

Montgomery, D. A.

Effects of Selected Task Performance Criteria at Initiating Adaptive Task Reallocations

Demaris A. Montgomery, Ph.D.
Department of Psychology
Bradley University, Peoria, IL 61625

January 2001

Final Report
NASA Langley Research Center

Table of Contents

1.	Summary.....	3
2.	General Introduction.....	4
3.	Experiment 1	
	a. Introduction.....	5
	b. Method.....	7
	c. Results and Discussion.....	11
	d. Conclusions.....	18
4.	Experiment 2	
	a. Introduction.....	21
	b. Method.....	22
	c. Results and Discussion.....	24
	d. Conclusions.....	32
5.	General Conclusions.....	33
6.	References.....	35

Summary

In the current report various performance assessment methods used to initiate mode transfers between manual control and automation for adaptive task reallocation were tested. Participants monitored two secondary tasks for critical events while actively controlling a process in a fictional system. One of the secondary monitoring tasks could be automated whenever operators' performance was below acceptable levels. Automation of the secondary task and transfer of the secondary task back to manual control were either human- or machine-initiated. Human-initiated transfers were based on the operator's assessment of the current task demands while machine-initiated transfers were based on the operators' performance. Different performance assessment methods were tested in two separate experiments.

Experiment 1

In the first experiment, human-initiated transfers were compared to machine-initiated transfers that were based on either primary task performance or a combination of primary and secondary task performance (joint assessment). Moreover, each assessment method was tested given machine-initiated transfers to automation only and machine-initiated transfers to both automation and manual control. Altogether, there were five switching methods tested: completely human-initiated, machine-initiated transfers to automation only based on primary or joint assessment, and machine-initiated transfers to both automation and manual control based on primary or joint assessment. The five switching methods produce similar performance on the primary task measures, but there were differences among the secondary task measures. Machine-initiated transfers to automation coupled with human-initiated returns to manual control and joint performance assessment produced the best system performance, but these gains depended on a high reliance on automation. In addition, there was a higher proportion of mode errors (i.e., accidental responses while in automation) given machine-initiated transfers to automation, particularly given machine-initiated transfers to both automation and manual control.

Experiment 2

In the second experiment, similar switching methods as those used in experiment 1 were tested, but the switching method that involved machine-initiated transfers to both automation and manual control was modified. With this method the operator was signaled when to implement a mode change rather than being simply informed of the change. More importantly, two performance assessment criteria were tested. Mode transfers depended on an absolute threshold value similar to the joint performance threshold criteria found in experiment 1 or on evidence of a continued change in performance beyond the threshold criteria. First, including the human operator in machine-initiated transfers reduced the proportion of mode errors produced by transitions to both automation and manual control relative to machine-initiated transfers to automation only. Second, the assessment method that required a change in performance produced performance advantages relative to the absolute threshold criterion without a heavy reliance on automation. There was a small decrement in secondary task performance, but also evidence that the number of mode errors decreased given the change in performance criterion.

Effects of Selected Task Performance Criteria at Initiating Adaptive Task Reallocations

Recently, technological advances have made it possible to automate many functions such that tasks can be assigned to either the human or the machine (e.g., expert systems used in decision making and autopilots in an aircraft). Under some circumstances, tasks are automated to reduce operator workload and increase system reliability. In other situations, the human operator controls the tasks when unpredictable, dynamic changes in the system have to be addressed or when optimal levels of operator engagement have to be maintained in vigilance tasks.

Ideally, system performance is optimized by assigning tasks to the appropriate mode (e.g., manual control or automation) depending on the situation demands. Systems that allow this type of dynamic mode adjustment represent adaptive automation. An important issue to be addressed regarding adaptive automation involves assessing the mechanisms used to initiate task reallocations (Scerbo, 1996).

Task reallocations may be initiated through a variety of methods where the general objective of most of these methods is to vary the mode or level of automation in order to maintain the optimal level of operator workload (Scerbo, 1996) and total system performance. Some methods determine the demands placed on the operator by monitoring operator performance (e.g., Rencken & Durrant-Whyte, 1993; Parasuraman, Mouloua, & Molloy, 1996, Kaber & Riley, 1999) or psychophysiological measures (e.g., Byrne & Parasuraman, 1996). These measures are monitored in real time and evaluated against some standard to determine the appropriate mode. Thus, they are truly responsive to the current demands and the particular individual operating the system. There are, however, incidental costs. They are computationally demanding and they can produce highly reactive systems (i.e., rapidly cycling between modes) which contributes to impaired operator mode awareness.

Alternative methods include operator performance modeling and monitoring mission activities (Parasuraman, Mouloua, & Molloy, 1996; Scerbo, 1996). With these latter methods operator performance on the relevant tasks is monitored in advance to identify the particular points during operation of the system where performance suffers. Then, mode shifts are initiated at the pre-specified periods or when the critical mission events are detected during subsequent system activity. The predictability of the mode shifts reduces the aforementioned problems. However, in situations where system activity is not predictable (e.g., operation start up or an emergency) and requires active operator involvement, these latter methods may be less useful than the physiological and performance assessment methods.

