1996

Task switching efficiencies within and across flight related functional categories

Allen Goodman
San Jose State University

TASK SWITCHING EFFICIENCIES WITHIN AND ACROSS FLIGHT RELATED FUNCTIONAL CATEGORIES

A Thesis

Presented to

the Faculty of the Department of Psychology

San Jose State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

by

Allen Goodman

August, 1996

UMI Number: 1392810

APPROVED BY THE DEPARTMENT OF PSYCHOLOGY

(Dr. Kevin Jordan, chairperson)

(Robert J. Shively, committee member)

(Dr. Laree Huntsman, committee member)

APPROVED FOR THE UNIVERSITY

ABSTRACT

TASK SWITCHING EFFICIENCIES WITHIN AND ACROSS FLIGHT RELATED FUNCTIONAL CATEGORIES

by Allen D. Goodman

In a recent analysis of future flight deck requirements, Abbott and Rogers (1993) assert that more effort and attention are required of pilots to change tasks across functional categories than within functional categories. Implications of this assertion go beyond the problems of increased processing demands to the dangers of subjecting flight crews to a series of "cognitively disjointed" tasks eventually leading to a loss of overall situational understanding.

A low-fidelity part-task simulation using general aviation pilots was used to evaluate performance for flight control, communications, and system's management tasks following specific switching conditions. It was found that functional continuity and inter-task interval play a role in task completion times and error rates with differing effects depending on task type. The predictive value of previously displayed information is suggested as a mediating variable which is positively correlated to task switching performance and would account for such differences.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Task Switching Efficiencies Within and Across Flight Related Functional Categories

Allen D. Goodman
San Jose State University

Running head: TASK SWITCHING EFFICIENCIES

Footnotes

Request for reprints should be sent to Allen D. Goodman, Department of Psychology, San Jose State University, San Jose, California 95192

Abstract

In a recent analysis of future flight deck requirements, Abbott and Rogers (1993) assert that more effort and attention are required of pilots to change tasks across functional categories than within functional categories. Implications of this assertion go beyond the problems of increased processing demands to the dangers of subjecting flight crews to a series of "cognitively disjointed" tasks eventually leading to a loss of overall situational understanding. A low-fidelity part-task simulation using general aviation pilots was used to evaluate performance for flight control, communications, and system's management tasks following specific switching conditions. It was found that functional continuity and inter-task interval play a role in task completion times and error rates with differing effects depending on task type. The predictive value of previously displayed information is suggested as a mediating variable which is positively correlated to task switching performance and would account for such differences.

Task Switching Efficiencies Within and Across Flight Related Functional Categories

A guiding principle of human factors design is to minimize the demands placed on the operator while achieving effective system performance. This goal is especially important in dynamic environments such as the modern flight deck where pilots routinely face intervals of intense, multiple task demands. At times, the effort and attention needed to cope with these fast paced streams of interrelated activity can exceed the capacity of the pilot and lead to degraded performance and increased error. Wiener (1985), for example, has pointed out that two-thirds of commercial transport accidents occur during terminal area operations (i.e., taxi, take-off, and landing) when task demands are generally agreed to be at their peak. This is despite the fact that these phases account for only six percent of total flight time. Such breakdowns in performance are even more likely in non-normal or emergency situations when pilots are additionally burdened with diagnostic and assessment activities. As noted in an analysis by Nagel (1988), pilots in such circumstances are particularly vulnerable to errors of misinterpretation (e.g., confusion over flight director modes during a missed approach) and errors of oversight (e.g., failure to monitor altitude while trouble shooting a malfunction).

Within aviation, as in other process control domains, such consequences are often referred to as "human error" when they may reflect systems and procedures which prove overly demanding under certain operational contingencies. From a design perspective, these are systems and procedures which do not fully take into account the processing limitations of the operator. The observed deterioration of pilot performance associated with periods of intense activity is an example of operational demands in conflict with human processing limitations. It follows that if such conflicts can be minimized, pilots will be better able to move with ease and efficiency through their tasks, avoid errors, and maintain reserve capacity for emergency situations.

Towards this objective, the current study was initiated to investigate a substantially ignored factor which recently has been speculated (Abbott & Rogers, 1993) to impact pilot performance: namely, the functional continuity between tasks within the sequencing of flight deck activity. At the heart of this research is the common sense notion that when confronted with different types of tasks, it is usually more efficient to group and perform similar tasks together rather than switch back and forth between task types. A now classic confirmation of this phenomenon was provided by Jersild in 1927 (as cited in Garcia-Ogueta, 1993) who compared average performance on a simple addition task (A) and a simple subtraction task (B) when presented in alternating trial blocks (i.e., ABABAB . . .), with performance in pure trial blocks (i.e., AAAA . . . or BBBB. . .). Jersild found task completion times an average of 620 ms longer in the alternating condition which he attributed to the increased "costs" associated with switching between different task types.

In much the same vein, this research focused on detecting differences in the speed and accuracy by which flight related tasks could be performed dependent on the functional similarity or dissimilarity of the immediately preceding activity. It was predicted, for example, that equivalent navigation tasks would be performed faster and/or with fewer errors when preceded by other navigation tasks than when preceded by communications or system's management tasks. As with Jersild's work, it was believed that such differences in task performance would reflect differences in task switching costs. Knowledge concerning the effects of functional continuity on task switching efficiencies might provide useful insights into the underlying causes of performance deterioration during intense multi-task conditions.

Multi-Task Performance

Many human performance limitations do not concern difficulties associated with performing a single task but, rather, the constraints on a person's ability to meet multiple, possibly concurrent, task demands. As a result, a great deal of research has been directed toward those behaviors which are specific to dealing with multi-task situations: time-sharing, task selection and scheduling, and task switching. Each of these processing activities imposes separate, but not mutually exclusive demands which cumulatively impact operator performance as described below.

Time-Sharing. By far, the most studied aspect of multi-task performance concerns the mechanisms and limitations which are operative when a person performs two or more tasks concurrently. Fundamental to most models of time-shared performance is the concept of limited resources. This hypothetical construct suggests differentiated pools of processing capacity available to the operator across specific dimensions.

Broadbent (1958) first formalized some of these ideas in a model of human operators as limited-capacity, single-channel information processors. However, this conceptualization did not recognize that operators can do more than one thing at a time; for example, people usually have little difficulty in carrying on a conversation while driving a car. Accounting for such instances of concurrent processing while still maintaining notions of limited capacity, Navon and Gopher (1979) proposed a multiple resource theory. As refined by

Wickens (1980; 1984), this theory assumes limited amounts of available resources within each of three processing dimensions: modalities (auditory versus visual perception); codes (verbal versus spatial); and stages (perceptual and cognitive versus response). Accordingly, tasks can be effectively time-shared only to the extent that they do not compete for the same resources within a dimension. This explanation has served as a good predictor of dual-task performance in numerous laboratory studies (see Damos, 1993). Likewise, it is consistent with everyday experience. People who converse so easily while driving would have great difficulty carrying on a conversation while reading a technical report, i. e., time-sharing two verbally coded activities.

The multiple resource model suggests that increased similarity between tasks along the three dichotomous dimensions of the model will increase task interference. However, there is also evidence that similarity between more elemental or structural aspects of tasks can impact time-shared performance. Some forms of this phenomenon, collectively known as similarity effects, can facilitate time-shared performance while other forms can have the opposite effect. When tasks share common processing subroutines or timing mechanisms performance seems to benefit. For example, Chernikoff, Duey, and Taylor (1960) found significant timesharing efficiencies when two tracking tasks shared identical control dynamics. Similarly, Klapp (1979) demonstrated superior time-sharing performance of two rhythmic activities when the rhythms were the same, rather than different. Conversely, performance is likely to suffer when semantic or representational elements of one task become confused in the processing of another task because of their similarity. Navon (1974) demonstrated that listeners would mistakenly attribute the words of one speaker to those of another speaker in a dichotomous listening task when voice quality and message were quite similar. This effect would all but disappear when the two speakers' physical and semantic characteristics differed. Hirst and Kalmar (1987) reported that timesharing between a spelling and arithmetic task was easier than timesharing between two spelling or two arithmetic tasks.

It should be noted that not all multi-task performance is time-shared. Research has shown (Adams, Tenney, & Pew, 1991) that operators in dynamic environments employ a mix of both concurrent task processing and single-task, serial execution in managing their task flows. Moreover, the distinction between serial versus time-shared behavior may be somewhat blurred. There is no question that people often work on more than one task at a time. This does not necessarily mean the cognitive components of such tasks are being processed in parallel. It could be the case, as argued by Schweikert and Boggs (1984), that focused attention is modular and that time-sharing represents a specialized form of serial processing involving rapid switching between interleaved subroutines of multiple in-progress tasks. This would suggest that factors which can be linked to serial task performance (such as functional continuity) might well extend to time-shared performance.

Task Selection and Scheduling. Within multi-task situations, deciding which tasks to initiate and when to initiate them is a necessary prerequisite to overt action. Factors contributing to the demands of this planning and decision making function include the number of tasks requiring attention, time constraints, level of situational ambiguity, clarity of goals, and criticality of responses. Researchers in this area have traditionally sought to develop human performance models which mathematically describe optimal selection and scheduling strategies for a given set of conditions (e.g., Kleinman & Curry, 1977). A more recent perspective identified as "strategic task management" has been espoused by Hart (1989) and Adams et al. (1991) which posits a real world view of operators who actively manage their time, energy, and available resources to achieve adequate performance while maintaining comfortable levels of task load.

