T. A. Varieties of fatigue. American Journal of Psychology, 1944, 57, 565-569.

185. Ryan, T. A. Work and effort. New York: Ronald Press, 1947.

186. Sampson, H. Pacing and performance on a serial addition task. Canadian Journal of Psychology, 1956, 10, 219-225.

187. Schreuder, O. B. Medical aspects of aircraft pilot fatigue with special reference to commercial jet pilot. Aerospace Medicine, Special Report Pub. 37 (4), 1966, 1-44.

188. Selye, H. The physiology and pathology of exposure to stress. Montreal: Acta, Inc., 1950.

189. Silber, R. H., and C. C. Porter. The determination of 17, 21-dihydroxy-20-ketosteroids in urine and plasma. Journal of Biological Chemistry, 1954, 210, 923-932.

190. Sollberg, A. Biological rhythm research. Amsterdam: Elsevier, 1965.

191. Stave, A. M. The effects of cockpit environment on long-term pilot performance. Human Factors, 1977, 19 (5), 503-514.

192. Steinberg, H. Effects of drugs on performance and incentives. In D. R. Laurence (Ed.), Quantitative methods in human pharmacology and therapeutics. London: Pergamon Press, 1959.

193. Surwillo, W. W. Cortical evoked response recovery functions: physiological manifestations of the psychological refractory period? Psychophysiology, 1977, 14, 32-39.

194. Teichner, W. H. Interaction of behavioral and physiological stress reactions. Psychological Review, 1968, 75, 271-291.

195. Teichner, W. H., and D. E. Olson. A preliminary theory of the effects of task and environmental factors on human performance. Human Factors, 1971, 13, 295-344.

196. Telford, C. W. The refractory phase of voluntary and associative responses. *Journal of Experimental Psychology*, 1931, *14*, 1-36.

197. Thackray, R. I. *Patterns of physiological activity accompanying performance on a perceptual-motor task (AM 69-8)*. Oklahoma City, Ok: Federal Aviation Administration, Office of Aviation Medicine, Civil Aeromedical Institute, April, 1969. (NTIS No. AD 697945)

198. Thackray, R. I., J. P. Bailey, and R. M. Touchstone. *Physiological, subjective, and performance correlates of reported boredom and monotony while performing a simulated radar control task (FAA-AM-75-8)*. Oklahoma City, Ok: Federal Aviation Administration, Office of Aviation Medicine, Civil Aeromedical Institute, August, 1975. (NTIS No. AD-A025426/8)

199. Thackray, R. I., K. N. Jones, and R. M. Touchstone. Personality and physiological correlates of performance decrement on a monotonous task requiring sustained attention. *British Journal of Psychology*, 1974, *65*, 351-358.

200. Threatt, D. H., and L. P. Perelli. *Computer program listings, flight plans and data files for the GAT-1 flying performance and physiological measurement system*. Brooks AFB, Tx: School of Aerospace Medicine, Aerospace Medical Division, August, 1977.

201. Trumbull, R. Diurnal cycles and work-rest scheduling in unusual environments. *Human Factors*, 1966, *8*, 385-398.

202. U.S. Civil Aeronautics Administration. *The role of fatigue in pilot performance* (Report No. 61). Washington, D.C.: Division of Research, Civil Aeronautics Administration, May, 1946.

203. Vernon, M. *Irritability*. New Haven: Yale University Press, 1913.

204. Vince, M. A. The intermittency of control movements and the psychological refractory period. *British Journal of Psychology*, 1948, *38*, 149-157.

205. Wadeson, R. N., J. W. Mason, D. A. Hamburg, and J. H. Handlon. Plasma and urinary 17-OHCS responses to motion pictures. *Archives of General Psychiatry*, 1963, *9*, 146-156.

206. Waller, A. D., and G. E. De Decker. The physiological cost of work in various departments of "The Times" printing house. *Journal of Physiology*, 1919-1920, *53*, cv-cvi.

207. Wargo, M. J. Time-estimation skill assessment: a demonstration of adaptive assessment techniques. *Human Factors*, 1971, *13*, 163-172.

208. Warren, N., and B. Clark. Blocking in mental and motor tasks during a 65-hour vigil. *Journal of Experimental Psychology*, 1937, *21*, 97-105.

209. Weiskotten, T. F., and J. E. Ferguson. A further study of the effects of loss of sleep. Journal of Experimental Psychology, 1930, 13, 247-266.