Research has substantiated the effectiveness of using physiological measure of arousal for initiating task reallocations in adaptive automation (e.g., see Byrne & Parasuraman, 1996). The results of these types of investigations have been used to develop a *closed-loop biocybernetic*

system at NASA, which is being developed and tested as a part of the Crew Hazards and Error Management (CREW) project. This system monitors electroencephalographic (EEG) activity to determine the level of automation needed to maintain optimal operator engagement.

Performance assessment of operator workload is also an effective means for triggering mode shifts in adaptive automation. For example, Rencken and Durrant-Whyte (1993) tested a quantitative queuing system for adaptive task allocation on a human operated surveillance system. They found that task reallocations initiated by changes in operator response times and error rates on a variety of tasks produced improved overall system performance. Similarly, Parasuraman, Mouloua, and Molloy (1996) had participants perform three simulated flight tasks in which one of the tasks (an engine status task) could be performed manually or it could be automated. They found that when a performance criterion was used to return the engine status task back to operator control for a brief period participants' error detection rates improved relative to a continuous automation condition. Kaber and Riley (1999) also found evidence indicating that a secondary task measure of workload in a dual-task setting was an effective means of triggering adaptive task reallocations.

Further research should address the use of performance assessment for triggering adaptive task reallocations since these methods are relatively non-intrusive. Performance measures of response times and error rates are easily measurable and can be incorporated in existing control models of complex systems. In addition, the performance assessment can be coupled with physiological measures to produce more effective multi-attribute assessment methods for initiating task reallocations.

The current investigation will assess the effectiveness of various performance threshold assessment methods of operator workload as a means of delegating a task to either automation or human control in a complex system. Specifically, operators will actively control a process in a fictional system while they simultaneously monitor secondary task displays for critical events. When the task demands are high, as indicated by a decline in operator performance, one of the monitoring tasks can be automated. When performance returns to acceptable levels the monitoring task can be returned to manual control. Various approaches to measuring operator performance as a basis for initiating mode shifts will be examined in two experiments to determine the optimal level of engagement by the human operator.

Experiment 1

The first experiment examined the efficacy of using primary and secondary task measures of operator workload as the basis of adaptive task reallocations. Measures of primary task performance (i.e., performance on the main task of interest) will generally decrease as workload increases, but this is not always true. The operator may direct additional resources toward the primary task in order to maintain high performance. Thus, performance on a concurrent secondary task is often measured to assess reserve capacity (Kaber & Riley, 1999; Wickens, Gordon, & Liu, 1998). It is assumed that as additional resources are devoted to the primary task fewer resources are available for the secondary task and secondary task performance reflects the

increase in workload.

In a previous pilot study, Montgomery & Gronert (1999) tested various triggering mechanisms based on performance threshold assessment of increases in operator workload. The participants in their study actively controlled a similar process as described in the proposed study. When performance dropped below specified threshold levels one of the monitoring tasks became automated. In four separate conditions, the mode shift depended one of the following: (1) primary task performance, (2) secondary task performance, (3) primary and secondary task performance combined, or (4) the operator's assessment of the current task demands.

They found that operator performance in controlling the primary task and in detecting secondary critical events were both highest when both primary and secondary task performance criteria were used to initiate mode shifts. Furthermore, although the effect was not statistically significant, the joint assessment method produced the most points accumulated. Thus, overall system performance seemed to benefit most from the use of a joint threshold assessment method. This is consistent with other results showing that multivariate assessments produce better adaptive systems (see Byrne & Parasuraman, 1996, p. 253).

Similar to the previous pilot study (Montgomery & Gronert, 1999), operators in the proposed study will be controlling a fictional process while simultaneously monitoring secondary task displays for critical events. In addition, as with the previous study either the human operator or the machine will automate one of the monitoring tasks. The human-initiated switch will be completely controlled by the operator and based solely the operator's assessment of the current task demands. With the machine-initiated switches the computer will automate the task when performance drops below specified criteria for either the primary task alone or both the primary and secondary tasks combined.

Moreover, in this study additional conditions will be included that should reduce the total time in automation for machine-initiated transfers to automation. In the Montgomery and Gronert (1999) study the observer controlled the transfer of the secondary task back to manual control. The heavy reliance on automation for the machine-initiated transfers found in their study suggests that operators tended to leave the secondary task in automation. When the performance threshold criterion is used only to automate a task, only increases in workload are being considered and it is assumed that the operator will exercise good judgment about the appropriate time to return to manual control.

Alternatively, when performance threshold criteria are used to produce shifts to both automation and a return to manual control, the computer is considering both overload and underload in the operator. To maintain optimal operator involvement and workload levels it is important to both reduce demands when they are too heavy, but also increase demands when they are too low (Parasuraman, Mouloua, & Molloy, 1996). Thus, optimal periods in automation and total system performance may result from workload assessment that produces mode shifts to both automation and manual control.