Task Switching. A second intrinsic feature of multi-task performance, and one most pertinent to this study, is the continual need for operators to switch from a current task (usually when completed) to a new task. This transitioning must occur irrespective of whether tasks are processed individually in sequence or concurrently in a time-shared manner. This research focused on factors which might be associated with the costs of such transitions. These costs can be characterized as the time and effort required to disengage from a current task, reorient to a new task, and to engage that task. An underlying premise of the present research is that all task shifts involve time and effort, but that these costs increase as tasks become farther apart in terms of certain categorical or global properties. In this study, the categorization of tasks by their highest-order functional purpose is evaluated as one such differentiating property.

Given that operator activities (other than completely reflexive, automatic tasks) require some attentional focus, then support for the above premise can be inferred from the literature regarding attentional shifts. For example, modality switching asserts that it takes longer to switch attention between two discrete stimuli across modalities than within modalities (Laverg, VanGelder, & Yellott, 1971). Similarly, studies in discourse processing (Bower, 1982) have shown greater time and effort are required for mental shifts between, rather than within, topics or semantic domains. Other support comes from studies of "interrupted task" performance (viewed here as an index of switching efficiency from primary task to interruption) which show that interruptions are handled more easily to the extent that they are similar in goal to the task they displace (Gillie & Broadbent, 1989).

Rationale for the Present Study

The impetus for the present study was provided in a NASA technical report (Abbott & Rogers, 1993) which presented a detailed taxonomy of flight deck functions for a generic, advanced civil aircraft. The authors defined functions as the category descriptors which most economically parse flight activities at a given level by their intended goal. For example, flight management, communications management, systems management, and task management were described as the highest-order flight deck functions sufficient to support the overall mission goal of safely and efficiently moving passengers and cargo from point of departure to destination. Interestingly, in the discussion section of this report, the authors caution that, " . . . more effort and attention are required to change tasks across functional categories than within functional categories" (p. 12). The implications of this assertion, they note, go beyond the adverse consequences of increased task demands to the dangers of subjecting flight crews to a series of "cognitively disjointed" tasks eventually leading to a loss of overall situational understanding.

Despite its common sense appeal, empirical support within aviation research for the increased costs associated with task switching across functional categories is unavailable. Yet, there are reasons for believing that such effects might be especially pronounced within the flight deck. In this highly evolved procedural domain, pilots have been trained to think in terms of shared functional hierarchies as in the classic "aviate, navigate, and communicate." With pilots encouraged to categorize their activities into groupings of cognitively proximal elements (functions), switching between functionally different tasks should require greater attention displacement and, consequently, increased time and effort.

Abbott and Rogers' (1993) functional taxonomy would seem to provide a logical approximation for pilots' internal organization of flight activities. This view is substantially supported in studies by Adams and Pew (1990) and Jonsson and Ricks (1993) which utilized multi-dimensional scaling and cluster analysis of semantic sorting and ranking tasks to more veridically capture pilots' categorization schema (see Table 1). From these findings the following three flight related functional categories were chosen for inclusion in this study: (1) flight control, (2) navigation and reference, and (3) systems management. It was determined that these particular top-level functions were best represented by the available simulation hardware.

Experimental Approach

The primary goal of the present study was to evaluate the effects of functional continuity on task switching efficiencies in a flight related environment. It was hypothesized that experienced pilots in a low fidelity simulation would exhibit faster, more accurate performance on a range of discrete tasks when the immediately preceding activity involved tasks from the same rather than different functional category. To limit the influence of other factors which might also effect task switching efficiencies, it was determined that the testing protocol should incorporate these features:

1. A within-subjects design (to minimize individual differences)
2. Test conditions set exclusively within context of non-overlapping,

Table 1

Top-Level Pilot Functional Categories as Noted in Recent Studies

Study	Functional Categories
Abbot & Rogers (1993)	1. Flight management 2. Communications Management 3. Systems Management 4. Task Management
Adams & Pew (1990)	1. Local navigation, guidance & control of aircraft 2. Macro-planning & navigation (route planning /replanning) 3. Remote communications 4. Flight crew resource management 5 Cabin management 6. Management of physical equipment, resources, & systems 7. Management of flight management computer 8. Bridging activities
Jonsson & Ricks (1993)	1. Flight control 2. Navigation & reference 3. Systems 4. Communications 5. Emergency

sequential task performance (to control for time-shared phenomena such as resource competition and similarity effects)

3. Tasking sequences scripted by experimenter (to ensure that specific switching conditions are tested and to eliminate demands on pilots for selecting, scheduling, and prioritizing tasks)
4. Maximum equivalency of tasks (to limit performance differences resulting from variability in task difficulty)

Additionally, the study manipulated two aspects of the preceding task condition which were hypothesized to play a mediating role in functional continuity effects during task switching. The first of these factors was the temporal proximity of preceding activity. It was believed that the effects of functional continuity would attenuate as the time interval between preceding activity and the initiation of a new task increased. Such results would reflect those found in semantic priming studies (a somewhat analogous experimental paradigm) in which both facilitory and inhibitory effects rapidly deteriorate as the delay interval between priming stimulus and target increases (Neely, 1977).

A final manipulation would examine whether the effects of functional continuity vary with the intensity of preceding attentional focus. It was believed that performing a series of like-function tasks, relative to a single task, should invoke a deeper level of processing or attentional focus within a given functional area. As a preceding condition, this deeper level of focus was predicted to increase both the facilitation for within-function transitions and the amount of interference for between-function transitions. In part, the performance gains for pure trial blocks in Jersild's (1927) study might reflect the added facilitation of consistent same-task repetitions.

Method

Participants

Twelve right handed male general aviation pilots were recruited and paid to participate in this study. They ranged in age from 20 to 49, with a median age of 34.5. Individual flight time ranged from 100 hours to 3,000 hours with the median for the group at approximately 500 hours. Four of the twelve pilots were multi-engine rated; three of these four held instructor certificates.

Equipment and Apparatus

Data were collected utilizing the Workload/PerformANce Simulation (Window/PANES) system developed at NASA-Ames Research Center. The software portion of this system runs on an IBM compatible PC and configures a 13 inch color monitor into four continuously displayed, experimenter controlled windows (depicted in Figure 1) as follows:

1. A text message window
2. An interactive tracking window with numeric readouts of ownship speed, altitude, and heading
3. A systems window with gauges
4. A static North-up map display representing a 25 x 40 mile view on which ownship movements are tracked

The system includes a custom response box (see Figure 2) containing a two-axis joy-stick (heading and altitude control), a slide potentiometer speed controller, a row of four buttons utilized in four-choice responses, and a separate pair of buttons for "yes" and "no" responses. During simulation runs, all button responses are recorded and time stamped while flight control tasks (i.e., those tasks using the slide potentiometer or joy stick) are recorded

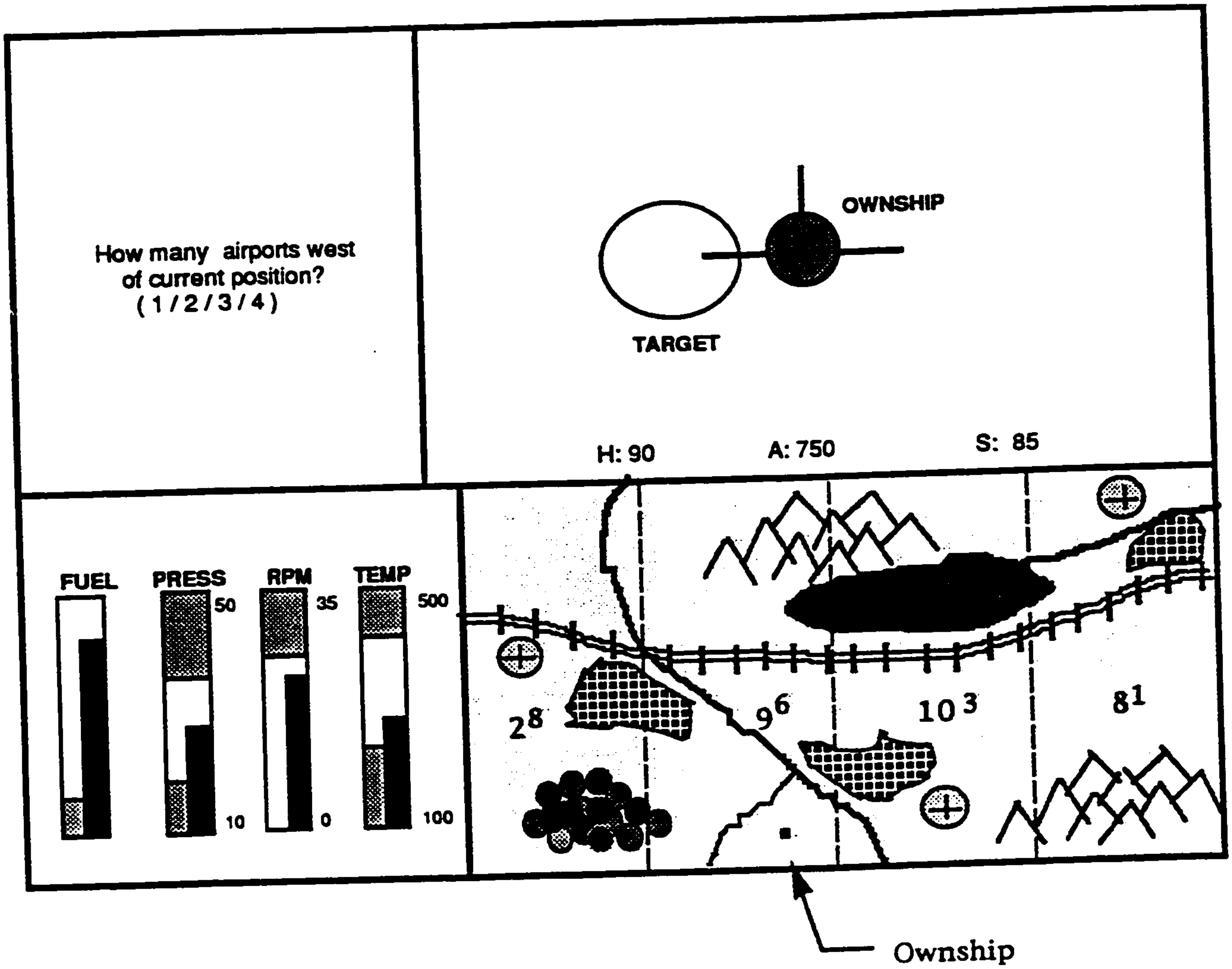


Figure 1. Gray scale depiction of Window/PANES display as configured for present study. Note that map layout differed for each testing block during data collection.

