CREWMEMBER PERFORMANCE BEFORE, DURING, AND AFTER SPACEFLIGHT

THOMAS H. KELLY, ROBERT D. HIENZ, TROY J. ZARCONE,
RICHARD M. WURSTER, AND JOSEPH V. BRADY

UNIVERSITY OF KENTUCKY,
THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE,
UNIVERSITY OF ROCHESTER MEDICAL CENTER,
AND ALPHAINSIGHT, INC.

The development of technologies for monitoring the welfare of crewmembers is a critical requirement for extended spaceflight. Behavior analytic methodologies provide a framework for studying the performance of individuals and groups, and brief computerized tests have been used successfully to examine the impairing effects of sleep, drug, and nutrition manipulations on human behavior. The purpose of the present study was to evaluate the feasibility and sensitivity of repeated performance testing during spaceflight. Four National Aeronautics and Space Administration crewmembers were trained to complete computerized questionnaires and performance tasks at repeated regular intervals before and after a 10-day shuttle mission and at times that interfered minimally with other mission activities during spaceflight. Two types of performance, Digit-Symbol Substitution trial completion rates and response times during the most complex Number Recognition trials, were altered slightly during spaceflight. All other dimensions of the performance tasks remained essentially unchanged over the course of the study. Verbal ratings of Fatigue increased slightly during spaceflight and decreased during the postflight test sessions. Arousal ratings increased during spaceflight and decreased postflight. No other consistent changes in rating-scale measures were observed over the course of the study. Crewmembers completed all mission requirements in an efficient manner with no indication of clinically significant behavioral impairment during the 10-day spaceflight. These results support the feasibility and utility of computerized task performances and questionnaire rating scales for repeated measurement of behavior during spaceflight.

Key words: NASA, spaceflight, operant behaviour, performance measures, self-report measures, humans

---

The behavioral effects of spaceflight have been of abiding interest and concern since the earliest man-made satellites transported living organisms beyond earth's atmosphere (Lindsley, 1972). The nonhuman animal pretest flights for Project Mercury in the late 1950s and early 1960s demonstrated that the stable parameters of schedule-controlled performances established prior to the mission remained intact during flight (Brady, 1990). Although the reports of monitored space crews during early missions (e.g., Apollo and Skylab) following the inaugural flight of Gagarin in 1961 were, for the most part, anecdotal in nature (e.g., Leonov & Lebedev, 1973; Taylor, 1989), they did indicate clearly that humans were capable of maintaining high functional levels during at least short-duration space ventures. As flight durations have been extended, however, concerns about crewmember compliance and motivation concerns have begun to emerge (Covault, 1988). Behavioral issues, however, have not been a high priority in confronting the many technological challenges of spaceflight. Relatively few systematic studies of human performance under such conditions have been reported, and the

Financial support for this study was provided by NASA Flight Research Contract NAS9-19167 from the National Aeronautics and Space Administration. Cleeve Emurian and Danielle Pyle provided expert support in ground-based preliminary studies, data analysis, and manuscript preparation. Karen Lawrence, Herbert Anderson, and George Ishmael, all from Lockheed-Martin, and Gary Gutschewski of NASA provided invaluable assistance in experiment integration and protocol implementation. Special thanks are due the flightcrew, Terry Wilcutt, Joe Edwards, Bonnie Dunbar, and Mike Anderson, for their patience and perseverance over the many months of this project.

This article is dedicated to the memory of Mike Anderson whose life was unexpectedly and tragically shortened in a subsequent mission.

Correspondence concerning this article should be addressed to either Thomas H. Kelly, Department of Behavioral Science, College of Medicine, University of Kentucky, Lexington, Kentucky 40536-0086 (e-mail: thkelly@uky.edu), or Joseph V. Brady, Division of Behavioral Biology, The Johns Hopkins University School of Medicine, Behavioral Biology Research Center, 5510 Nathan Shock Drive, Suite 3000, Baltimore, Maryland 21224 (e-mail: jvb@jhmi.edu).

doi: 10.1901/jeab.2005.77-04

behavioral effects of spaceflight remain, in large part, undetermined (Taylor, 1989).

Advances in propulsion and life-support technologies over the past several decades have increased the feasibility of extended-duration spaceflight. In developing the requirements for such exploratory initiatives (e.g., Mars [''Bioastronautics Roadmap,'' n.d.]), significant concerns that have yet to be resolved include the development of a technology that will assess adequately the behavioral effects of long-term extraterrestrial missions as well as establish and maintain safe and productive performances by humans living in confined and isolated microsocieties under extreme environmental conditions (Brady, 1990). Behavior analytic approaches provide an efficient yet rigorous methodological framework for studying the performance of individuals and groups (e.g., Brady, 1990, 1993). Brief and repeated behavioral measurements can be conducted under both control and experimental conditions (i.e., A-B-A designs) with even a small number of participants to identify the effects of spaceflight conditions on performance within the challenging context of a National Aeronautics and Space Administration (NASA) mission. The primary purpose of the present study was to examine the feasibility of using a behavior analytic approach with a brief computerized test battery to assess the effects of spaceflight on crewmembers' performance.

A number of conditions that could disrupt the stability of human behavior during spaceflight or in other similar extreme environments have been identified (e.g., Christensen & Talbot, 1986; Kanas, 1987; Stuster, 1996). Several reports, for example, have documented the adverse effects of motion sickness (e.g., Ratino, Repperger, Goodyear, Potor, & Rodriguez, 1988; Thornton, Moore, Pool, & Vanderploeg, 1987), altered sleep cycles (Frost, Schumate, Salamy, & Booher, 1976), and physiological changes (e.g., Day, Allen, Mohajerani, Greenisen, Roy, & Edgerton, 1995; West, 1984) on the behavior of individual crewmembers during spaceflight. In this context, several dimensions of human behavior have been examined, including perception, gross movement, and man–machine interactions (for review see Manzey & Lorenz, 1998). Gross body movements have been shown to change under conditions of weightlessness (e.g., Gerathewohl, Stallings, & Strughold, 1957), and alteration in the discrimination of weight and mass have been reported in microgravity environments (Ross, Schwartz, & Emmerson, 1987). Visual performance, however, appears generally unaffected during spaceflight (e.g., Leone, Lipshits, Gurfinkel, & Berthoz, 1995a, 1995b). Small changes in topography, but not accuracy, of fine movement of the hand and fingers also have been documented upon initial entry into microgravity (e.g., Newman & Lathan, 1999; Ross et al., 1987; Sangals, Heuer, Manzey, & Lorenz, 1999). These effects have been attributed to proprioceptive changes associated with the altered perceptual response to weight and mass. Adaptation to these effects has been reported, as have rebound effects upon reentry to normal gravity conditions following adaptation during short-term flight performance.