Altogether, machine-initiated mode shifts should produce better total system performance than human-initiated mode shifts, particularly when the machine initiates shifts to both automation and manual control. This advantage will be minimized, though, if the operator is slow to recognize the current mode of automated functions. A heavy reliance on automation tends to produce lower operator situation awareness (Endsley & Kiris, 1995) and, thus, an increased likelihood of mode errors. These mode errors may be most likely when the modes are transferred more frequently for machine-initiated shifts in both directions.

Finally, all conditions will be tested under two time constraint conditions. In one case, there will be no time constraint. That is, the operator has complete control over the task rate. In the second case, there will be a time constraint on the operator to alter the levels of the reservoirs or reset the gauges. Any operator input to the system entered before a time limit will be accepted before a system update occurs. Based on pilot data, the time constraints may influence the predicted differences among the triggering methods tested.

Method

Participants Twenty individuals were tested on a preliminary set of trials to identify those individuals who were able to meet performance criteria established from previous pilot data. Seventeen individuals (11 females and 6 males) were selected who produced mean reaction times less than 5000 msec and a total point accumulation greater than 0 on the preliminary trials. All participants were undergraduate students enrolled at Bradley University. The participants were paid minimum wage for their involvement in the study. Any individuals who did not complete all experimental sessions were paid for any time in which they did participate.

Experimental Task Displays Graphical images depicting state changes for three different sub-tasks of a fictional system were presented on a color monitor driven by a Pentium III-500 computer. Figure 1 provides an example of the display of the three sub-tasks presented to participants. First, the primary task, located in the upper left corner of the monitor, involved controlling the water levels of reservoirs A and B that fed water into a third reservoir. White lines, representing the reservoir levels, were positioned at the center of each reservoir at the beginning of a monitoring period. During a monitoring period, at randomly determined times, water was drawn from reservoir A or B.

The amount drawn from a given reservoir depended on the difficulty of the task. For low task difficulty 2, 4, or 6 units of water were drawn, depending on the value randomly selected. Given medium task difficulty, 4, 8, or 12 units were drawn and for high task difficulty the units were 6, 12, or 18. For half of the display updates no perturbations occurred on either reservoir (i.e., 0 units of water were drawn from both reservoirs). For the other half of the display updates, n units of water were drawn from either reservoir A or B. Thus, the likelihood that a particular amount of water (e.g., 2, 4 or 6 units) was drawn from a given reservoir (e.g., reservoir A) was approximately .083. Altogether, the reservoir levels could vary from 0 to 100 units. Finally, there are two red critical level indicator lines at the 25 and 75 unit levels.