210. Welford, A. T. Fatigue and monotony. In O. G. Edholm and A. L. Bacharach (Eds.), The physiology of human survival. New York: Academic Press, 1965.

211. Welford, A. T. Fundamentals of skill. London: Methuen, 1968.

212. Welford, A. T. The psychologist's problem in measuring fatigue. In W. F. Floyd and A. T. Welford (Eds.), Symposium on fatigue. London: H. K. Lewis & Co., 1953.

213. Welford, A. T., R. A. Brown, and J. E. Gabb. Two experiments on fatigue as affecting skilled performance in civilian air crew. British Journal of Psychology, 1950, 40, 195-211.

214. Weybrew, B. B. Patterns of psychophysiological response to military stress. In M. H. Appley and R. Trumbull (Eds.), Psychological stress: issues in research. New York: Appleton-Century-Crofts, 1967.

215. Whiting, H. F., and H. B. English. Fatigue tests and incentives. Journal of Experimental Psychology, 1925, 8, 33-49.

216. Wickelgren, W. A., and D. A. Norman. Strength models and serial position in short-term recognition memory. Journal of Mathematical Psychology, 1966, 3, 316-347.

217. Wiener, E. L., and F. L. Keeler. Adaptive strategies in vigilance research. Ergonomics, 1975, 18, 403-414.

218. Wilkinson, R. T. Aftereffect of sleep deprivation. Journal of Experimental Psychology, 1963, 66, 439-442. (a)

219. Wilkinson, R. T. Effects of up to 60 hours of sleep deprivation of different types of work. Ergonomics, 1964, 7, 175-186.

220. Wilkinson, R. T. Interactions of lack of sleep with knowledge of results, repeated testing, and individual differences. Journal of Experimental Psychology, 1961, 62, 263-271.

221. Wilkinson, R. T. Interaction of noise with knowledge of results and sleep deprivation. Journal of Experimental Psychology, 1963, 66, 332-337. (b)

222. Wilkinson, R. T. Rest pauses in a task affected by lack of sleep. Ergonomics, 1959, 2, 373-380.

223. Wilkinson, R. T. Sleep deprivation. In O. G. Edholm and A. L. Bacharach (Eds.), The physiology of survival. New York: Academic Press, 1965.

224. Wilkinson, R. T. Sleep deprivation: performance tests for partial and selective sleep deprivation. In L. A. Abt & B. F. Riess (Eds.), Progress in Clinical Psychology (Vol. 8). New York: Grune, 1969, 28-43.

225. Wilkinson, R. T., and W. P. Colquhoun. Interaction of alcohol with incentive and with sleep deprivation. Journal of Experimental Psychology, 1968, 76, 623-629.

226. Williams, H. L., and A. Lubin. Speeded addition and sleep loss. Journal of Experimental Psychology, 1967, 73, 313-317.

227. Williams, H. L., A. M. Granada, R. C. Jones, A. Lubin. and J. C. Armington. EEG frequency and finger pulse volume as predictors of reaction time during sleep loss. Electroencephalography and Clinical Neurophysiology, 1962, 14, 64-70.

228. Williams, H. L., C. F. Gieseking, and A. Lubin. Some effects of sleep loss on memory. Perceptual and Motor Skills, 1966, 23, 1287-1293.

229. Williams, H. L., O. F. Kearney, and A. Lubin. Signal uncertainty and sleep loss. Journal of Experimental Psychology, 1965, 69, 401-407.