Some reports have suggested that modest cognitive and psychomotor performance changes also occur during spaceflight (e.g., increased reaction times, impaired tracking efficiency, and reduced time-estimation accuracy). In most instances, however, methodological factors constrain the interpretation of these data. Substantial variability in performance during spaceflight, or changes in baseline performance before and after flight, make it difficult to attribute behavioral changes that occur during such missions to the spaceflight environment per se (i.e., Eddy, Schiflett, Schlegel, & Shehab, 1998; Manzey, Lorenz, Schiewe, Finell, & Thiele, 1995; Matsakis, Lipshits, Gurfinkel, & Berthoz, 1993; Monk, Buysse, Billy, Kennedy, & Willrich, 1998). Other studies have established stable performance through extensive training prior to spaceflight but have not used objective quantitative or statistical criteria to verify the significance of changes during flight (Benke, Koserenko, Watson & Gerstenbrand, 1993; Manzey, Lorenz, Heuer, & Sangals, 2000). As a result, findings are seldom replicated across studies (e.g., Manzey, Lorenz, & Poljakov, 1998; Manzey et al., 1995; Manzey, Lorenz, Schiewe, Finell, & Thiele, 1993). The bulk of the available evidence suggests that human behavior remains largely unchanged during spaceflight (e.g., Newman & Lathan, 1999; Ratino et al., 1988; Sangals et al., 1999), but it also is important to recognize that this supposition is based on a relatively small number of methodologically constrained studies.

A second purpose of this study was to expand the range of behavioral measures that have been used to evaluate the effects of spaceflight on human performance. The behavioral tasks were selected from among those that have been used successfully in clinical psychopharmacology research. They are easily acquired, require little time to complete, and engender performances that are relatively stable across repeated assessment yet sensitive to subtle manipulations of factors known to alter human behavior (e.g., drugs, sleep, and nutrition).

METHOD

Participants

Four adult crewmembers (3 males, 1 female) who participated in the 10-day NASA space shuttle mission STS-89 participated in the study. They included both flight crew (commander and pilot) and mission specialists (payload personnel). All participants had completed college degrees and had received extensive postgraduate and other professional training prior to the study. All participated in a consent procedure approved by the NASA Institutional Review Board prior to providing written consent.

Apparatus

The behavioral tasks were presented on a Macintosh® Powerbook 170 with an attached Kensington Keypad (Model CA-07941624, San Mateo, CA). Crewmembers were asked to record daily food and fluid intake, medication usage, and sleep duration in logbooks with individual identification and study-phase information imprinted on the cover.

Task Battery

The six tasks composing the behavioral test battery were chosen on the basis of previous research showing that performance on the tasks remained stable over repeated testing following minimal training and in the absence of formal environmental perturbations. Previous research had established as well that the performance battery required only a brief amount of time to complete and was sensitive to manipulations, such as drug administration and changes in nutrition, known to affect human behavior (e.g., Foltin et al., 1990; Kelly, Foltin, & Fischman, 1993; Kelly, Foltin, Rolls, & Fischman, 1994; Kelly, Foltin, Serpick, & Fischman, 1997; Ward, Kelly, Foltin, & Fischman, 1997). During each test session, the tasks were presented in the following order:

Profile of Mood States (POMS). The POMS is a self-report questionnaire designed to measure current mood (McNair, Lorr, & Droppleman, 1971). Each of the 72 items in a research version of the POMS was presented sequentially on the computer display screen above a row of boxes numbered 1 to 5, corresponding to Keys 1 to 5 on the Kensington keypad. Crewmembers responded to presentation of each item by pressing Key 1 for ''not at all,'' Key 2 for ''a little,'' Key 3 for ''moderately,'' Key 4 for ''quite a bit,'' or Key 5 for ''extremely.'' These labels were displayed above each box on the screen. The corresponding box on the screen display turned dark immediately upon activation of a key on the keypad. Answers could be changed by pressing a different key number before pressing the Enter key on the keypad to record the answer. The amount of time required to complete all 72 items, as well as the time required to complete each item, was recorded. Anxiety, Anger, Fatigue, Depression, Vigor, and Confusion scale scores were derived from answers to individual items. The research version of the POMS also provided modified Friendliness and Elation scale scores. Additionally, an Arousal scale was derived by subtracting the sum of the Fatigue and Confusion scale scores from the sum of the Anxiety and Vigor scale scores. A Positive Mood scale also was derived by subtracting the Depression scale score from the Elation scale score. Crewmembers typically completed the POMS in 1.5 to 3.5 min.

Visual Analogue Scale (VAS). The VAS is a questionnaire designed to measure current interoceptive conditions (Foltin & Fischman, 1991). Each of 11 words, euphoric, stressed, fatigue, anxious, thirsty, stimulated, lethargic, hungry, faint, sleepy, and nauseous, was presented in sequence above a line labeled ''not at all'' on the left end and ''extremely'' on the right end. Crewmembers used a track ball to move a cursor to a position along the line and then clicked the trackball button to select the line position. Responses could be changed by repositioning the cursor along the line and clicking again before clicking a ''Next'' box on the display screen below the line to record the position and display the next word in the sequence. The number of discrete units from the left line endpoint to the cursor mark on the line determined the rating of each item (score range, 0 to 100). The response time for each word as well as the amount of time required to complete all 11 words was recorded. Crewmembers typically completed the VAS in 0.5 to 2.0 min.

Differential-reinforcement-of-low-response-rate (DRL). DRL tasks have been used to examine the consistency and accuracy of time estimation (e.g., Kelly et al., 1994). Crewmembers pressed the ''0'' key on the keypad and earned a point on a counter each time the key press was separated in time by 12 s or more from the preceding key press. Responses separated by less than 12 s reset the timer and earned no points. A 0 key display on the screen turned dark in parallel with activation of the 0 key on the keypad, a counter displayed on the screen increased by one with each correct press, and a gauge with 15 discrete units increased by one unit following each correct press. The total number of responses, the proportion of responses meeting the temporal requirement, and the mean and distributions of interresponse times were measured. The task was presented for 3 min.

Repeated acquisition of response sequences (RA). The RA task has been used to examine the process by which new behavior is learned (Boren & Devine, 1968; Fischman, 1978). During this 3-min task, crewmembers were required to learn a new 10-response sequence using four keys (i.e. 1, 3, 7, and 9) on the keypad. A new sequence (e.g., 3-7-7-1-9-1-3-9-3-1) was randomly chosen by the computer each time the task was presented and remained unchanged throughout the duration of that exposure to the task. As each correct response in the sequence was emitted, a position counter on the screen increased in steps from 0 to 9. Incorrect responses were followed by a 1-s time-out, during which the screen was blank, but did not change the position step counter or the sequence. After the 10th and final correct response in a sequence was emitted, a trial (point) counter on the screen increased by one, and the position counter was reset to zero signaling a return to the beginning of the response sequence.

During the first trial, 60% of responses, on average, were incorrect. As crewmembers acquired the response sequence, the number of incorrect responses occurring during the completion of the sequence decreased. To evaluate task performances, patterns of correct and incorrect responding during the task were quantified using an index of curvature (Fry, Kelleher, & Cook, 1960) to assess acquisition efficiency during each session. Additionally, the session rates of both correct and incorrect responding were examined.