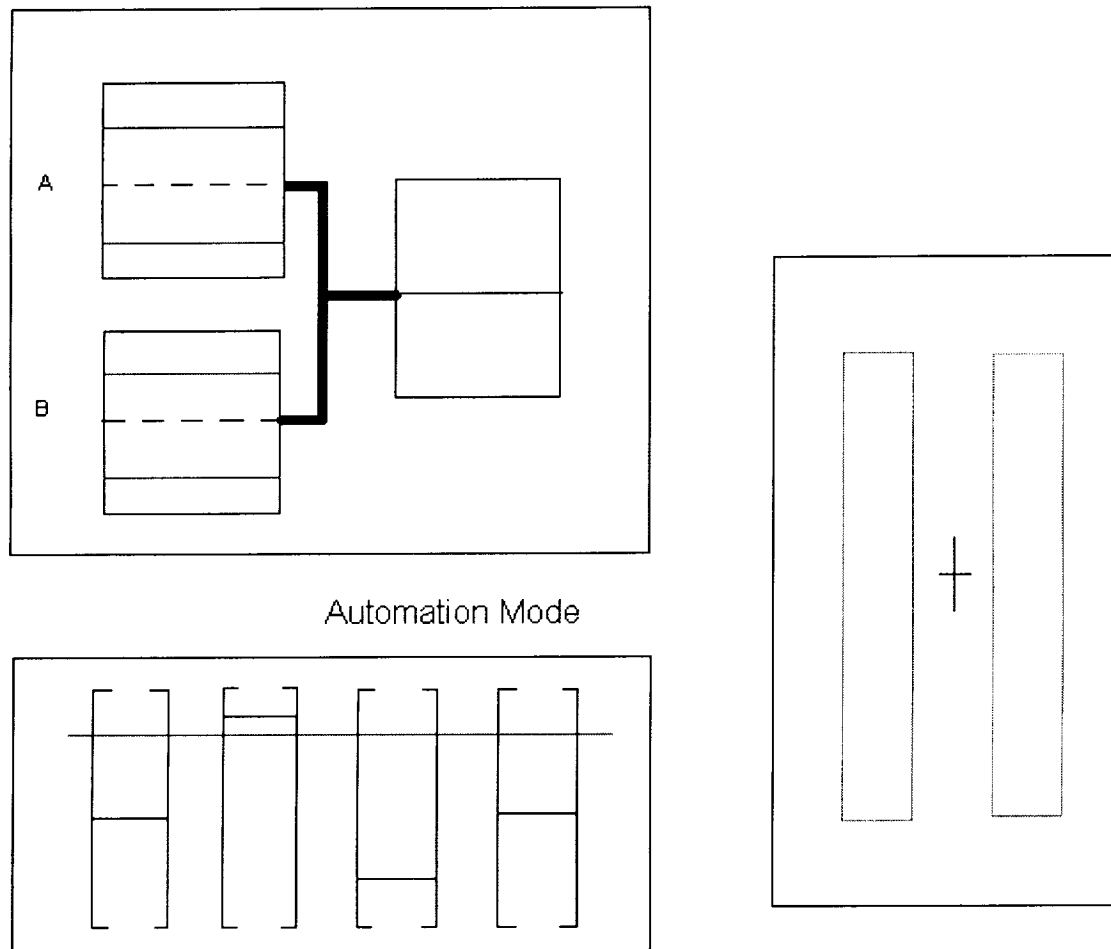


Figure 1. For the primary task, depicted in the top left corner, the dashed lines represent the reservoir levels and the solid lines represent the critical levels. The two monitoring task displays are found on the right side and the lower left of the primary task display. (The gauge task is currently in the Automation Mode.)

Below the reservoir display, four gauges were presented on the monitor. Each gauge was composed of two parallel white lines and magnitude information was depicted by the vertical displacement of a white horizontal marker. At the beginning of each display update, the magnitude of each gauge depended on the value independently sampled from a normal distribution with $\mu = 50$ and $\sigma = 20$. Analogous to the reservoirs, the gauge levels could vary from 0 to 100 units and a red line presented at the top of the gauges demarcated a critical level at 75 units. Since the values were independently sampled from a normal distribution, anywhere from zero to four of the gauges could have values that exceeded the critical value on a given display update. Thus, the probability of a critical event was approximately .42.

Finally, the display for the third task was located on the right side of the monitor. At the beginning of a monitoring period, two green bars appeared in this display. After a randomly

determined delay of 250, 500, 1000, or 1500 ms (all equally likely), one of the bars turned red. In most cases, the operator had to enter a response for this display before any of the component displays were updated and the bar returned to green. (Under constraint conditions, as described below, there were some conditions where the computer automatically updated the display independent of an operator initiated reset.)

Procedure: Sub-tasks Participants were informed that their primary task was to keep the reservoir levels between 25 and 75 units in order to maintain the appropriate flow rate into the third reservoir. If the operator allowed the reservoir levels to exceed the critical values by 20 units, the system would shut down temporarily for an emergency reset. Thus, as long as the levels were between the critical level indicators the flow rates were optimal and emergency resets avoidable.

Pressing keys located in the lower left corner of the keyboard controlled the reservoir levels. The level of reservoir A was increased and decreased by pressing keys labeled A+ and A-, respectively. The level of reservoir B was controlled by pressing keys labeled B+ and B-. The operator's performance on the primary task was assessed by keeping track of both the number of display updates in which the level of either reservoir A or B exceeded the critical values and the number of emergency resets that occur during a monitoring period.

While performing the primary task the operator concurrently responded to events in the other two displays. For the gauge task displayed in the lower left corner, the operator had to respond whenever a gauge indicator exceeded the 75 unit critical level indicator. The operator responded to these events by pressing a corresponding key at the top of the computer keyboard (e.g., the key labeled "2" to reset the second gauge). Given the correct response, the gauge value changed to a newly selected random value during the next screen update; otherwise, the marker remained fixed in position until the correct response was made. The operators' performance was based on his hit-to-signal ratio (i.e., the number of critical events correctly reset / total number of critical events).

Finally, the participants' third task involved responding to changes in the display located on the right side of the monitor. For this display, whenever one of the two bars turned red the operator used the mouse to click on the red item. The time elapsed between the onset of the red bar and the operator's reset was recorded. When there was no constraint on the participants' responses, the participant had to reset the red item before a display update would occur. Thus, in the absence of a constraint the participants controlled the update rate of the displays and could take as much time as they needed to respond to the other tasks.

Alternatively, when a constraint was present the participant controlled the update rate as long as his time to respond did not exceed a time limit. The time limit depended on the participant. During the first session, participants completed multiple practice sets (as described below). From this practice data each participants' reaction time data was used to calculate a mean and standard deviation. The constraint consisted of a time limit based on the participants' particular mean reaction time plus one standard deviation. For example, if a participant's mean

reaction time for the practice data was 2000 msec and his standard deviation was 500 msec, then he was allowed to respond to all displays up to a limit of 2500 msec. If the participant did not respond by 2500 msec, the display updated based on any responses entered before the limit was exceeded.

Procedure: General The operator was always in control of the primary task and the mouse-based reset task, but the gauge-monitoring task could be automated. A small display above the gauges indicated the current mode, "Manual Mode" or "Automation Mode." In automation the computer monitored the gauges for the operator, allowing the operator to devote full attention to the other two tasks. Three methods of triggering a switch to automation were tested: (1) human-initiated, (2) machine-initiated based on primary task performance, and (3) machine-initiated based on primary and secondary task performance.