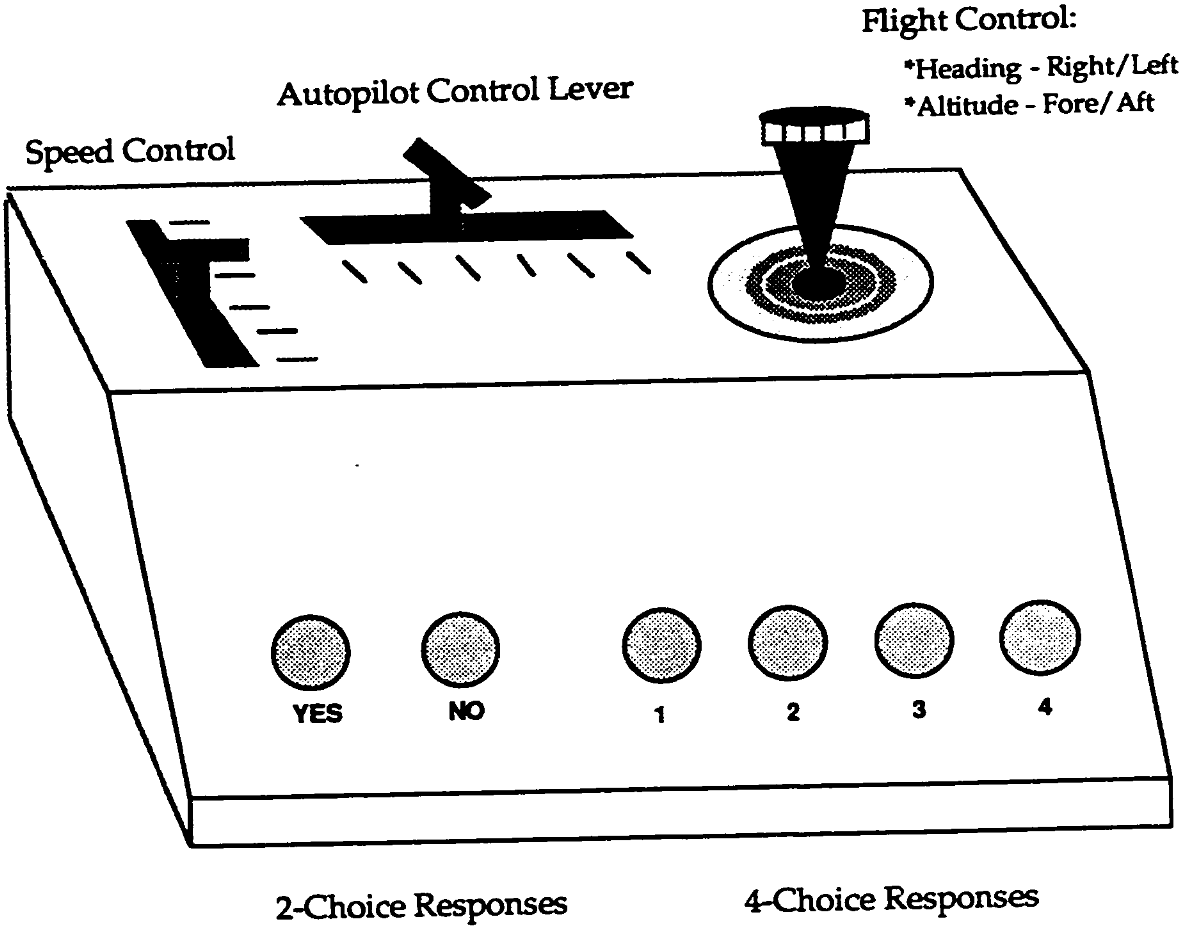


Figure 2. Depiction of response box used by pilots during experimental trials. Note that the flight and speed controllers were used exclusively for executing discrete, commanded changes in heading, altitude, or speed, and not continuous tracking movements.

as ownship's deviations from commanded speed, altitude, or heading at a 10 Hz sampling rate. Together, the Window/PANES hardware and software provides a low fidelity simulation of a flying environment in which discrete and continuous task performance can be measured.

Tasks

A total of 96 short commands and queries were created to serve as a pool of pseudo air traffic control (ATC) data-linked messages which could be programmed to appear in the text window during a Window/PANES simulation. These messages were designed to elicit pilot activity which could be categorized into one of three basic flight functions (see Table 2). One third of the messages were commanded changes of speed, heading, and altitude which required monitoring of the tracking window while making joystick or slide potentiometer adjustments. Such actions are representative of routine flight control tasks. Another third of the messages were queries concerning the number and location of certain ground features (e.g., cities, airports, highways) or the location of such features in relation to ownship's current or projected position. These messages required pilots to evaluate the map display and then make an appropriate button response. Such activities are characteristic of navigation/reference tasks common to general aviation. Lastly, there were queries regarding the status of engine parameters which required pilots to scan the systems window and determine the current reading, the relationship among readings, or the directional trend of one or more of the four vertical gauges representing fuel, rpm, manifold pressure, and engine temperature. Pilot assessments of this sort are typical of normal system management tasks.

All tasks shared the same basic structure: a short text message cued a

Table 2

Task Function Categories with Sample Messages

Flight Control	Navigation/Reference	Systems Management
"Turn left heading 080"	"How many airports in map view ? (1 / 2 / 3 / 4)"	"Which gauge is rising? (1 / 2 / 3 / 4)"
"Climb to 700 feet"	"In which map quadrant is fork in the river? (1/ 2 /3 / 4)"	"Are PRESS/RPM running in sync? (Y or N)"
"Increase power to 75%"	"Will current heading cross rail tracks? (Y or N)"	"How many gauges in the red zone? (1 / 2 / 3 / 4)"
"Turn right heading 210"	"How many quadrants above 5,000 feet? (1 / 2 /3 / 4)"	"Is TEMP gauge falling? (Y or N)"

simple visual assessment of a display which, in turn, determined a discrete psychomotor response. Assessments involving the map and gauge display windows were carefully constructed so as to have only one clear and unambiguous correct response. To verify the suitability of candidate tasks, a pilot study using two general aviation pilots (who did not participate in the experimental trials) was conducted. Only those tasks which were consistently and accurately performed in under 4 s were included in the pool of 96 experimental tasks.

Within this study, tasks were used in two ways. A single task or a cluster of three tasks (always from the same functional domain) could serve as an immediately preceding condition. Alternately, an individual task could serve as the behavior of interest or "target task" which follows such preceding conditions. It was important for the validity of this study that target tasks within each of the three functional categories be as equivalent as possible. Special care was taken, therefore, in ensuring that the message prompts (cueing stimuli) were of the same length and phrasing. Additionally, the salience of needed display information and the required processing were matched in complexity. Lastly, the response outputs were identical across task types. In the case of flight control, for example, only heading changes of 10° were used as target tasks. All other heading changes, as well as speed and altitude commands, were used as preceding tasks. For navigation/reference and system's management tasks only 4-choice summation queries of the form "How many map quadrants are above 5,000 feet?" or "How many gauges are within operational limits?" were used as target tasks. Other 4-choice queries invoking self-terminating searches such as "Which map quadrant ... " or "Which gauge....", as well as all 2-choice queries were used exclusively as preceding tasks. Consequently, adjusting the slide potentiometer speed controller was never compared to joy stick movements. Nor were 2-choice responses compared to 4-choice responses.

Pilot performance during experimental trials was measured in terms of speed and accuracy of task execution. For navigation/reference and system management tasks, completion times were measured in milliseconds from the onset of a cueing message (which was always signaled by a brief tone) to the initiation of a button response. In the case of flight control tasks, it was recognized that both the slide potentiometer and the joy-stick are adjustable control mechanisms with no "hard stops." Throughout this study, however, these controls were used to make specific, stepped changes to speed, altitude, and heading, not continuous tracking movements. Consequently, the flight control tasks were discrete, though they did necessitate a different performance metric. For these tasks, the root mean square error (RMSe) was calculated for deviations from commanded heading, altitude, or speed over a four second interval beginning with the commanded change. This provided a measure which was sensitive to both response latency and control accuracy.