Number recognition (NR). This modified delayed matching-to-sample task has been used to examine memory (Sternberg, 1966). The 5-min task consisted of a series of trials, each signaled by a ''Ready'' cue on the screen. When the two keys labeled ''Y'' (Yes) and ''N'' (No) on the keypad were both depressed, the Ready message was replaced by a one- to six-digit sample number display presented for 1.5 s. The number array was then removed, and after a 2 s blank-screen delay, a single test digit was presented for 2.5 s. Within 2.5 s, crewmembers were required to lift the Y key finger if the single test digit was contained within the sample display, or lift the N key finger if it was not. A counter displayed on the screen increased by one after each correct response. Both trial accuracy and response time (on correct trials) were analyzed. Response times also were analyzed as a function of trial type (yes or no) and number of digits in the sample stimulus (1 to 6).

Digit-symbol substitution task (DSST). A modified version of the computerized DSST (McLeod, Griffiths, Bigelow, & Yingling, 1982), adapted from the Wechsler Adult Intelligence Scale (e.g., Wechsler, 1997), was used to assess psychomotor performance. During this 2-min task, nine 3-row by 3-column arrays of boxes, labeled 1 to 9 from left to right, were displayed at the top of the computer screen. Each row contained one black and two white boxes, with the location of the black box determined randomly. A random number, between 1 and 9, was displayed in the center of the screen during each trial, indicating which of the nine arrays was to be reproduced in a 3-row by 3-column pattern on the keypad. One response in each of the three rows, corresponding to the position of the black box, was required per trial, and the third response per trial generated a new random number indicating the next array to be reproduced. A point counter displayed on the screen increased by one following each correct trial. A new random pattern of black and white boxes was presented following the completion of each block of 25 trials. Trial rate and accuracy were recorded.

Completion of all six tasks required less than 20 min. After the DSST, a 15-s screen display provided feedback on performance by displaying the number of points acquired during each of the tasks as well as the total number of points earned during the session. There were no financial contingencies associated with these points, although in poststudy debriefing, crewmembers reported that they had discussed and compared session points. Crewmembers also were instructed to record daily food and fluid intake, medication usage, and sleep duration in logbooks immediately following each test session.

Experimental Phases

The study consisted of four phases. At the start of the study and prior to each phase, a separate schedule was developed for each crewmember such that sessions would be completed at regular intervals throughout each phase at times that would not interfere with other mission activities. Unexpected schedule changes and conflicting mission priorities, however, resulted in occasional session postponement or cancellation. Under such conditions, the actual number and time of sessions during each phase varied unavoidably among crewmembers. The several phases of the experiment, however, occurred in the following order:

Training. The initial training phase consisted of orientation and repeated practice sessions, beginning 5 months prior to the flight. During orientation, crewmembers were given written instructions describing the operation of the computer as well as the procedures for completing the computer-based questionnaires and performance tasks. The initial orientation session was completed in the presence of study investigators who answered questions pertaining to the operation of the computers and tasks, as requested. Additional practice sessions were completed in assigned offices with the crewmembers seated in chairs before a laptop computer on the desk. The number and timing of additional practice sessions during training varied among crewmembers (a range of 8 to 16 sessions), but all received multiple exposures to each component of the battery, and performances were demonstrably proficient (i.e., high response rates with few errors) before commencement of succeeding experimental phases. No programmed contingencies were associated with performance outcomes.

Preflight. Crewmembers were quarantined in quarters together during a 10-day interval immediately prior to the flight during the preflight phase of the study. During this phase, sessions were completed in a shared office with the crewmember seated in a chair before the computer on top of a table. Two crewmembers completed five sessions and 2 completed six sessions during the preflight phase.

Inflight. The inflight phase of the study was conducted in the course of STS 89, a 10-day spaceflight mission aboard space shuttle Endeavor. No performance sessions were programmed during the first 2 or the last 2 days of the mission. The experimental performance measures during the inflight phase were obtained during the intervening 6 days under zero gravity conditions. All sessions were conducted in the commander's seat on the space shuttle flight deck with the Velcro-mounted computer and keypad attached to an aluminum lap desk secured to the crewmember's upper thighs with Velcro straps. The flight deck provided a functional working space that minimized distractions under conditions that were physically separated from the main work station areas of the shuttle. Two crewmembers completed five inflight performance sessions, and the other 2 crewmembers completed 6 and 10 sessions, respectively. At least one of each crewmember's sessions was videotaped.

Postflight. Two crewmembers completed six postflight sessions during a 9-day interval beginning 4 days after the shuttle's return landing, and the other 2 crewmembers completed nine sessions each. As with the preflight phase of the study, postflight sessions were completed in shared crew quarters with the crewmembers seated in a chair before the computer on top of a table.

Data Analysis

Although mission demands imposed variability in the frequency and timing of sessions both within and between study participants, each crewmember completed at least five sessions during each phase of the study. Task performance and questionnaire responses during the final five sessions of the preflight phase and the initial five sessions of the inflight and postflight phases of the study were examined. A two-way repeated-measures analysis of variance (ANOVA) with study phase (preflight, inflight, postflight) and session (1 to 5) was conducted on each outcome measure using the Tukey-A (HSD) procedure to follow up significant main effects and simple-effects analyses to evaluate significant interactions. Results were considered significant at $p < .05$. The number of factors in the ANOVA was expanded to examine the influence of additional task factors when necessary. A four-way ANOVA, for example, was used to analyze the effects of trial type (yes and no trials) and number of digits in the sample stimulus (1 to 6) on accuracy and response times during the NR task.

RESULTS

Self-Report Measures

The Fatigue scales of both the POMS, $F(2, 6) = 15.32$, $p < .05$, and the VAS, $F(2, 6) = 13.16$, $p < .05$, varied significantly across study phases. Scores during the postflight phase (POMS: 1.20 ± 0.40, mean ± SEM; VAS: 12.90 ± 4.39) were lower than those during either preflight (POMS: 3.85 ± 0.78; VAS: 27.65 ± 5.46) or inflight (POMS: 4.95 ± 0.76; VAS: 34.20 ± 6.54) phases. The Arousal scale of the POMS also varied significantly across study phases, $F(2, 6) = 6.67$, $p < .05$, but in contrast to the Fatigue scores, Arousal ratings were higher during the postflight phase (13.15 ± 1.45) than during the other two phases of the experiment (preflight: 9.90 ± 2.24; inflight: 8.0 ± 2.01), with differences between inflight and postflight phases reaching statistical significance ($p < .05$). No session or phase-by-session interactions were observed on Arousal or Fatigue ratings. Figure 1 illustrates the Fatigue and Arousal ratings for a representative crewmember (S3).

Individual crewmember ratings on other POMS and VAS items also varied in a systematic manner across study phases. VAS Hungry and Thirsty rating scores, for example, increased during the inflight and/or postflight phases for 3 of the 4 crewmembers (S2, S3, and S4), whereas VAS Sleepy rating scores decreased during the postflight phase for 3 of the 4 crewmembers (S1, S2, and S3). There were no other consistent POMS or VAS rating scale changes either within or between crewmembers across phases of the study, and no other statistically significant rating scale effects were observed.

DRL Task Measures

The number of responses that did and did not meet the IRT requirement (12 s) during each session of DRL task performance are shown in the left column of Figure 2. Performance accuracy was maintained at a high level throughout all phases of the experiment. Over 92% of responses during each 3-min performance interval met the IRT requirements. There were no systematic changes in performance as a function of session or study phase on this task.