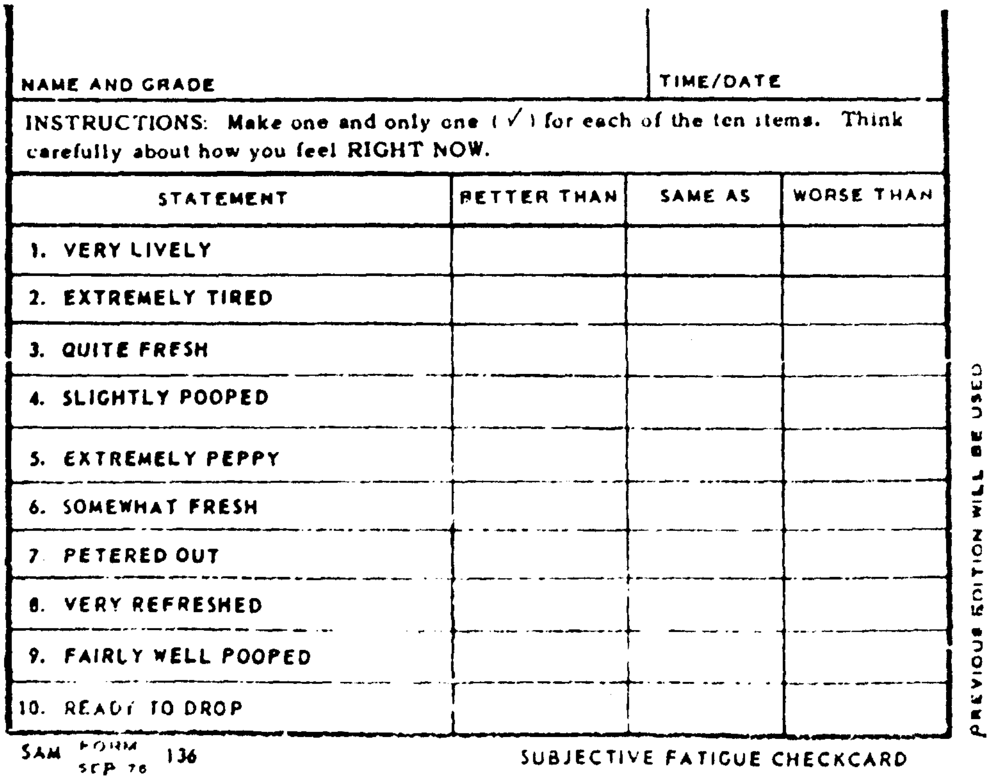
230. Williams, H. L., A. Lubin, and J. D. Goodnow. Impaired performance with acute sleep loss. Psychological Monographs, 1959, 73 (Whole No. 484).

231. Woodward, D. P., and P. D. Nelson. A user oriented review of the literature on the effects of sleep loss, work-rest schedules, and recovery on performance (ACR-206). Arlington, Va: Office of Naval Research, Biological and Medical Sciences Division Physiology Program, December, 1974. (NTIS No. AD-A009 778/2GI)

232. Woodworth. R. S., and F. L. Wells. Association tests. Psychological Monographs, 1911, 13 (5, Whole No. 57).

233. Yerkes, R. M., and J. D. Dodson. The relation of strength of stimulus to rapidity of habit-formation. Journal of Comparative Neurology and Psychology, 1908, 18, 459-482.

APPENDIX A



NAME AND GRADE	TIME/DATE

INSTRUCTIONS: Make one and only one (✓) for each of the ten items. Think carefully about how you feel RIGHT NOW.

STATEMENT	BETTER THAN	SAME AS	WORSE THAN
1. VERY LIVELY			
2. EXTREMELY TIRED			
3. QUITE FRESH			
4. SLIGHTLY POOPED			
5. EXTREMELY PEPPY			
6. SOMEWHAT FRESH			
7. PETERED OUT			
8. VERY REFRESHED			
9. FAIRLY WELL POOPED			
10. READY TO DROP			