Experimental Design

To investigate the performance consequences of different switching conditions a 2 (functional continuity: shifts within or between domains) X 2 (temporal proximity: 4 s or 8 s inter-task-intervals) X 2 (attentional focus: 1 or 3 preceding tasks) within-subjects factorial design was implemented (see Figure 3). Completion times and error rates for target tasks within navigation/reference ana system's management, as well as RMSe for target flight control tasks, were separately compared across the eight conditions of the fully crossed factorial design. Additionally, baseline measurements of

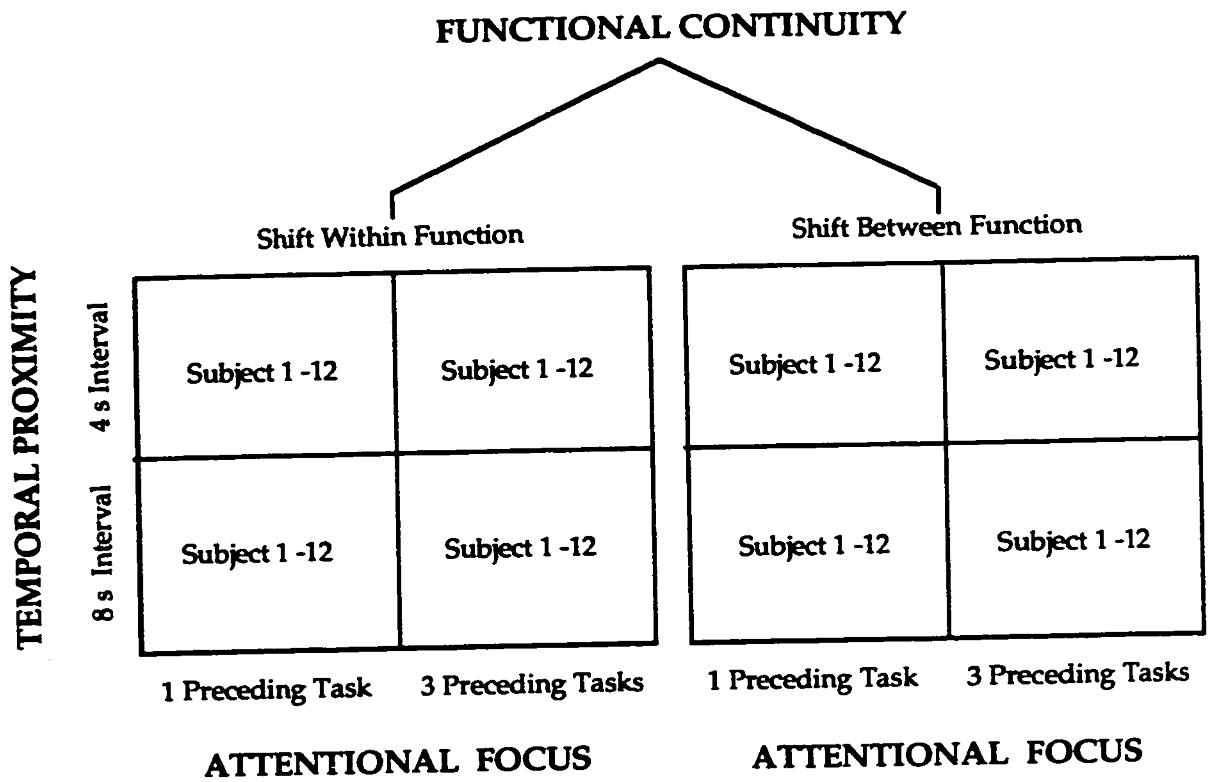


Figure 3. Schematic of within subjects fully crossed factorial design. This design was implemented separately to evaluate completion times and error rates for targeted navigation/reference and systems management tasks and for RMSe on targeted flight control tasks.

target tasks in all three functional domains were taken to determine if any of the experimental conditions led to significantly inferior or superior performance compared to tasks executed in isolation.

Six different scenarios of 6-min length were constructed on the Window/Panes system. Each scenario presented pilots 4 flight control, 4 navigation/reference, and 4 systems management tasks in randomized order following different switching conditions. Additionally, a single task from each functional category was presented in isolation (i.e., following 15 s of total scenario inactivity) to establish baseline performance. All scenarios shared a common timing structure with message prompts on-screen for 4 s and each experimental condition separated by 15 s.

The eight experimental conditions for each of the three task types were evenly distributed between every two scenarios forming three matched-pairs of scenarios or testing blocks. Each testing block provided one full replication of the design, plus two baseline data points per functional category (see Figure 4). The map display, as well as gauge dynamics, remained constant within the two scenarios comprising a given testing block. However, no individual message prompt was repeated within that block.

Testing Procedure

Pilots were brought in individually and told that the Window/Panes system represented a mock-up of a future general aviation cockpit environment featuring both an electronic map and a data-linked text display. During a scenario run pilots were to initially set the auto-pilot and, thereafter, take action only as directed by the data-link messages. Pilots were instructed to respond to messages as quickly as possible without sacrificing accuracy. They were also instructed to attend only to the most current message, i.e., if a new

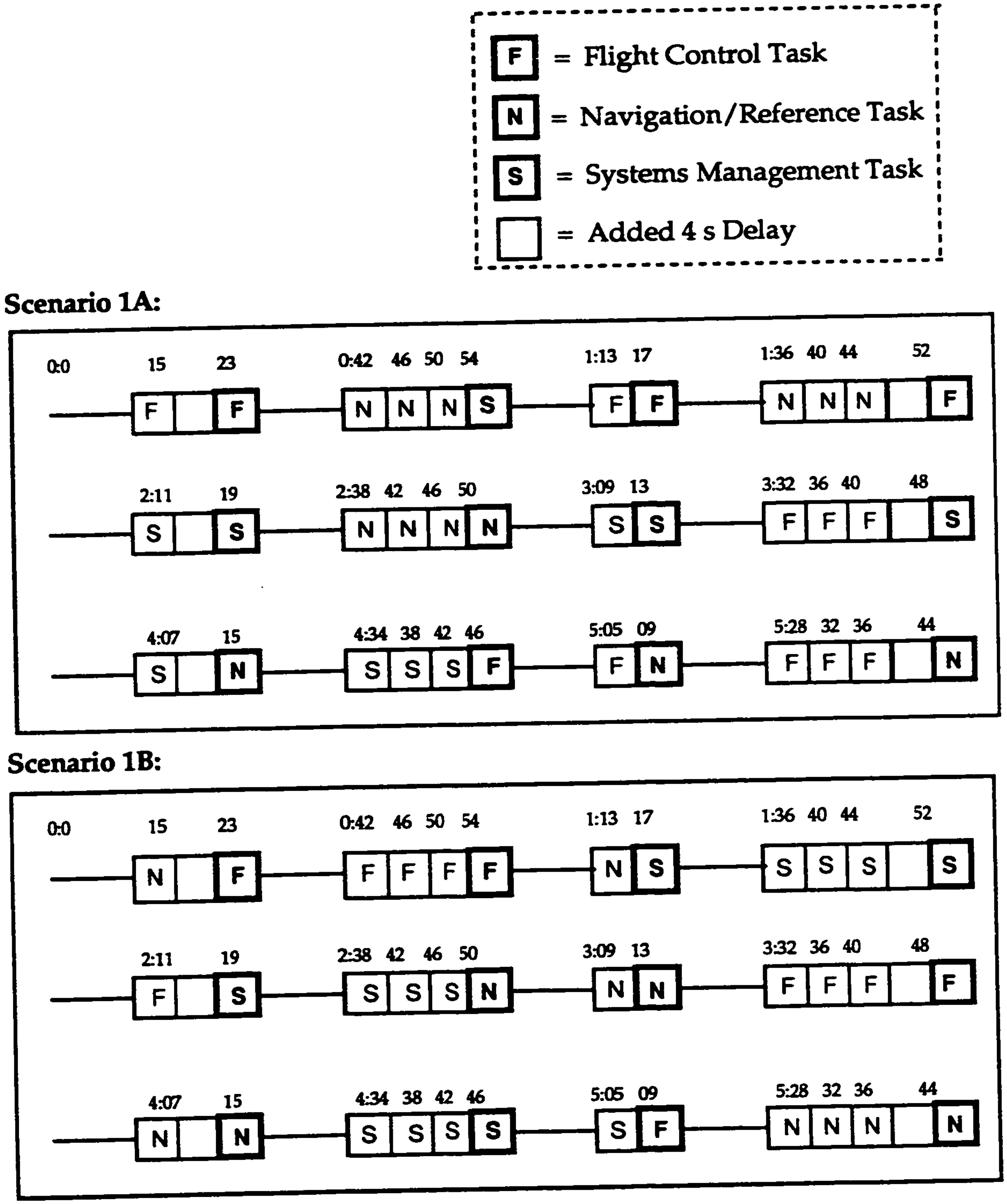


Figure 4. Timeline of Embedded Task Conditions for Two Matched 6-Minute Scenarios Constituting a Testing Block.

message appeared prior to their completing a task, they were to forego the earlier task and concentrate immediately on the new task.

Pilots were then introduced to the Window/PANES system and the task demands of the study through a graduated series of four minute training scenarios which presented flight control, navigation/reference, and systems management tasks in isolation and then in series with a 6 s, then 5 s, and finally, 4 s inter-task-interval. Though messages used in the training scenarios were identical to those in the experimental trials, task responses differed as flight routes, map presentations, and gauge behaviors were unique to the training scenarios.

Upon completion of training participants were given a short break before beginning the first of three testing blocks in which experimental data was collected. Total session time including breaks was approximately three hours.

Results

Data Screening

To insure the sequential nature of assessed task performance, scenario data files were carefully reviewed in order to discard any trial in which the immediately preceding activity (priming task) was not completed. This avoided the possibility of temporal overlap with target tasks and the confounding of processing and completion times. A total of 19 of the 864 experimental trials (8 conditions X 3 task types X 3 repetitions X 12 subjects) were discarded for this reason. The completion times for the remaining priming tasks ranged from 1277 ms to 3980 ms, with a mean duration of 2991 ms. On average, this put the temporal proximity between the completion of a priming task and the onset of a target task at 1009 ms in the short interval condition (4 s inter-task interval) and 5009 ms in the long interval condition

(8 s inter-task interval).

Additionally, 29 trials with incorrect responses in the 4-choice response time tasks were discarded, as well as 3 flight control trials in which the deviation scores indicated no response movement whatsoever.