RA Task Measures

The right column of Figure 2 presents response rate and efficiency during sessions of RA task performance. High rates of correct responding were observed across the entire study (overall mean of 1.3 ± 0.11 responses per second). The variability in rate across sessions was related to the time associated with acquisition of the response sequence; once the sequence had been acquired, crewmembers completed the sequence at a high rate. No systematic changes in response rate were observed as a function of session or study phase, although the response rate by Participant S4 increased across the study. When overall response rate remains high, an index of curvature for incorrect responses can be used as a measure of acquisition efficiency. An index of curvature value less than zero reflects negative acceleration in errors across trials, and smaller values represent enhanced acquisition efficiency. Acquisition efficiency varied across phases, $F(2, 6) = 6.79$, $p < .05$, with values during the inflight and postflight phases being significantly lower ($p < .05$) than during the preflight phase. There was no evidence in these data for a selective change

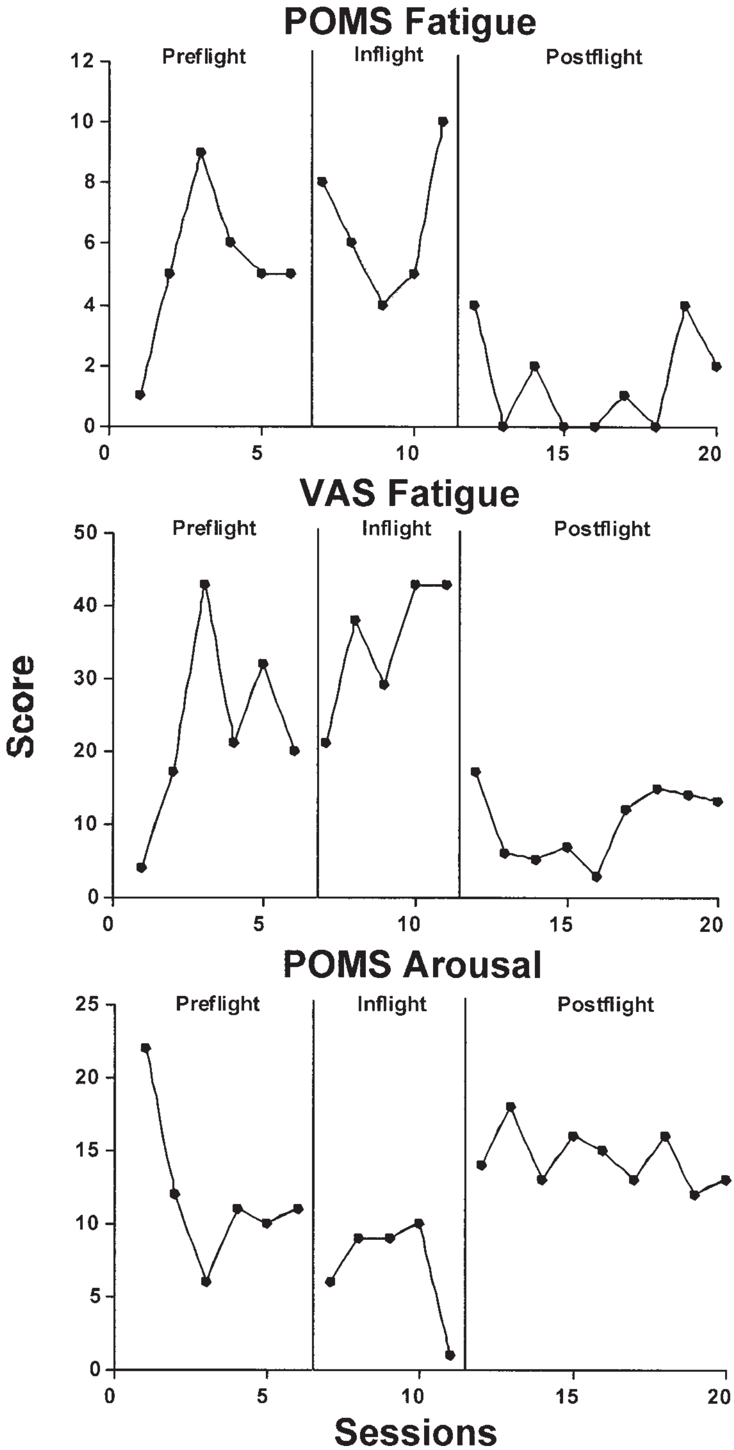

Fig. 1. Profile of Mood State (top panel), Visual-Analog Scale Fatigue (middle panel), and Profile of Mood State Arousal (bottom panel) ratings in successive sessions during the preflight, inflight, and postflight phases of the study for NASA flight crewmember S3.