PREVIOUS EDITION WILL BE USED

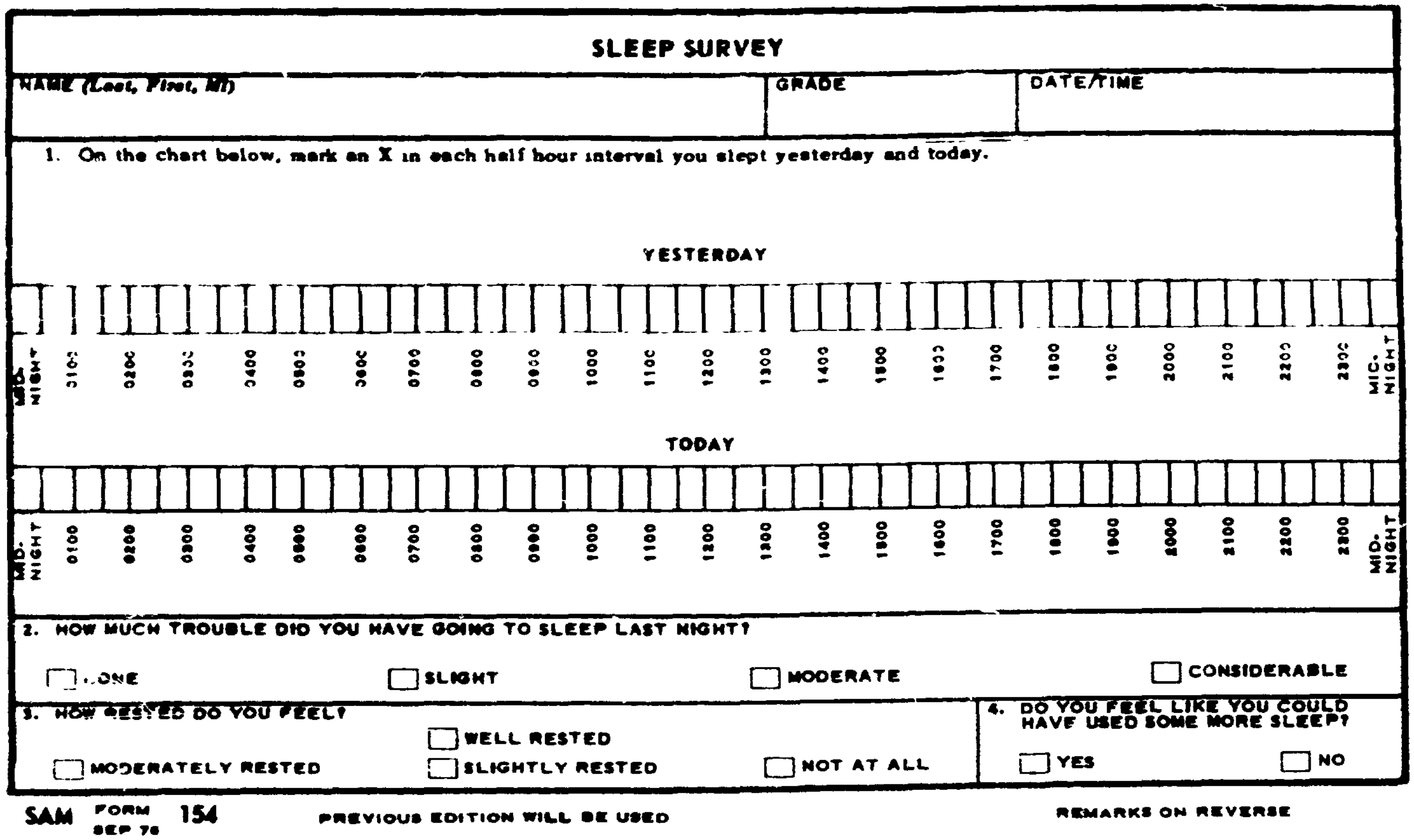
SAM FORM SEP 76 136

SUBJECTIVE FATIGUE CHECKCARD

Figure A-1. USAF School of Aerospace Medicine "Subjective Fatigue Checkcard." (The Checkcard is scored by adding 2 points for each check in the "better than" column, and 1 point for each check in the "same as" column. Checks in the "worse than" column are not counted.)

1. Feeling active and vital; alert; wide awake.
2. Functioning at a high level; not at a peak but able to concentrate.
3. Relaxed; awake; responsive, but not at full alertness.
4. A little foggy; let down; not at peak.
5. Foggy; slowed down; beginning to lose interest in remaining awake.
6. Sleepy; woozy; prefer to be lying down; fighting sleep.
7. Almost in reverie; sleep onset soon; losing struggle to remain awake.

Figure A-2. Modified "Stanford Sleepiness Scale" (103). [Subjects reported the number corresponding to their current subjective estimate of how sleepy they felt at the time of each data collection period. This number was entered on their SAM Subjective Fatigue Checkcard, which was filled out at the same time. (Note: Wording modified slightly from original scale for adaptation to present study.)]



SLEEP SURVEY

NAME (Last, First, MI)
GRADE
DATE/TIME

1. On the chart below, mark an X in each half hour interval you slept yesterday and today.

YESTERDAY

MID-NIGHT 0100 0200 0300 0400 0500 0600 0700 0800 0900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000 2100 2200 2300 MID-NIGHT

TODAY

MID-NIGHT 0100 0200 0300 0400 0500 0600 0700 0800 0900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000 2100 2200 2300 MID-NIGHT

2. HOW MUCH TROUBLE DID YOU HAVE GOING TO SLEEP LAST NIGHT?

☐ NONE ☐ SLIGHT ☐ MODERATE ☐ CONSIDERABLE

3. HOW RESTED DO YOU FEEL?

☐ WELL RESTED
☐ MODERATELY RESTED ☐ SLIGHTLY RESTED ☐ NOT AT ALL

4. DO YOU FEEL LIKE YOU COULD HAVE USED SOME MORE SLEEP?

☐ YES ☐ NO

SAM FORM SEP 76 154 PREVIOUS EDITION WILL BE USED REMARKS ON REVERSE

Figure A-3. USAF School of Aerospace Medicine "Sleep Survey Form."