The analyses presented below are based on the mean scores of pilots across replications for baseline and experimental conditions within each of the three task types.

Flight Control Task

Flight control performance was assessed by calculating the root mean square error (RMSe) in degrees of heading deviations across a four second interval starting with the onset of a commanded change. This metric captured both response latencies and control accuracy with smaller RMSe values indicating better performance. Means and standard deviations of pilot performance in each of the experimental conditions are presented in Table 3.

A 2 (functional continuity) X 2 (temporal proximity) X 2 (attentional focus) within-subjects analysis of variance was conducted on pilots' RMSe scores. There was no main effect of functional continuity for flight control tasks. However, a significant main effect of attentional focus [F (1,11) = 15.11, p = .003] was found. An examination of the pattern of cell means suggested that the locus of this effect was exclusively in the larger control deviations associated with long interval transitions from three other flight control tasks (see Figure 5). This observation was confirmed by a significant three-way interaction of attentional focus by functional continuity by temporal proximity [F (1,11) = 9.7, p = .010].

A corollary concern of this study was to assess whether any of the eight experimental conditions differed significantly from baseline performance.

Table 3

Flight Control Performance: Means and Standard Deviations of Pilots' RMSe Deviation in Degrees.

	Functional Continuity	
	Shift Within-Function	Shift Between-Function
One Preceding Task		
4 s Interval	2.72 (SD = .93)	2.44 (SD = 69)
8 s Interval	2.51 (SD = .68)	2.61 (SD = .69)
Three Preceding Tasks		
4 s Interval	2.60 (SD = .78)	3.10 (SD = .94)
8 s Interval	3.55 (SD = .88)	3.09 (SD = 1.27)

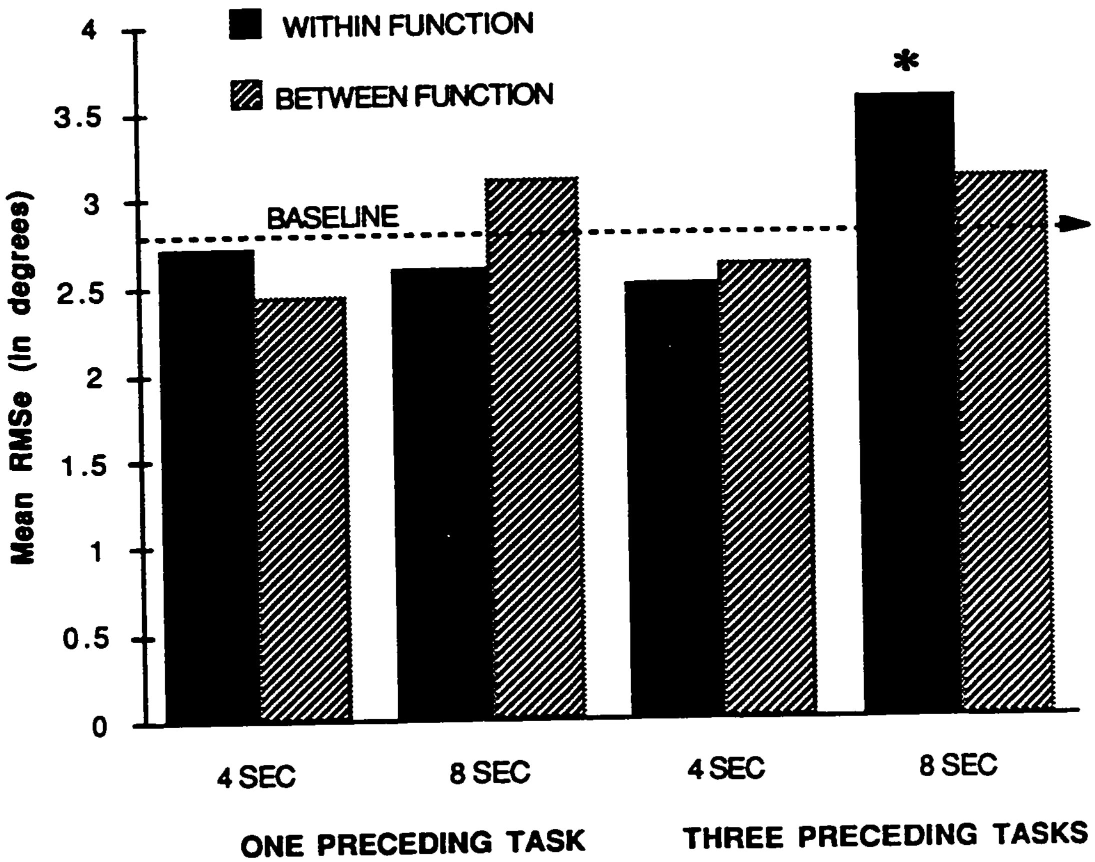


Figure 5. Mean heading deviation for flight control tasks following different experimental switching conditions.

*Condition differs significantly from baseline

Pilots had a mean baseline rate of 2.81 degrees (SD = .46) deviation for heading changes executed in isolation (i.e., following 15 s of scenario inactivity). Complimentary paired *t* -tests indicated that the aforementioned long interval transition from three flight control tasks with a mean deviation of 3.55 degrees was the only condition in which performance was reliably different [t_{paired} (1,11) = 3.75, $p < .01$] from baseline. The adverse effects to performance of this three-way interaction involving task switching from repeated like-function tasks cannot be explained in terms of this studies' hypotheses.

Navigation/Reference Task

Navigation/reference task performance was assessed by response times and error rates to 4-choice queries regarding current positional information or feature locations as displayed in the map window of the simulation. Means and standard deviations of pilot response times to this task are presented in Table 4.

A 2 (functional continuity) X 2 (temporal proximity) X 2 (attentional focus) within-subjects analysis of variance of was performed on the response time data. As hypothesized, a significant main effect of functional continuity [F (1,11) = 34.58, $p < .001$] was found. A review of cell means showed that navigation/reference task response times were consistently greater for between-function transitions (see figure 6). As further hypothesized, a significant two-way interaction between functional continuity and temporal proximity [F (1,11) = 6.13, p =.031] was found. However, rather than attenuating the effects of functional continuity as hypothesized, long interval conditions for navigation/reference tasks seems to amplify the adverse effects

Table 4

Navigation/Reference Performance: Means and Standard Deviations of Pilot Response Times in Milliseconds.

	Functional Continuity	
	Shift Within Function	Shifts Between Function
One Preceding Task		
4 s Interval	3202 (SD = 361)	3458 (SD = 708)
8 s Interval	2858 (SD = 445)	3822 (SD = 720)
Three Preceding Tasks		
4 s Interval	2769 (SD = 416)	3376 (SD = 578)
8 s Interval	2993 (SD = 406)	3589 (SD = 434)

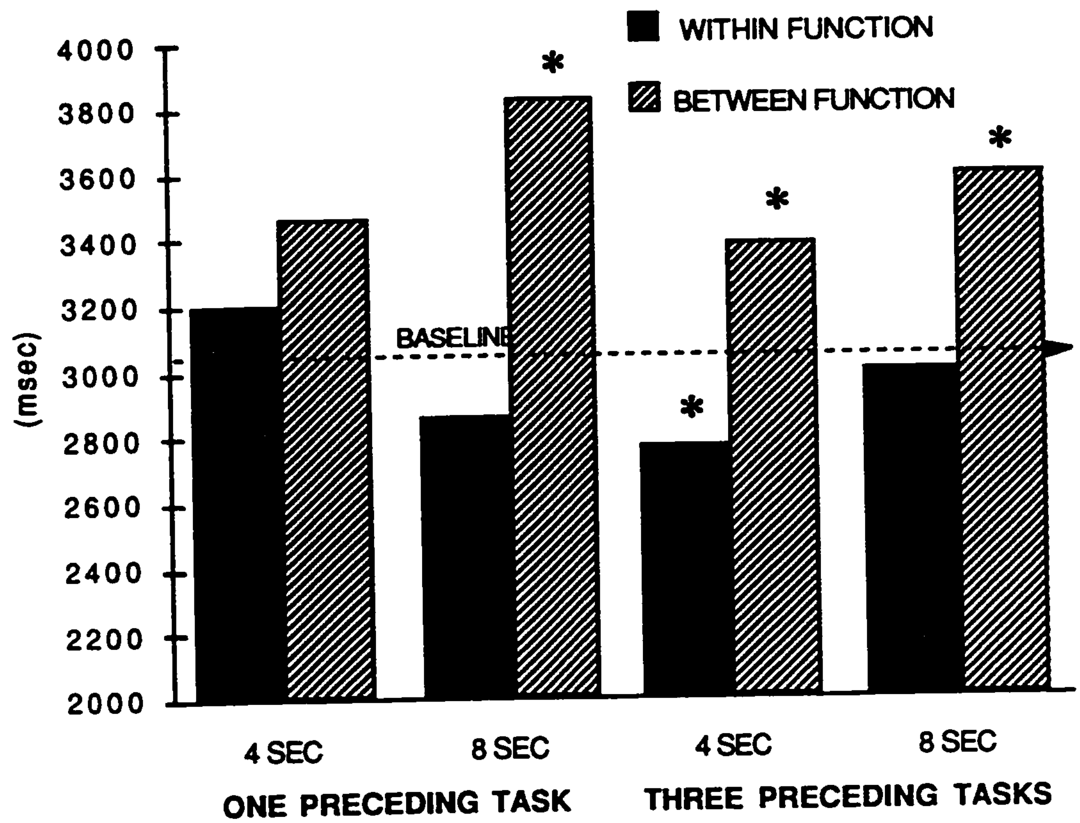


Figure 6. Mean response times for navigation tasks following different experimental switching conditions.