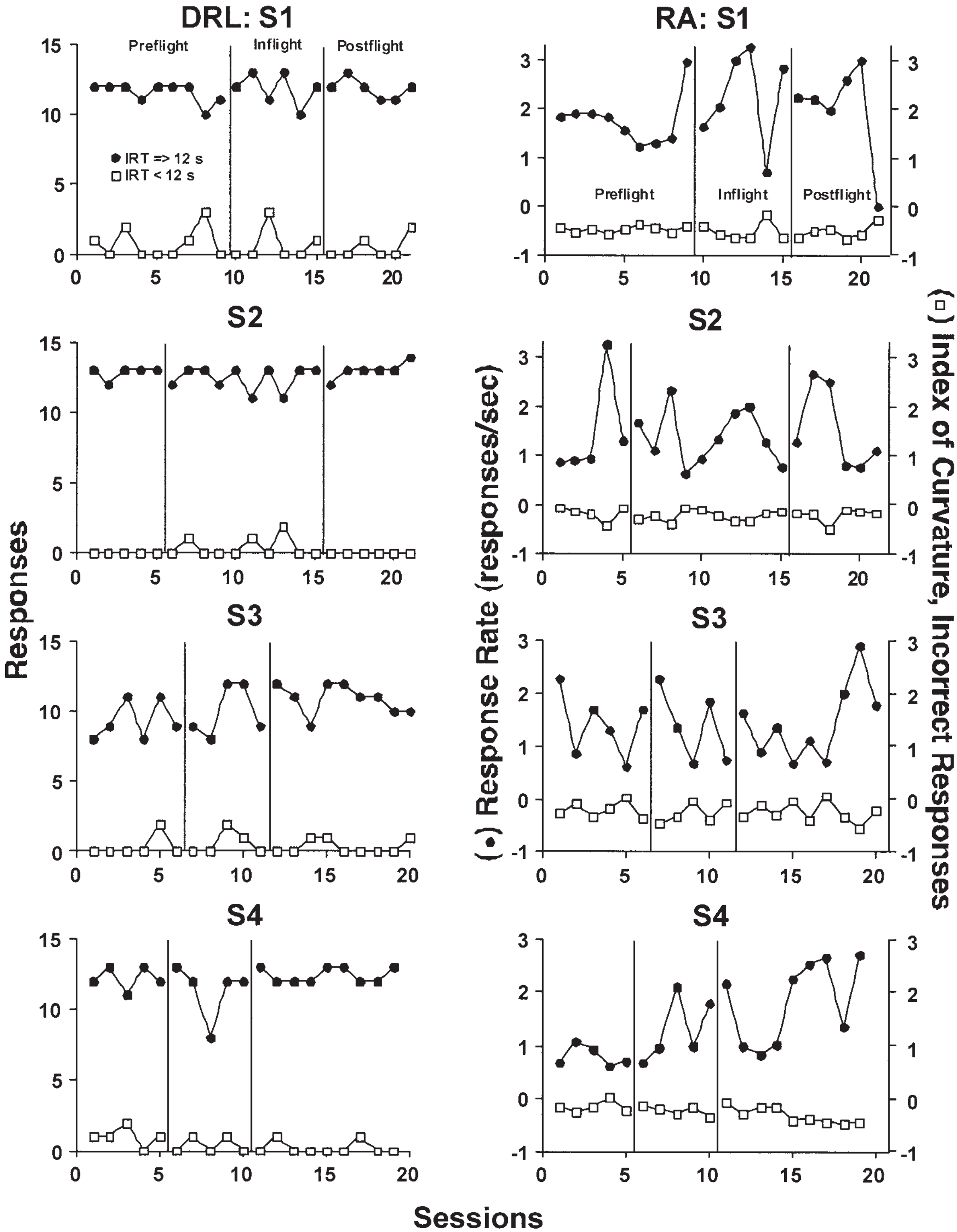

Fig. 2. Number of interresponse times that did (IRT = > 12 s) or did not (IRT < 12 s) engender point increments during the differential-reinforcement-of-low-rate 12-s task (left column), and total response rate (responses per second) and index of curvature for incorrect responses during the repeated acquisition task (right column) on successive sessions during the preflight, inflight, and postflight phases of the study for each of the 4 NASA flight crewmembers.

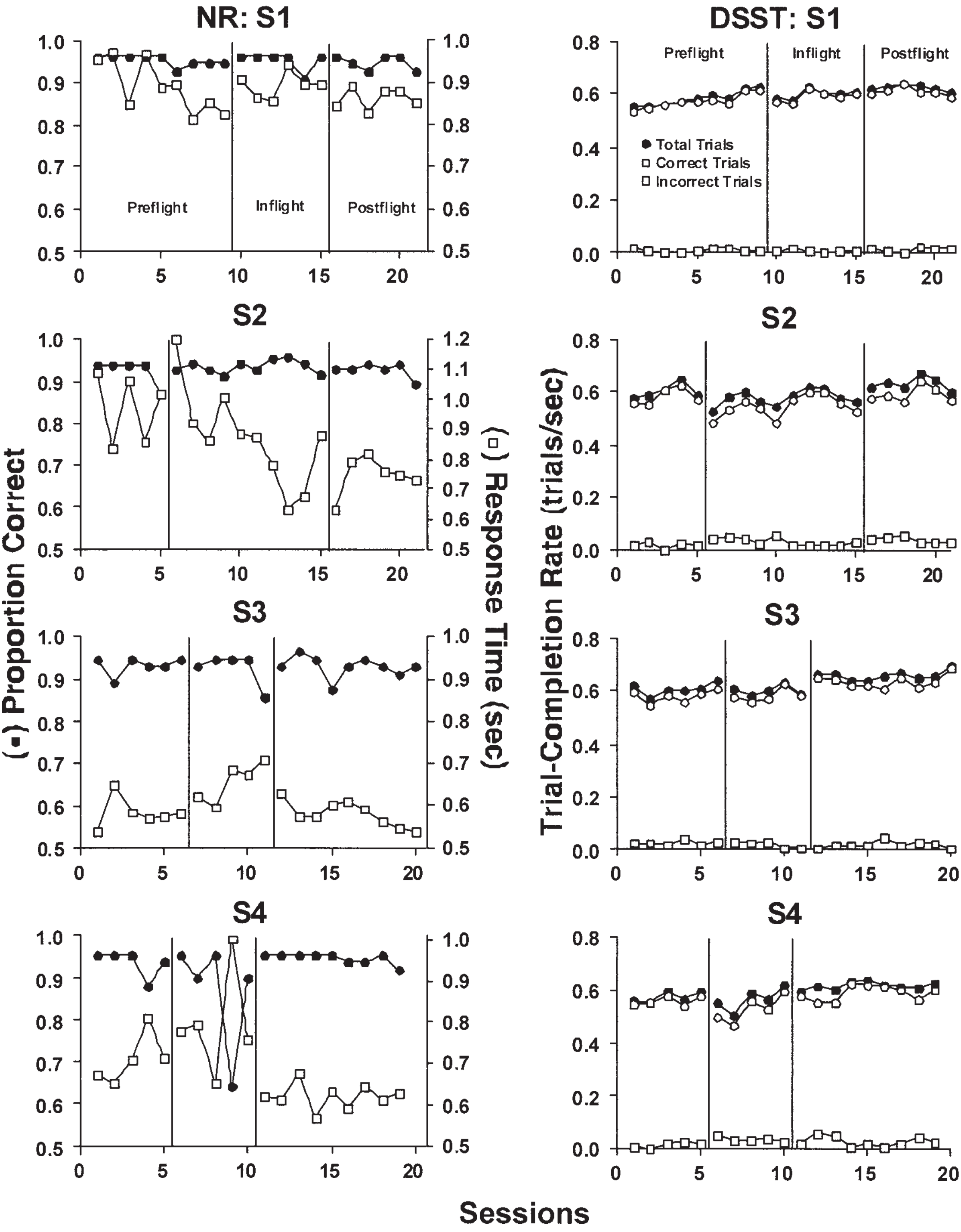

Fig. 3. Proportion correct and mean response time (in seconds) during the number recognition task (left column) and total, correct, and incorrect digit-symbol substitution task trial completion rate (trials per second; right column) on successive sessions during the preflight, inflight, and postflight phases of the study for each of the 4 NASA flight crewmembers.

in learning performance during spaceflight because the improved efficiency observed during inflight was either sustained or further enhanced during postflight.

NR Task Measures

Figure 3 (left column) presents accuracy and response times on NR task trials across the study. More than 90% of the NR trials were accurate during all phases of the study, with no systematic changes in accuracy observed across session, study phase, or trial type.

Response time varied among crewmembers. Response times decreased across sessions for Participant S2 and were lower during postflight relative to preflight and inflight phases for Participant S4. Figure 4 presents mean response times on Yes and No trials as a function of the number of digits in the sample display during the preflight, inflight, and postflight phases of the experiment for each of the 4 crewmembers. Response times during Yes trials were significantly shorter than during No trials, $F(1, 3) = 36.21$, $p < .05$, and both Yes and No response times increased as a function of the number of digits contained in the sample stimulus, $F(5, 15) = 13.85$, $p < .05$. A significant three-way interaction was observed as a function of study phase, trial type, and number of digits contained in the sample stimulus, $F(10, 30) = 2.40$, $p < .05$. Simple-effects analyses indicated that differences in response times between Yes and No trials increased significantly as a function of the number of digits in the sample stimulus during only the inflight phase of the study ($p < .05$), and the two-way interaction of trial type by number of sample stimulus digits was significant only during the inflight phase ($p < .05$). Although the magnitude of the effect was modest, a similar interaction pattern was in evidence for each subject. These data suggest that response times during complex stimulus presentations increased to a greater extent during inflight than in either the preflight or postflight phases of the study.