DEPARTMENT OF THE AIR FORCE
USAF SCHOOL OF AEROSPACE MEDICINE (AFSC)
BROOKS AIR FORCE BASE, TEXAS 78235

DEPARTMENT OF DEFENSE
UNITED STATES OF AMERICA

SUBJECT: Consent of Volunteer

1. I hereby volunteer to participate as a test subject in the study, "Effects of Fatigue Stressors on Psychobiological Indices of Air Crew Performance." I understand that the fatigue stressors employed will be within known human tolerance limits and are not designed to test unknown physiological reactions.

2. Capt Layne P. Perelli (USAFSAM/VNE) has discussed with me to my satisfaction the reasons for this experiment and its possible adverse and beneficial consequences including the effects of extended sleep deprivation, such as production of temporary abnormal EKG, and EEG tracings and possible precipitation of seizures in epileptics. I understand that during tests I will be required to wear EKG leads and to have my body core temperature recorded from a temperature probe inserted into my rectum. I understand that I may be required to stay awake for periods of up to 36 hours and perform a complex task for which I have been previously trained. During the recovery period I will be observed and required to obtain adequate rest before being allowed to resume my normal duties. I further understand the importance of following my assigned work-rest cycle as closely as possible to insure the results of the study are meaningful.

3. This consent is voluntary and has been given under circumstances in which I can exercise free power of choice. I have been informed that I may at any time revoke my consent and withdraw from the experiment without prejudice and that the investigator or physician may terminate the experiment at any time regardless of my wishes.

4. I certify that I have read and understand the hazards, constraints, and responsibilities expressed above, and agree to participate as a test subject in programs pursuant to this protocol. I understand that before my use as a test subject, I must inform the principal investigator and project physician of any change to my medical status. This information will include any medications I have taken and medical or dental care/treatment received since my last use as a test subject.

(Signature of Officer Who Advised of Possible Consequences)	(Signature of Volunteer)
(Signature of Witness)	(Typed Name, Grade, Serial #)

Date: ____________________

Figure A-4. An example of the volunteer consent form signed by each subject participating in this study.

ABBREVIATIONS, ACRONYMS, AND SYMBOLS

AFOQT	Air Force Officer Qualification Test
AFR	Air Force Regulation
AHRV	average heart rate variability
B	Block
BD	block day
bpm BPM	beats per minute
BRAC	basic rest-activity cycle
CA	course adjustment
CCF	critical fusion frequency
CHR	changes in heart rate
CRT	changes in rectal temperature
CTSE	combined total seconds of error
df	degrees of freedom
DIPT	Discrete Information Processing Test
E	experimenter
ECG	electrocardiogram
E-paced	Experimenter-paced
FD	flight day
FDP	Flight Director Panel
FMT	Flight Maneuver Test
FPM	feet per minute
G	Group
GAT-1	Link General Aviation Trainer-1
HR	heart rate
HRV	heart rate variability
Hz	hertz
K	potassium
kg	kilogram(s)
KR	knowledge of results
L	leg(s)
L/P	legs within period

ABBREVIATIONS, ACRONYMS, AND SYMBOLS (CONT'D)

mg	milligram(s)
μg	microgram(s)
MMPI	Minnesota Multiphasic Personality Inventory
mph	miles per hour
ms	millisecond(s)
MS	mean square
MTPB	Multiple-Task Performance Battery
N	nitrogen
Na	sodium
P	period
R	Recovery day
rpm	revolutions per minute
RT	reaction time
S	subject
SAM	USAF School of Aerospace Medicine
SDC	subjective fatigue and SSS data collection
SDBPM	standard deviation units of beats per minute
17-OHCS	17-hydroxycorticosteroids
SF	subjective fatigue
SLT	Straight and Level Test
S-paced	Subject-paced
Ss	subjects
SSS	Stanford Sleepiness Scale
t	time
T/O	takeoff
WAIS	Wechsler Adult Intelligence Scale