*Condition differs significantly from baseline

**Condition differs near- significantly from baseline

The baseline rate for pilots performing navigation/reference tasks was 3051 ms (SD = 442). Paired comparisons revealed that all four between-function switching conditions were associated with significantly increased response times over baseline rates: (1) when there was one such task with a short interval [*t*-paired (1,11) = 2.22, p = .049], (2) when there was one such task with a long interval [*t*-paired (1,11) = 3.91, p = .002], (3) when there are three such tasks with a short interval [*t*-paired (1,11) = 2.94, p = .014], and (4) when there are three such tasks with a long interval [*t*-paired (1,11) = 4.51, p = .001]. Conversely, the within-function, three-task, short interval switching condition was significantly faster than baseline [*t*-paired (1,11) = 3.91, p = .002] with pilot's requiring an average of 282 ms less time to respond under such conditions. These findings suggest that for navigation/reference tasks the relative performance differences between conditions are generally (but not exclusively) a product of between-functions shifts generating interference rather than within-function shifts producing facilitation

A floor effect severely skewed error rate data for navigation/reference tasks, with zero errors occurring in several conditions. This limited the utility of inferential analysis. Nonetheless, the actual rates are presented in Figure 7 and seem to coincide with the reaction time data in that between-function shifts produced greater errors, especially under long interval conditions. Consequently, there is no indication of a speed-accuracy trade-off.

Systems Management Tasks

Systems management performance was assessed by response times and error rates to 4-choice queries regarding current gauge readings or trends as displayed in the systems window of the simulation. Means and standard deviations of pilot response times to this task are presented in Table 5.

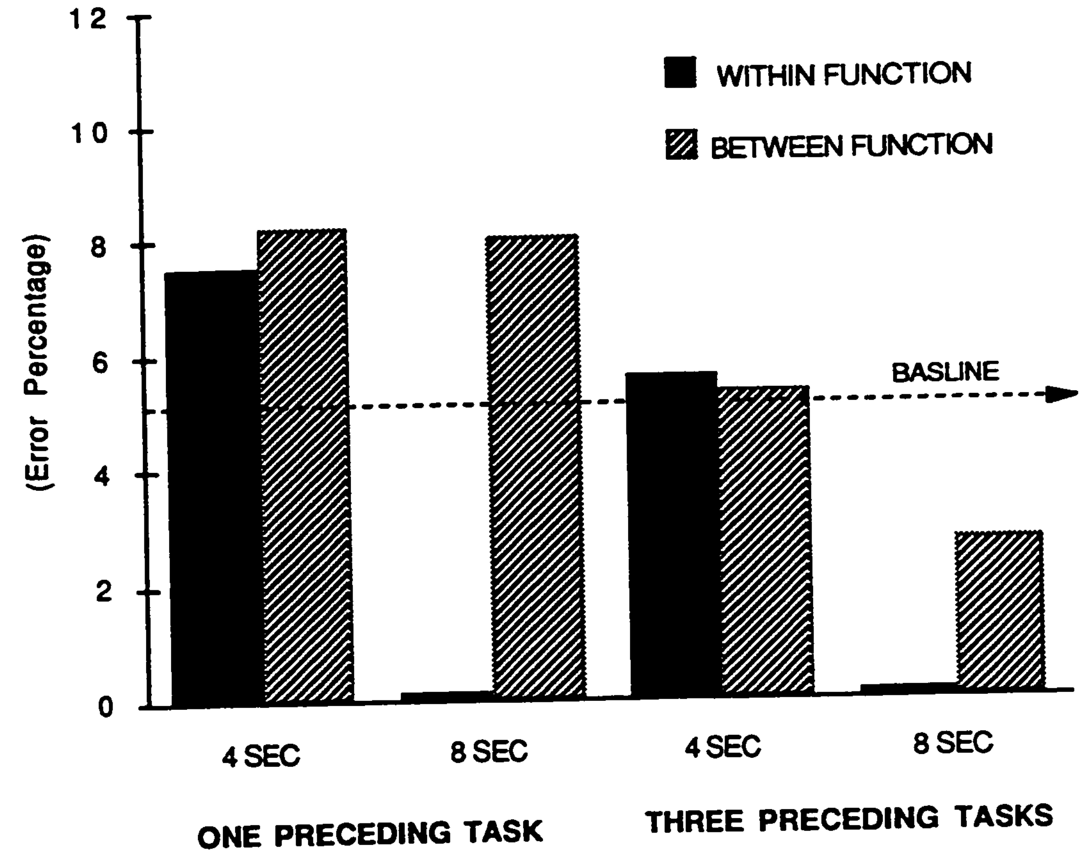


Figure 7. Mean error rates for navigation tasks following different experimental switching conditions.

Table 5.

System Management Performance: Means and Standard Deviations of Pilot Response Times in Milliseconds.

	Functional Continuity	
	Shift Within Function	Shifts Between Function
One Preceding Task		
4 s Interval	2870 (SD = 573)	2798 (SD = 382)
8 s Interval	2774 (SD = 440)	2753 (SD = 482)
Three Preceding Tasks		
4 s Interval	3154 (SD = 599)	2984 (SD = 450)
8 s Interval	3027 (SD = 265)	2661 (SD = 430)

A 2 (functional continuity) X 2 (temporal proximity) X 2 (attentional focus) within-subjects analysis of variance was performed on the reaction time data for systems management tasks. As hypothesized, a main effect of functional continuity was found [$F(1,11) = 6.19$, $p = .03$]. However, the direction of this effect was opposite that which had been predicted. For systems management tasks the longest response times consistently occurred for within-function transitions. There was also a main effect of temporal proximity [$F(1,11) = 6.98$, $p = .023$] with all long interval conditions associated with faster response times (see figure 8). This finding might seem at odds with the attenuating effects predicted for long interval conditions. But faster response times in each of the long interval conditions represents a return towards baseline performance, or an "attenuation" of those factors which may have increased response times.

Average baseline reaction time for all pilots was computed to be 2670 ms (SD = 386) for this task type. Paired *t* -test comparisons of experimental rates to baseline revealed three conditions, all involving three preceding tasks, in which reaction times were significantly increased over baseline: (1) within-function transitions following a short interval [$t_{\text{-paired}}(1,11) = -4.30$, $p = .001$], (2) between-function transition with a short interval [$t_{\text{-paired}}(1,11) = -4.51$, $p = .001$], and (3) within-function transition with a long interval [$t_{\text{-paired}}(1,11) = -4.98$, $p < .001$]. These findings, again, suggest that significant performance differences between conditions reflect differences in relative interference, i.e., which combination of factors create more or less task switching interference.

As there was some variability in error rates for system management tasks, a 2 (functional continuity) X 2 (temporal proximity) X 2 (attentional

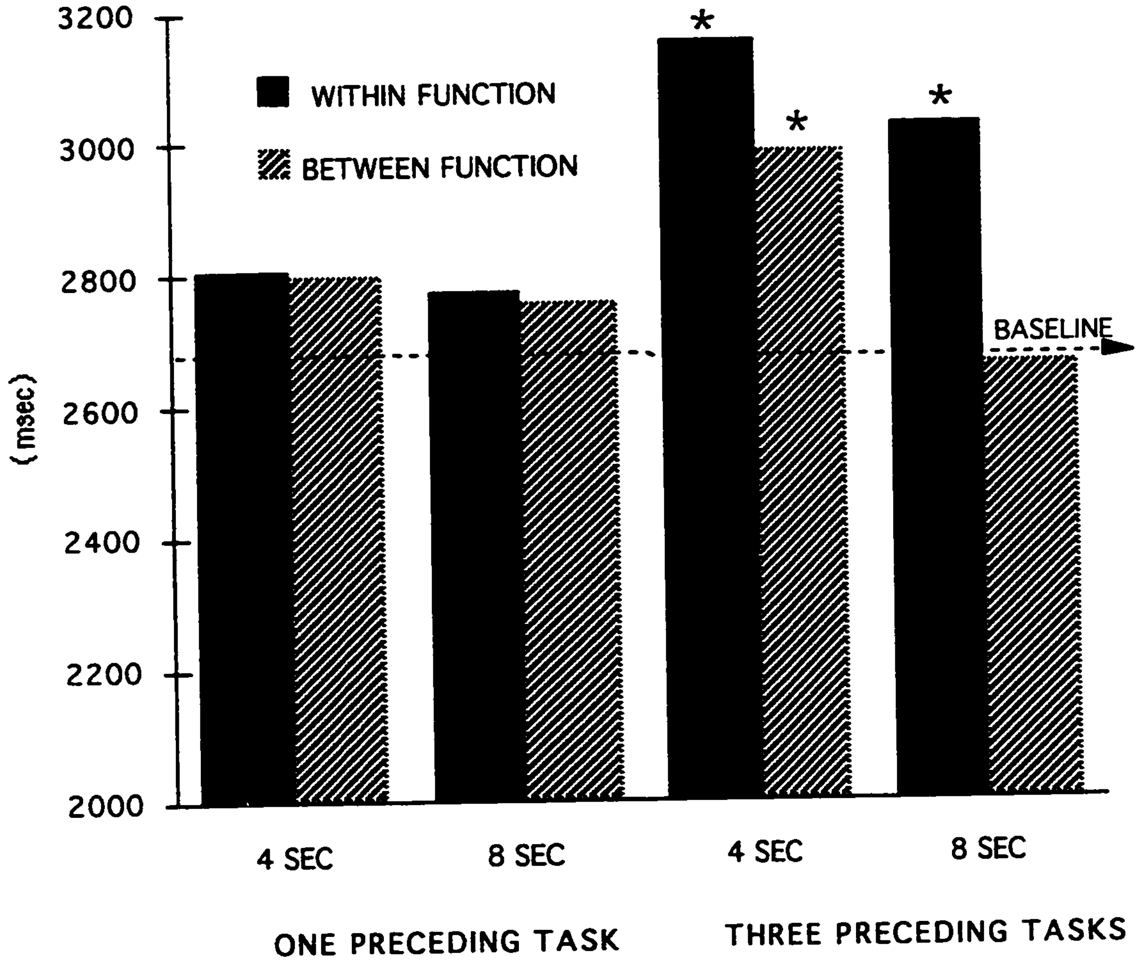


Figure 8. Mean reaction times for system management tasks following different experimental switching conditions.