DSST Measures

Figure 3 (right column) presents the rate of trial completion (trials per second) for correct, incorrect, and total DSST trials for each of the 4 crewmembers across consecutive sessions during the preflight, inflight, and postflight phases of the experiment. A high level of accuracy was sustained throughout the experiment. Incorrect trial rates remained consistently low, with no changes across study phases. Trial completion rate increased with repetition across the study. Significant differences between phases were observed in both total, $F(2, 6) = 19.11$, $p < .05$, and correct, $F(2, 6) = 8.68$, $p < .05$, trial rates. Significant increases were observed during postflight relative to both the preflight and inflight phases ($p < .05$). Slight decreases in total and correct trial rates at the onset of the inflight phase, relative to the preflight phase, were not statistically significant.

Logbook Entries

The findings from the logbook entries for individual crewmembers across the phases of the experiment revealed no obvious relations between the task performance measures and medication, food, and/or fluid intake. The logbook entries for all 4 crewmembers, however, did reflect a consistent reduction in reported sleep time during the inflight phase of the study. By comparison with the preflight logbook entries, there was approximately a 15% reduction in sleep time inflight, though no crewmember reported less than 5 hr of sleep on any daily logbook entry.

DISCUSSION

The primary objective of this study was to examine the feasibility of using a behavior analytic approach for assessing crewmember performance during spaceflight. Crewmembers were trained to complete computerized tasks and self-report questionnaires in a single session, and stable measures of performance were established with a limited number of training sessions. In addition, both the rate and the accuracy of performance were sustained or enhanced across the study. The only programmed consequences associated with task performance in this experiment were the computer screen displays of point totals during each task and at the end of a session. It is likely that sustained accurate and high performance rates also reflect rule-governed behavior (e.g., Hayes, 1989) engendered during mission training and participation in the broader NASA flight program. In debriefing

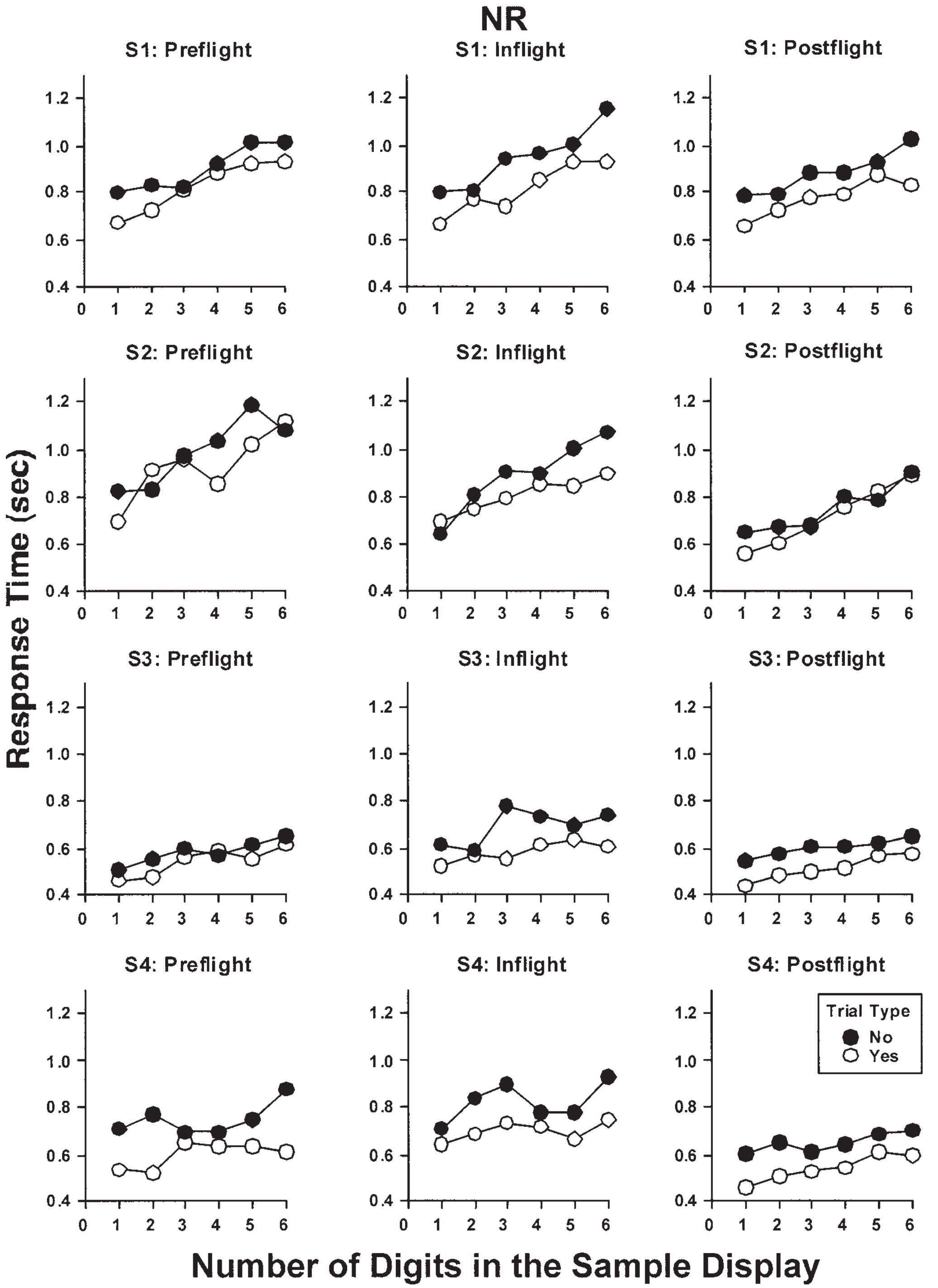

Fig. 4. Mean response times on Yes and No trials during the number recognition task as a function of the number of digits contained in the sample display during the preflight, inflight, and postflight phases of the study for each of the 4 NASA flight crewmembers.

after the study, crewmembers also described positive verbal interactions and friendly competition regarding point totals throughout the course of the study. Although the specific controlling variables remain obscure, task performance and other behaviors were maintained over many months, and this study demonstrated the feasibility of using behavior analytic approaches for assessing performance during spaceflight.

The sensitivity and predictive utility of using performance measures to detect or predict behavioral changes of relevance to the successful completion of the mission also was evaluated in this study. The sensitivity of the Fatigue and Arousal self-report ratings on both the POMS and the VAS was confirmed, with postflight levels being significantly different from levels during either preflight or inflight phases. These findings suggest that the ratings may be influenced by similar controlling variables because there was a high inverse correlation between POMS Arousal and Fatigue scores for each crewmember (i.e., −0.72, −0.79, −0.91, −0.87). No other VAS or POMS ratings changed consistently over the course of the study. Mood scales have been used in previous spaceflight experiments (e.g., Eddy et al., 1998; Manzey et al., 1998; Manzey et al., 2000), and as in the current study, few systematic changes as a function of study phase have been reported. The predictive validity of these self-report ratings was supported to a limited extent by the logbook entries indicating a 15% reduction in inflight sleep, although no decremental effects were apparent in performance of mission activities as a consequence of this minimal reduction in sleep duration (i.e., mission objectives were completed successfully).