Under the human-initiated condition, the operator controlled mode shifts by pressing a key labeled "AU/M" based on the operator's discretion. The operator could override automation and return to manual control by pressing the "AU/M" key again. For the machine-initiated shift based on primary performance, the computer automated the gauge task when the reservoir levels moved beyond the critical level indicators. When the secondary task criteria were added, response times on the mouse-based reset task and error rates on the gauge task also lead to gauge task automation. When the observer's response time exceeded their mean response time by one standard deviation ($M_{RT} + SD_{RT}$) for four consecutive display updates on the mouse-reset task or their error rate on the gauge task exceeded 10%, the gauge task was automated. When the limits were exceeded and the computer automated the gauge task, the operator was signaled by a tone and a by change in the visual display indicating the current mode (i.e., Automation Mode).

Among the conditions that involved a machine-initiated switch to automation, the return to manual control was either human or machine initiated. In one case the participant returned to manual control by pressing the "AU/M" key, based on the operator's discretion as found in the human-initiated condition. Alternatively, the machine could return control of the gauge task to the human operator when performance returned to acceptable levels, based on the criteria described above. For this latter condition the operator was signaled by a tone and a by change in the visual display indicating the current mode (i.e., Manual Mode).

Altogether there were five methods for transferring modes. A completely manual method, all transfers were initiated by the participant. Four machine-initiated transfers, two used the primary criterion only and two used the joint criteria (primary and secondary task performance). Moreover, two of the four machine-initiated transfers produced transfers to automation only (i.e., the operator controlled transfers back to manual control) and two produced transfers to both automation and manual control. Thus, there were five methods tested: Manual, Automation-Primary, Both-Primary, Automation-Joint, and Both-Joint.

The number of display updates while in automation and the number of accidental responses to the gauge task during automation (reflecting participant mode awareness) were also recorded. Finally, operators received points for keeping the reservoirs within the critical limits

(+20 points) and for identifying critical gauge events (+20 points). Similarly, points were subtracted when the reservoir levels exceed the limits (-20 points), gauge critical events were missed (-20 points), and emergency resets occur (-250 points).

As stated earlier, all individuals who were interested in participating in the experiment began by completing a preliminary test set. Those who met the selection criteria were scheduled for seven subsequent sessions. During the first session, participants learned how to control the three sub-tasks and they completed one training period for each switching method. The primary task difficulty was set at low for all practice sets and the order in which the switching methods were presented to the participants during the practice session was determined by the row assigned in a Latin square.

Five monitoring periods were completed during each of the six remaining experimental sessions. Thus, there were thirty experimental monitoring periods, altogether, composed of a factorial combination of the three independent variables: switching method (manual, Automation-Primary, Both-Primary, Automation-Joint, and Both-Joint), primary task difficulty (high, medium, and low), and response constraint (present and absent). The order of presentation for the response constraint was counterbalance. Half of the participants performed the task under a time constraint during the first three experimental sessions and with the time constraint absent during the remaining three experimental sessions. It was reversed for the other half of the participants. The order in which participants viewed both the switching method and the primary task difficulty was randomized.

Results and Discussion

First of all, the factorial combination between switching method (Manual, Automated, and Both) and performance measure (Primary and Joint) was incomplete since manual switches to automation did not include a computer-based assessment of performance. Thus, ANOVAs were first performed comparing five switching methods; the Manual method was compared to the factorial combination of the two computer-based switches with two performance measures (i.e., Automation-Primary, Automation-Joint, Both-Primary, and Both-Joint).

Subsequently, ANOVAs were also performed with the Manual condition removed, treating switching method and performance measures as two separate variables. Most significant results found in the first analysis, with the Manual condition included, were replicated in the second analysis. Thus, the results from the second analysis are reported. However, when there is an exception the results of the first analysis are reported, as well.

Primary Task and Total System Performance The number of trials (display updates) in which either reservoir level was beyond the critical level lines was calculated and the means are depicted in Table 1 for the three levels of task difficulty. In general, the number of times operators exceeded unsafe levels increased as the magnitude of the level perturbations increased, $F(2,32) = 50.46$, $p < .001$. Analytic t-tests with a Bonferroni adjustment to the error rate indicated that performance declined with each increment in task difficulty ($p < .001$).

The means for the number of emergency resets and the total points accumulated are also depicted in Table 1. ANOVAs performed on both the number of emergency resets and the total number of points accumulated indicated significant main effects of task difficulty for both measures, $F(2,32) = 42.24$, $p < 0.001$ and $F(2,32) = 126.08$, $p < .001$, respectively. Subsequent analytic comparisons indicated an increase in the number of emergency resets and a decrease in the total points with each increment in task difficulty ($p \leq .001$).

Table 1. Mean performance measures for the primary task and total points accumulated.

Primary Task Difficulty	Number of Critical Reservoir Events		Number of Emergency Resets		Total Points	
	<u>M</u>	<u>SE</u>	<u>M</u>	<u>SE</u>	<u>M</u>	<u>SE</u>
Low	7.52	1.69	0.64	0.26	1975.80	207.73
Medium	11.77	1.79	0.91	0.27	1540.62	183.54
High	15.59	1.79	1.78	0.37	1068.91	213.66

There were no other significant effects for any of the primary and total system performance measures. However, in the first analysis, with the manual condition included, there was a significant effect of the switching method on the total points accumulated, $F(4,64) = 3.88$, $p = .007$. The highest points were accumulated when operators had complete control over switches to automation in the manual condition ($M = 1713.92$, $SE = 210.17$). Conversely, the lowest points were accumulated when mode transfers were to automation only and depended on the joint task assessment method ($M = 1367.36$, $SE = 193.35$). Analytic t-tests with a Bonferroni adjustment to the error rate indicated that the difference between the Manual and Automation-Joint conditions was the only significant difference ($p < .005$).