*Condition differs significantly from baseline focus) within-subjects analysis of variance was performed on the error data. A significant three-way interaction was found [F (1,11) = 4.97, p = .048]. The source of this effect is revealed quite clearly in Figure 9 in which the 16% error rate for the within-function, three-task, short interval condition is more than triple the baseline rate of 4.7%. This coincides with the fact that this condition also produced the longest response time. Once, again, there is no evidence of a speed-accuracy trade-off.

Discussion

There are four important findings in the present study. First, there is evidence that functional continuity does affect task switching efficiencies for flight related tasks. Second, the presence and direction of such effects depends on the nature of the tasks being performed. Third, temporal proximity can mediate these effects. And, last, despite the costs which are assumed to be inherent to all task switching, there are conditions under which performance is actually enhanced over baseline rates. These findings will be discussed in turn.

Functional Continuity

The primary goal of this study was to examine Abbott and Rogers' (1993) assertion that it requires more time and effort to transition between functionally dissimilar flight tasks than between functionally similar ones. In fact, the data from this research clearly show that performance of both navigation/reference tasks and systems management tasks is sensitive to the functional domain of immediately preceding activity. This was evidenced by a main effect of functional continuity on reaction times for both task types. As predicted, performance on navigation/reference tasks was significantly slower when immediately preceding activity was functionally different. These effects,

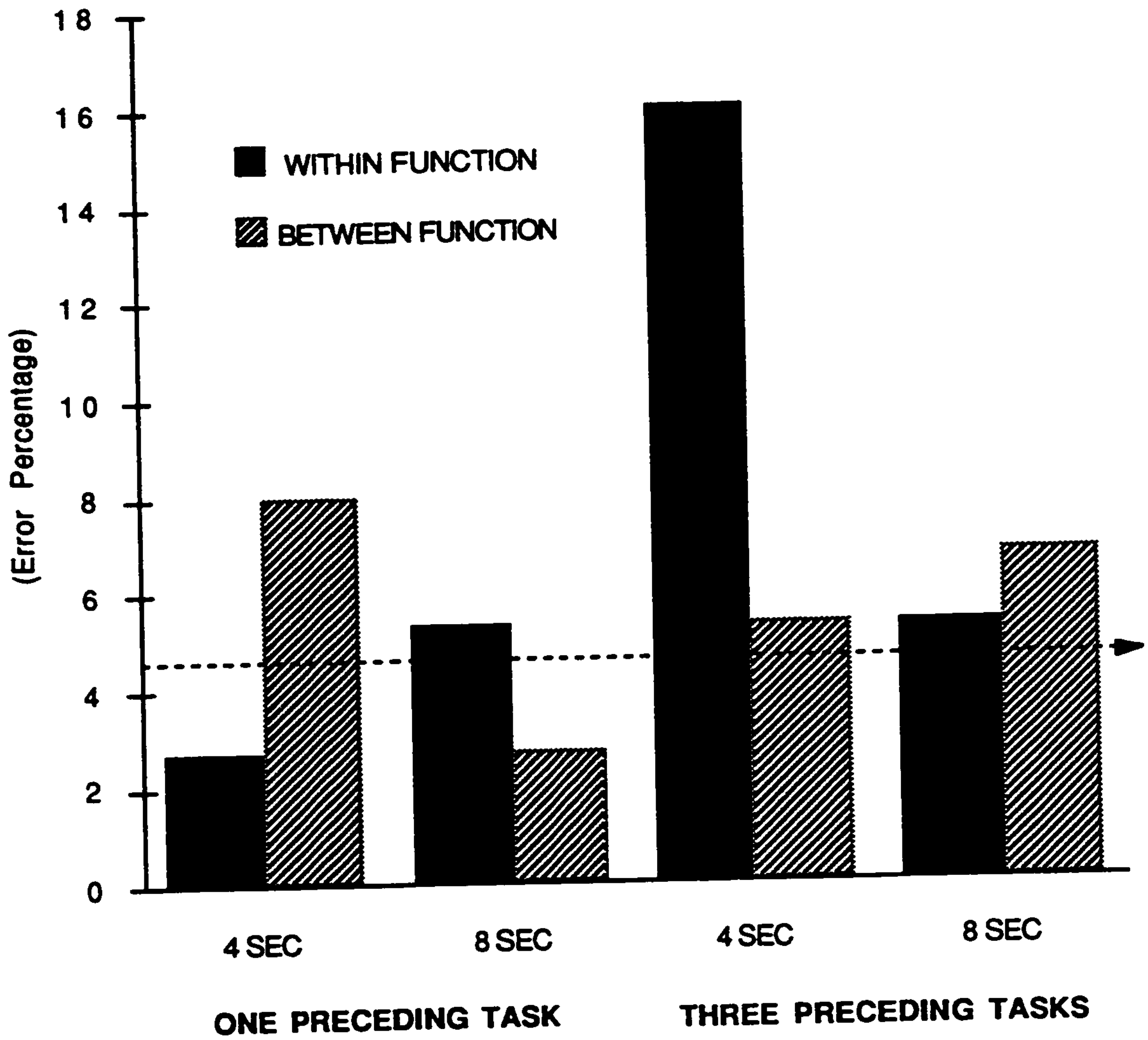


Figure 9. Mean error rates for system management tasks following different experimental switching conditions.

however, were reversed for systems management tasks in which consistently slower performance followed transitions from other systems management tasks, i.e., the within-function conditions. In contrast, flight control tasks were unaffected by functional continuity [except as noted for a particular three-way interaction in which the significantly degraded performance may have reflected a "let-down" in vigilance during the long interval following three rapidly sequenced control tasks (see Kantowitz & Sorkin, 1983)].

Task Dependent Effects

As these varying effects were incompatible with attentional displacement as the operative mechanism of functional continuity, other explanations were explored. As noted earlier, performance for all task types was dependent on display information. What did vary between task types was the predictive value of the associated display information. In the case of navigation/reference tasks, the map display information remained static throughout a given scenario (with the exception of ownship's slow and steady track). Features, landmarks, and positional information gleaned in just completed navigation/reference tasks would offer highly stable, predictive information for future tasks. This would support faster performance times for navigation/reference tasks following within-function transitions when such valuable information might remain immediately available in working memory. Conversely, the high rate of change and abrupt directional reversals of gauge movements made the systems display information highly unstable with negative predictive value. Gauge values and directional trends encoded to working memory from just completed systems management tasks could interfere with current assessments when such information was in conflict. This discordance could require time and effort to overcome which would be revealed in slower performance of system management tasks following within-function transitions. In the case of flight control tasks, display information functioned as feedback to support response movements. Both the graphical tracking display and the digital readouts for speed, altitude, and heading were reactive informational elements driven by pilot input. As such, they were stable (owing to the auto-pilot feature) until the pilot made them unstable and had neutral predictive value. This would support the absence of functional continuity effects on flight control tasks.

Temporal Proximity

Temporal proximity was believed to play a mediating role in functional continuity effects. More precisely, it was predicted that functional continuity effects would diminish as the time between tasks increased. This hypothesis was supported by a main effect of temporal proximity for systems management tasks in which all long interval conditions had reduced reaction times which more closely approximated baseline rates (refer to Figure 8). As noted earlier, such effects represented a dampening of those factors which were driving performance away from baseline.

The significant two-way interaction of functional continuity and temporal proximity on the performance of navigation/reference tasks is further evidence for the mediating effects of temporal proximity. Contrary to prediction, however, long interval conditions increased (rather than attenuated) the departure of performance from baseline for between-function transitions. The unexpected direction of this interaction might, again, be explained in terms of the predictive nature of previously observed display information. Long interval conditions might simply decrease the availability, due to memory decay, for highly predictive navigation/reference information at the onset of such a task.

Facilitation Versus Interference

That differences in switching conditions can lead to differences in task performance has been established in this study with regard to functional continuity and in other studies with regard to predictability (Garcia-Ogueta, 1993), modalities (Laverg et al., 1971), and semantic domains (Bower, 1982). Such findings of superior or inferior performance between investigated conditions does not, in itself, specify the relative effect of a given condition to baseline performance. Consequently, a series of analytic comparisons was undertaken to categorize observed effects as facilitory, interfering, or neutral.

Given the stated premise that all task switching involves some costs, it was not surprising that nearly all conditions yielded either equivalent or degraded performance from baseline. However, performance for navigation/reference tasks following a short interval from a sequence of three other navigation/reference tasks was significantly improved over baseline (a combination of factors initially predicted to have optimal benefits). This exception is worth noting for it demonstrates that it is possible to structure immediately preceding activity to facilitate current task performance, rather than simply to minimize interference.

Design Implications and Future Research

Although preliminary, this study reflects the advantages of switching within a functional category when predictive information is available. When the predictive information is negative there is negative transfer. This may have important implications for cockpit design and procedural operations. The generalizability of these effects could be further explored in higher fidelity simulations and across time-shared behaviors.

References

Abbott, T., & Rogers, W. H. (1993). Functional categories for future flight deck design. (Report No. TM-109005). Hampton, Virginia: NASA Langley Research Center.