The results obtained with the DRL, RA, NR, and DSST performance tasks complement those of the self-report ratings. Both the DRL and RA performances were well maintained throughout all phases of the experiment, and there was no indication that spaceflight had any selective effect on the behaviors involved in either of these tasks. Performance on the DSST and NR tasks also were maintained at high rates of response and accuracy even though subtle changes in performance on the two tasks were observed inflight. Based on the results of this study, however, there was little or no indication of significant behavioral impairment associated with such short-duration spaceflight missions.

Throughout all phases of the experiment, reported food and fluid intake were well maintained, medication usage was infrequent or totally absent, and even the sleep-time reduction reported during inflight did not engender observable decrements in mission performance. Furthermore, despite the austere and threatening context of spaceflight, the crewmembers of this and other short-duration missions have maintained a remarkably consistent, high level of behavioral proficiency during spaceflight. As such, the absence of any marked changes in task performance during the battery was consistent with the uniformly high level of mission performance by crewmembers during short-term spaceflight. To the extent that crewmember performance during the short-term mission, including task performance, was sustained in part by instructional control emerging from extensive training and participation in the NASA flight program, concerns associated with maintaining performance over extended durations associated with space exploration are apparent. Instructional compliance is conditional upon reinforcing (and punishing) consequences and must be actively monitored (e.g., Cerutti, 1989); these conditions are difficult to sustain during extended spaceflight, and deterioration of crewmember performance of mission responsibilities have been reported under these conditions (e.g., Covault, 1988). The predictive utility of the current battery of tests should be tested in longer-duration spaceflight.

A second purpose of this study was to examine the effects of spaceflight on an expanded range of behavioral measures. The task components of the battery engendered a generally high degree of performance stability and precision, but small-magnitude effects of spaceflight on a subset of performance measures were observed. The most reliable and significant changes in performance that occurred during spaceflight were those in the NR task. Relative to Yes trials, response times on No trials increased significantly as a function of the number of digits in the sample stimulus during only the inflight phase of the study. Increases in response times on a modified recognition task involving letters rather than numbers and delayed rather than immediate matching-array presentations have been reported (Manzey et al., 1998; Newman & Lathan, 1999), but selective changes in response time as a function of stimulus conditions during spaceflight have not been described (e.g., Eddy et al., 1998; Manzey et al., 1993, 1995). Also, during the DSST, small but consistent decreases in total and correct trial rates were observed during the inflight phase relative to the preflight and postflight phases for each of the crewmembers (Figure 3), although this effect did not reach statistical significance.

Changes in performance during spaceflight have not consistently been reported in previous studies (cf. Newman & Lathan, 1999), so it is useful to consider why such changes were detected in this study. It seems unlikely that the differential sensitivity of these tasks was a function of performance topography or stimulus presentation conditions because the task procedures used in this study (i.e., numbers and letters presented on a computer screen, button presses) were not substantially different from those used in previous spaceflight studies. Specific contingencies do vary across tasks and studies, however, and it is possible that the effects of spaceflight were specific to behaviors associated with NR and DSST task performance. It also is possible that feedback conditions could have an important influence on such performance effects. In contrast to previous studies where feedback was absent (Benke et al., 1993; Eddy et al., 1998) or present only during initial training (Manzey et al., 1995; Ratino et al., 1988), the present study included feedback throughout all phases of the experiment. This procedural variation could be responsible for enhancing the intersession stability required to detect small magnitude performance changes in the present repeated-measures experimental design (Johnston & Pennypacker, 1993).

In addition to microgravity conditions, the spacecraft environment and mission requirements during spaceflight involve a range of factors that could affect human behavior adversely (e.g., Manzey & Lorenz, 1998). Determining the effects of vibration, ambient noise and alarms, frequent light-dark cycling, vestibular-induced motion sickness in a setting characterized by limited oxygen, water and food resources, as well as demanding communication and work requirements that include both tightly scheduled activities and mission emergencies was well beyond the scope of the present study. The demonstrated stability and maintained sensitivity of the measures that were used over extended preflight, postflight, and in-flight time intervals, however, confirmed their effectiveness in detecting even small-magnitude behavioral changes that may provide ‘‘early warning signs’’ prior to impairment of spaceflight duty performance that could impact mission safety and integrity. Early detection of impairment could be important for initiating countermeasures to minimize or prevent social and behavioral disruption among crewmembers. Continued monitoring of these behavioral changes also might be useful for evaluating the effectiveness of the countermeasures. The continued development of such technological approaches to the assessment of behavioral integrity for individuals and groups living under extreme environmental conditions is essential not only to ensure the success of extended spaceflight missions, but also to enhance safety and the quality of life in many applied settings (e.g., Kelly, Taylor, Heishman, & Crouch, 1998).

REFERENCES

Benke, T., Koserenko, O., Watson, N. V., & Gerstenbrand, F. (1993). Space and cognition: The measurement of behavioral functions during a 6-day space mission. *Aviation, Space, and Environmental Medicine, 64*, 376–379.

Bioastronautics roadmap. (n.d.) From http://criticalpath.jsc.nasa.gov

Boren, J. J., & Devine, D. D. (1968). The repeated acquisition of behavioral chains. *Journal of the Experimental Analysis of Behavior, 11*, 651–660.

Brady, J. V. (1990). Toward applied behavior analysis of life aloft. *Behavioral Science, 35*, 11–23.

Brady, J. V. (1993). Behavior analysis applications and interdisciplinary research strategies. *American Psychologist, 48*, 435–440.

Cerutti, D. T. (1989). Discrimination theory of rule-governed behavior. *Journal of the Experimental Analysis of Behavior, 51*, 259–276.

Christensen, J. M., & Talbot, J. M. (1986). A review of the psychological aspects of space flight. *Aviation, Space, and Environmental Medicine, 57*, 203–212.

Covault, C. (1988, January 4). Record Soviet manned space flight raises human endurance questions. *Aviation Week and Space Technology, 128*, 25.

Day, M. K., Allen, D. L., Mohajerani, L., Greenisen, M. C., Roy, R. R., & Edgerton, V. R. (1995). Adaptations of human skeletal muscle fibers to spaceflight. *Journal of Gravitational Physiology, 2*, P47–P50.

Eddy, D. R., Schiflett, S. G., Schlegel, R. E., & Shehab, R. L. (1998). Cognitive performance aboard the life and microgravity spacelab. *Acta Astronautica*, *43*, 193–210.

Fischman, M. W. (1978). Cocaine and amphetamine effects on repeated acquisition in humans. *Federation Proceedings*, *37*, 618.