Such differences in the total points most likely reflect participants' efforts to accumulate as many points as possible. The more time one spent in automation the fewer points he could accumulate from resetting the secondary gauge task. Thus, participants were instructed that they could accumulate more points by keeping the secondary task in manual control as long as they were able to avoid emergency resets and maintain relatively high accuracy in detecting critical gauge events. Evidence that they may be using this strategy is found among the secondary task measures, as described below.

Secondary Gauge and Reset Tasks First, as depicted in Table 2, the mean number of trials in automation tends to be greater when transfers were to automation only and when both primary and secondary task performance influenced transfers to automation. An ANOVA performed on the number of trials in automation indicated that there were significant main effects of switching method and performance assessment method. (The F values are also reported in Table 2.)

Moreover, from the first analysis, with the manual condition included, there was a significant effect of the switching method on the number of trials in automation, $F(4,64) = 28.92$, $p < .001$. Analytic t-tests with a Bonferroni adjustment to the error rate ($p \leq .005$) indicated that the Manual condition produced a significantly lower number of trials in automation ($M = 18.71$, $SE = 7.38$) than all other methods, except when machine-initiated transfers occurred in both directions and depended on primary task performance only (Both-Primary).

Table 2. Mean performance on the secondary task measures.

Switching Method	Number of Trials in Automation		Hit-to-Signal Ratio		Reaction Time		Proportion of Responses Automation	
	<u>M</u>	<u>SE</u>	<u>M</u>	<u>SE</u>	<u>M</u>	<u>SE</u>	<u>M</u>	<u>SE</u>
Switching Method								
Automation	52.93	5.68	.925	.011			.061	.012
Both	30.08	3.41	.869	.021	No Effect		.094	.012
<u>F</u> (1,16)	45.94 (<u>p</u> < .001)		9.80 (<u>p</u> = .006)				10.27 (<u>p</u> = .006)	
Task Assessment								
Joint	55.11	5.75	.937	.006	1452.55	114.13		
Primary	27.90	3.61	.857	.024	1531.25	112.32	No Effect	
<u>F</u> (1,16)	46.78 (<u>p</u> < .001)		16.57 (<u>p</u> = .001)		5.04 (<u>p</u> = .039)			

Thus, the slightly higher points accumulated with the Manual method, relative to the Automation-Joint criterion method, resulted from operators maintaining reasonably good control of the secondary gauge task without showing a significant increase in emergency resets or critical events on the primary task. However, there were repercussions for attempting to maintain control of the secondary gauge task, as evidenced in Table 2.

When participants maintained control of the secondary gauge task, it tended to produce lower hit-to-signal ratios and the slower mean reaction times. An ANOVA performed on the hit-to-signal ratios indicated significant main effects of both switching method and performance assessment. (Again, the F values are reported in Table 2.) The hit-to-signal ratios were lower for transfers to both automation and manual control and when primary assessment was used for transfers. Similar trends showed up in the first analysis, where there was a significant effect of switching method, $F(4,64) = 8.83$, $p < .001$. Analytic t-tests with a Bonferroni adjustment to the error rate ($p \leq .005$) indicated that the hit-to-signal ratio for the Automation-Joint ($M = .95$, $SE = .007$) and Both-Joint ($M = .93$, $SE = .008$) criterion methods were significantly greater than ratios for the Manual condition ($M = .86$, $SE = .032$).

An ANOVA performed on the reaction time data also indicated a significant main effect of performance assessment. As indicated in Table 2, mean reaction was higher for the primary assessment than the joint assessment. Similarly, in the first ANOVA there was a significant effect of switching method, $F(4,64)=3.08$, $p = .022$. Subsequent analyses with a Bonferroni adjustment to the error rate indicated significantly slower ($p < .001$) reaction times for the Manual method ($M = 1504.75$, $SE = 105.9$) relative to the Automation-Joint ($M = 1375.42$, $SE = 102.3$) method.

As participants spend more time in automation, though, there is an increased likelihood of an accidental response while in automation (i.e., mode error). Thus, in order to make comparisons across conditions which differed in the number of trials in automation, the proportion of mode errors were calculated (i.e., the number of accidental responses while in automation divided by the total number of display updates while in automation). Despite this adjustment, the first analysis which included that Manual condition indicated a significant main effect of switching method, $F(4,64) = 10.378$, $p < .001$, on the proportion mode errors. Analytic t-tests with a Bonferroni adjustment to the error rate ($p \leq .005$) indicated that the proportion of mode errors in the Manual condition ($M = .014$, $SE = .006$) was significantly lower than the other conditions.