Adams, R., & Pew, R. (1990). Situational awareness in the commercial cockpit -- A cognitive perspective. In Proceedings of Digital Avionics Systems Conference, 9th, (pp. 519-524). Virginia Beach, VA: Institute of Electrical and Electronic Engineers, Inc.

Adams, R., Tenny, Y., & Pew, R. (1991). State of the art report: Strategic workload and the cognitive management of advanced multi-task systems. (Report No. BBN-7650). Wright-Patterson Air Force Base, Ohio: CSERIAC.

Bower, G. H. (1982). Plans and goals in understanding episodes. In A. Flammer and W. Knitsch (Eds.), Discourse processing. (pp. 2-15). New York: North-Holland.

Broadbent, D. E. (1958). Perception and communications. London: Pergamon.

Chernikoff, R., Duey, J. W., & Taylor, F. W. (1960). Two dimensional tracking with identical and different coordinate dynamics. Journal of Experimental Psychology, 66, 95-99.

Damos, D. L. (1991). Dual-task methodology: Some common problems. In D. L. Damos (Ed.), Multi-task performance. New York: Taylor & Franks.

Garcia-Ogueta, M. I. (1993). Internal attentional switching: Effects of predictability, complexity, and practice. Acta Psychologica, 83, 13-32.

Gillie, T., & Broadbent, D. (1989). What makes interruptions disruptive? A study of length, similarity, and complexity. Psychology Research, 50, 243-250.

Hart, S. G. (1989). Crew workload management strategies: A critical factor in system performance. In Proceedings of the Fifth International Symposium on Aviation Psychology, 1, (pp. 2-27). Columbus, OH: The Ohio State University.

Hirst, W., & Kalmar, D. (1987). Characterizing attentional resources. Journal of Experimental Psychology: General, 116, 68-81.

Johsson, J. E., & Ricks, W. R. (1993). Categorization and prioritization of flight deck information. In Proceedings of the Seventh International Symposium on Aviation Psychology, 1, (pp. 126-131). Columbus, OH: The Ohio State University.

Kantowizt, B., & Sorkin, D. (1983). Human factors: understanding people-system relationships. New York: Wiley & Sons.

Klapp, S. (1979). Doing two things at once: The role of temporal compatibility. Memory and Cognition, 7, 375-381.

Kleinman, D. L., & Curry, R. E. (1977). Some new control theoretic models for human operator display monitoring. IEEE Transactions on Systems, Man, and Cybernetics, 7, 778-784.

Laverg, D., VenGelder, P., & Yellott, S. (1971). A cueing technique in choice reaction time. Journal of Experimental Psychology, 87, 225-228.

Navon, D., & Gopher, D. (1979). On the economy of the human processing system. Psychological Review, 86, 214-255.

Nagel, D. C. (1988). Human error in aviation operations. In E. L. Weiner & D. C. Nagel (Eds.), Human factors in aviation. New York: Academic Press.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. Journal of Experimental Psychology: General, 106, 226-254.

Schweickert, R., & Boggs, G. J. (1984). Models of central capacity and concurrency. Journal of Mathematical Psychology, 28, 223-81.

Wiener, E. F. (1985). Beyond the Sterile Cockpit. Human Factors, 27, 75-90.

Wickens, C. D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), Attention and performance VIII (pp. 239-57). Hillsdale, NJ: Lawrence Erlbaum.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman and R. Davies (Eds.), Varieties of attention (pp. 63- 101). New York: Academic Press.

Appendix A. Signed Approval Form

APPROVED BY THE MASTER'S THESIS COMMITTEE

(Dr. Kevin Jordan, chairperson)

(Robert J. Shively, committee member)

(Dr. Laree Huntsman, committee member)

SJSU SAN JOSE STATE UNIVERSITY

A campus of The California State University

Office of the Academic Vice President • Associate Academic Vice President • Graduate Studies and Research
One Washington Square • San Jose, California 95192-0025 • 408/924-2480

TO: Allen D. Goodman
128 Rincon St.
Santa Cruz, CA 95060

FROM: Serena W. Stanford *Serena W. Stanford*
AAVP, Graduate Studies & Research

DATE: March 14, 1996

The Human Subjects-Institutional Review Board has approved your request to use human subjects in the study entitled:

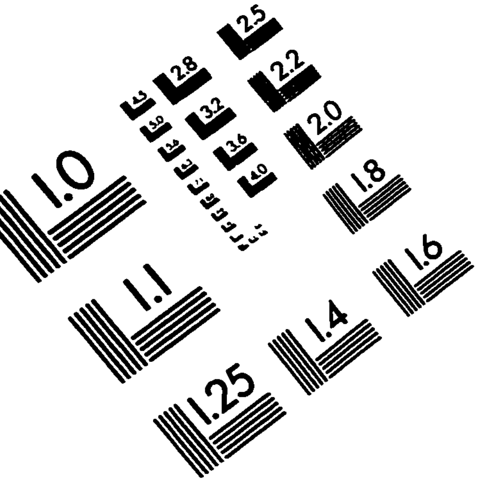
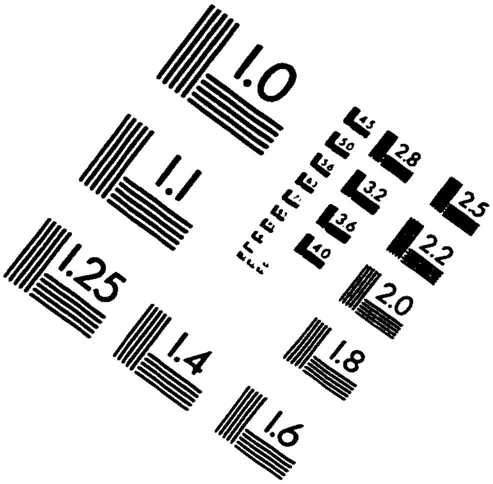
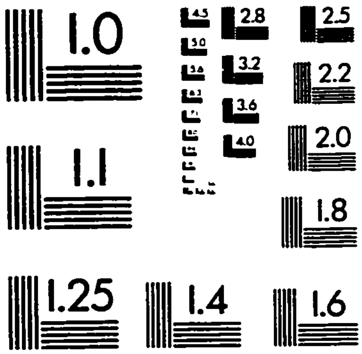
"Task Switching Efficiencies Within and Across Flight Related Functional Categories"

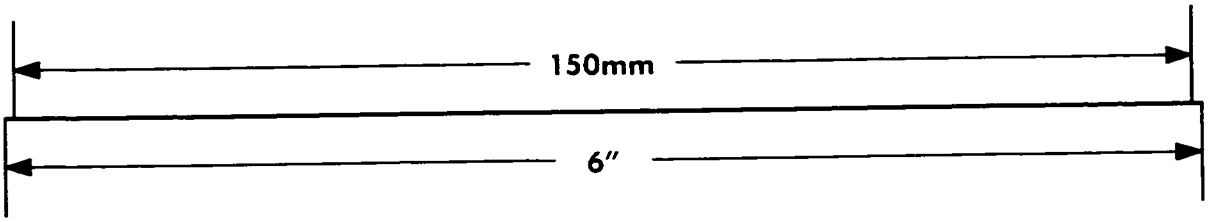
This approval is contingent upon the subjects participating in your research project being appropriately protected from risk. This includes the protection of the anonymity of the subjects' identity when they participate in your research project, and with regard to any and all data that may be collected from the subjects. The Board's approval includes continued monitoring of your research by the Board to assure that the subjects are being adequately and properly protected from such risks. If at any time a subject becomes injured or complains of injury, you must notify Serena Stanford, Ph.D., immediately. Injury includes but is not limited to bodily harm, psychological trauma and release of potentially damaging personal information.

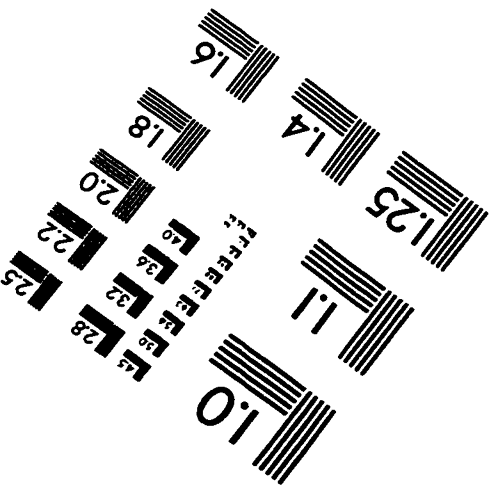
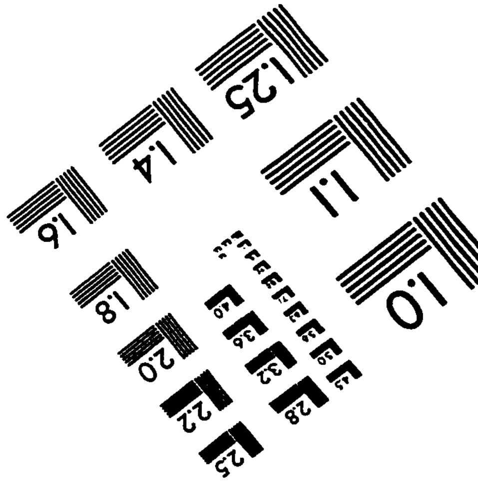
Please also be advised that each subject needs to be fully informed and aware that their participation in your research project is voluntary, and that he or she may withdraw from the project at any time. Further, a subject's participation, refusal to participate, or withdrawal will not affect any services the subject is receiving or will receive at the institution in which the research is being conducted.

If you have any questions, please contact me at (408) 924-2480.

IMAGE EVALUATION TEST TARGET (QA-3)



APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989