Foltin, R. W., & Fischman, M. W. (1991). Assessment of abuse liability of stimulant drugs in humans: A methodological survey. *Drug and Alcohol Dependence*, *28*, 3–48.

Foltin, R. W., Fischman, M. W., Brady, J. V., Bernstein, D. J., Capriotti, R. M., Nellis, M. J., & Kelly, T. H. (1990). Motivational effects of smoked marijuana: Behavioral contingencies and low-probability activities. *Journal of the Experimental Analysis of Behavior*, *53*, 5–19.

Frost, J. D., Shumate, W. H., Salamy, J. G., & Booher, C. R. (1976). Sleep monitoring: The second manned Skylab mission. *Aviation, Space, and Environmental Medicine*, *47*, 372–382.

Fry, W., Kelleher, R. T., & Cook, L. (1960). A mathematical index of performance on fixed-interval schedules of reinforcement. *Journal of the Experimental Analysis of Behavior*, *3*, 193–199.

Gerathewohl, S. J., Stallings, H. D., & Strughold, H. (1957). Sensomotor performance during weightlessness; Eye-hand coordination. *Journal of Aviation Medicine*, *28*, 7–12.

Hayes, S. C. (1989). *Rule-governed behavior: Cognition, contingencies and instructional control.* New York: Plenum Press.

Johnston, J. M., & Pennypacker, H. S. (1993). *Strategies and tactics of behavioral research: Second edition.* Hillsdale, NJ: Erlbaum.

Kanas, N. (1987). Psychological and interpersonal issues in space. *American Journal of Psychiatry*, *144*, 703–709.

Kelly, T. H., Foltin, R. W., & Fischman, M. W. (1993). Effects of smoked marijuana on heart rate, drug ratings and task performance by humans. *Behavioural Pharmacology*, *4*, 167–178.

Kelly, T. H., Foltin, R. W., Rolls, B. J., & Fischman, M. W. (1994). Effect of meal macronutrient and energy content on human performance. *Appetite*, *23*, 97–111.

Kelly, T. H., Foltin, R. W., Serpick, E., & Fischman, M. W. (1997). Behavioral effects of alprazolam in humans. *Behavioural Pharmacology*, *8*, 47–57.

Kelly, T. H., Taylor, R. C., Heishman, S. J., & Crouch, D. J. (1998). Performance measures of behavioral impairment in applied settings. In S. B. Karch (Ed.), *Handbook on drug abuse* (pp. 235–265). Boca Raton, FL: CRC Press.

Leone, G., Lipshits, M., Gurfinkel, V., & Berthoz, A. (1995a). Influence of graviceptives cues at different level of visual information processing: The effect of prolonged weightlessness. *Acta Astronautica*, *36*, 743–751.

Leone, G., Lipshits, M., Gurfinkel, V., & Berthoz, A. (1995b). Is there an effect of weightlessness on mental rotation of three-dimensional objects? *Cognitive Brain Research*, *2*, 255–267.

Leonov, A. A., & Lebedev, V. I. (1973). Psychological characteristics of the activity of cosmonauts (Role of cosmonaut in man-spacecraft system, and psychological aspects in training cosmonauts as operators). *NASA Technical Translation NASA-TT-F-727*, NASA: Washington, DC.

Lindsley, D. B. (1972). *Human factors in long duration spaceflight*, National Academy of Sciences: Washington, DC.

Manzey, D., & Lorenz, B. (1998). Mental performance during short-term and long-term spaceflight. *Brain Research Reviews*, *28*, 215–221.

Manzey, D., Lorenz, B., Heuer, H., & Sangals, J. (2000). Impairments of manual tracking performance during spaceflight: More converging evidence from a 20-day space mission. *Ergonomics*, *43*, 589–609.

Manzey, D., Lorenz, B., & Poljakov, V. (1998). Mental performance in extreme environments: Results from a performance monitoring study during a 438-day spaceflight. *Ergonomics*, *41*, 537–559.

Manzey, D., Lorenz, B., Schiewe, A., Finell, G., & Thiele, G. (1993). Behavioral aspects of human adaptation to space: Analyses of cognitive and psychomotor performance in space during an 8-day space mission. *Clinical Investigator*, *71*, 725–731.

Manzey, D., Lorenz, B., Schiewe, A., Finell, G., & Thiele, G. (1995). Dual-task performance in space: Results from a single-case study during a short-term space mission. *Human Factors*, *37*, 667–681.

Matsakis, Y., Lipshits, M., Gurfinkel, V., & Berthoz, A. (1993). Effects of prolonged weightlessness on mental rotation of three-dimensional objects. *Experimental Brain Research*, *94*, 152–162.

McLeod, D. R., Griffiths, R. R., Bigelow, G. E., & Yingling, J. (1982). An automated version of the digit-symbol substitution test (DSST). *Behavior Research Methods and Instrumentation*, *14*, 463–466.

McNair, D. M., Lorr, M., & Droppleman, L. F. (1971). *Profile of Mood States (Manual).* San Diego, CA: Educational and Industrial Testing Service.

Monk, T. H., Buysse, D. J., Billy, B. D., Kennedy, K. S., & Willrich, L. M. (1998). Sleep and circadian rhythms in four orbiting astronauts. *Journal of Biological Rhythms*, *13*, 188–201.

Newman, D. J., & Lathan, C. E. (1999). Memory processes and motor control in extreme environments. *IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews*, *29*, 387–394.

Ratino, D. A., Repperger, D. W., Goodyear, C., Potor, G., & Rodriguez, L. E. (1988). Quantification of reaction time and time perception during space shuttle operations. *Aviation, Space, and Environmental Medicine*, *59*, 220–224.

Ross, H. E., Schwartz, E., & Emmerson, P. (1987). The nature of sensorimotor adaptation to altered g-levels: Evidence from mass discrimination. *Aviation, Space, and Environmental Medicine*, *58*, A148–A152.

Sangals, J., Heuer, H., Manzey, D., & Lorenz, B. (1999). Changed visuomotor transformations during and after prolonged microgravity. *Experimental Brain Research*, *129*, 378–390.

Sternberg, S. (1966, August 5). High-speed scanning in human memory. *Science*, *153*, 652–654.

Stuster, J. (1996). *Bold endeavors: Lessons from polar and space exploration.* Annapolis, MD: Naval Institute Press.

Taylor, A. J. W. (1989). Behavioural science and outer space research. *Aviation, Space, and Environmental Medicine*, *60*, 815–816.

Thornton, W. E., Moore, T. P., Pool, S. L., & Vanderploeg, J. (1987). Clincial characterization and etiology of space motion sickness. *Aviation, Space, and Environmental Medicine*, *58*, A1–A8.

Ward, A. S., Kelly, T. H., Foltin, R. W., & Fischman, M. W. (1997). Effects of d-amphetamine on task performance and social behavior of humans in a residential laboratory. *Experimental and Clinical Psychopharmacology*, *5*, 130–136.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale - Third Edition*. San Antonio, Texas: The Psychological Corporation.

West, J. B. (1984). Spacelab - The coming of age of space physiology research. *Journal of Applied Physiology*, *57*, 1625–1631.

Received August 17, 2004
Final acceptance June 9, 2005