However, as shown in Table 2, when there was a higher number of trials in automation given machine-initiated transfers there was not a corresponding increase in the number of mode errors. An ANOVA performed on the proportion of responses while in automation indicated that there was only a significant effect of switching method. There were more mode errors given transfers to both automation and manual control relative to automation alone, despite the lower number of trials in automation. Moreover, the increased number of trials in automation for the joint assessment relative to primary assessment did not produce a significant change in the number of mode errors.

For the secondary task measures there were some additional significant effects for the number of trials in automation and the reaction time measures. First, ANOVA results indicated that there was a significant main effect of task difficulty on number of trials in automation, $F(2,32) = 41.51$, $p < .001$. Analytic t-tests with a Bonferroni adjustment to the error rate indicated that the number of trials in automation increased with each increment in task difficulty ($p \leq .016$), ranging from low ($M = 32.96$, $SE = 5.01$), to medium ($M = 41.93$, $SE = 4.36$), to high ($M = 49.62$, $SE = 4.06$). There was also a main effect of time constraint on participants' reaction times, $F(1,16) = 8.54$, $p = .01$. Mean reaction time was higher in the absence of a constraint ($M = 1400.97$, $SE = 93.30$) than in the presence of a constraint ($M = 1582.89$, $SE = 133.78$).

These effects support expectations. As performance declines on the primary task with increases in primary task difficulty, there should be an increased reliance on automation. In addition, there should be faster reaction times when there is a constraint on performance. If these effects were not present, there would be questions regarding the effectiveness of the manipulations of primary task difficulty and the constraint on performance.

Finally, there were four significant interactions, as well. The interaction between the switching method and difficulty for the number of trials in automation was statistically significant, $F(2,32) = 11.76$, $p < .001$. Figure 2 depicts the mean number of trials in automation for the two switching methods as a function of task difficulty. Analytic t-tests with a Bonferroni adjustment to the error rate ($p < .01$) indicated that for transitions to automation only, the number of trials in automation increased with each increment in task difficulty. However, for transitions in both directions there was a difference between the low difficulty level and the next two levels. (This interaction emerged in the first analysis as well, producing similar results. However, there was no effect of task difficulty for the Manual condition.) Thus, transitions to automation only seem to be more sensitive to variations in primary task difficulty than transitions to both automation and manual control.

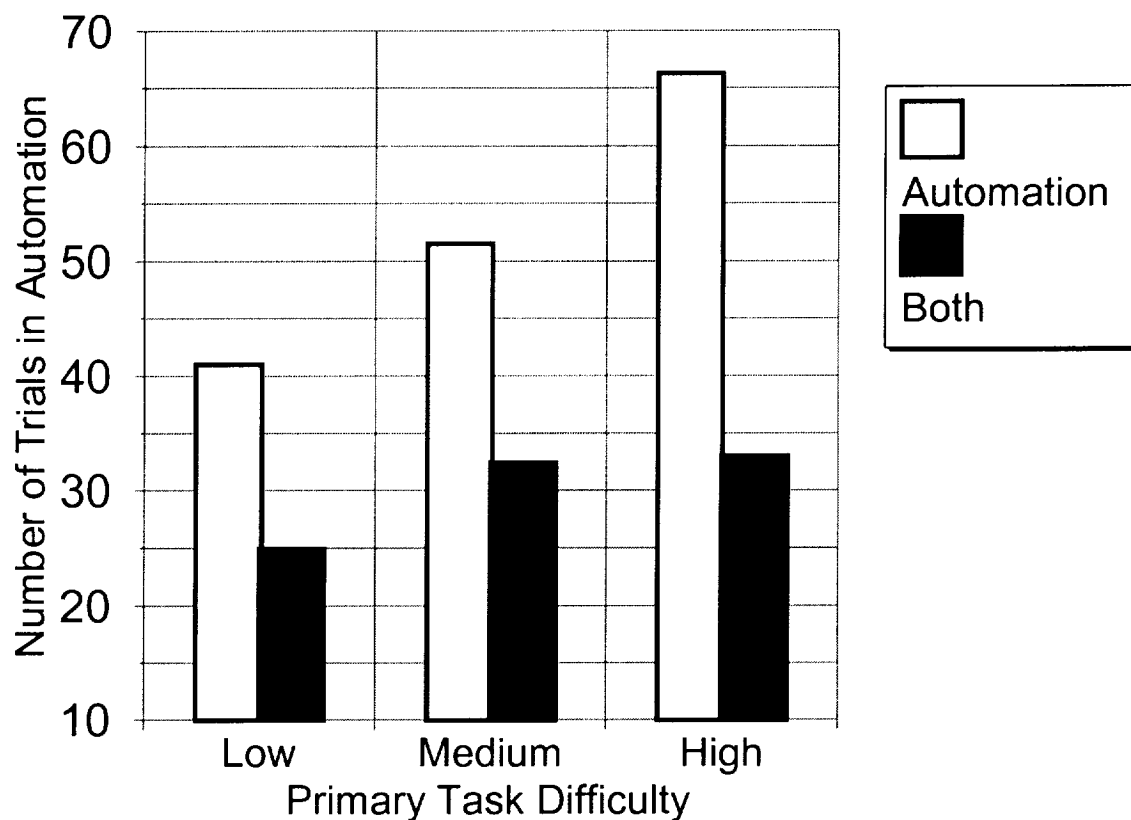


Figure 2. The mean number of trials in automation as a function of task difficulty given transitions to automation only (Automation) or transitions to both automation and manual control (Both).

Also, as depicted in Figure 3, the number of trials in automation was high for both primary and joint assessment for transitions to automation only. Alternatively, there was a difference in the number of trials in automation between the two assessment methods for transitions in both directions. The interaction between switching method and assessment method

was statistically significant, $F(1,16) = 8.98$, $p = .009$. Moreover, analytic tests indicated that the reliance on automation is more affected by type of assessment (primary vs. joint assessment) for transitions in both directions than transitions to automation alone ($p < .001$).

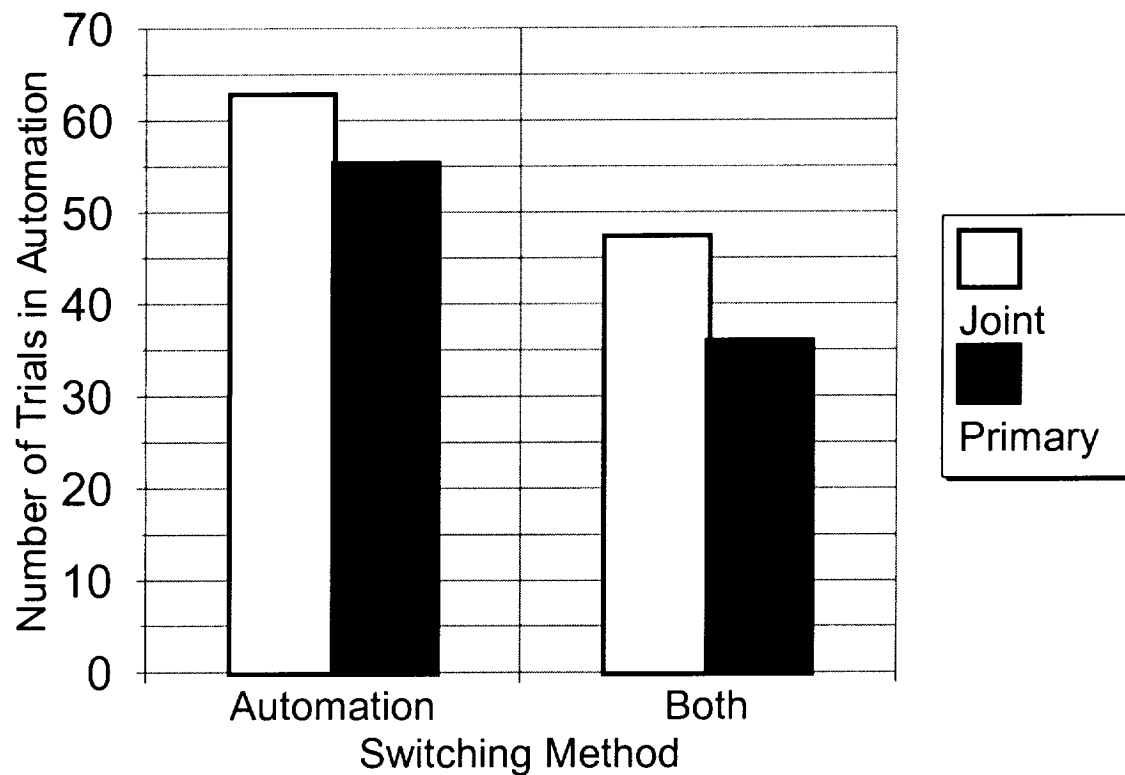


Figure 3. The mean number of trials in automation for transitions to automation only (Automation) or transitions to both automation and manual control (Both) given the two performance assessment methods, Primary task performance alone or a Joint assessment of primary and secondary performance.

The interaction between the switching method and the assessment method was also statistically significant for the hit-to-signal ratio, $F(1,16) = 9.38$, $p = .007$. Figure 4 depicts the mean hit-to-signal ratios for the two different switching methods and the two assessment methods. Again, analytic tests indicated that the hit-to-signal ratios were more affected by type of assessment (primary vs. joint assessment) for transitions in both directions than transitions to automation alone ($p < .001$). The higher hit-to-signal ratios for the joint method, relative to the primary assessment method, probably reflect differences in reliance on automation.

Finally, there was also a significant, $F(1,16) = 6.18$, $p = .024$, interaction between the time constraint and the switching method for mean reaction time. The mean response times for

the two switching methods when the constraint was present and absent are shown in Figure 5. Analytic tests indicated that there was no effect of switching method when the time constraint was absent. However, there were faster response times for shifts to Automation only relative to the shifts in both directions when the constraint was present, based on the same analytic t-tests with a Bonferroni adjustment to the error rate, $p \leq .005$. (The same interaction and pattern of results was present in the first analysis with the manual condition included.) Again, the difference in the reaction times is probably linked to differences in the number of trials in automation. There were no other significant effects.

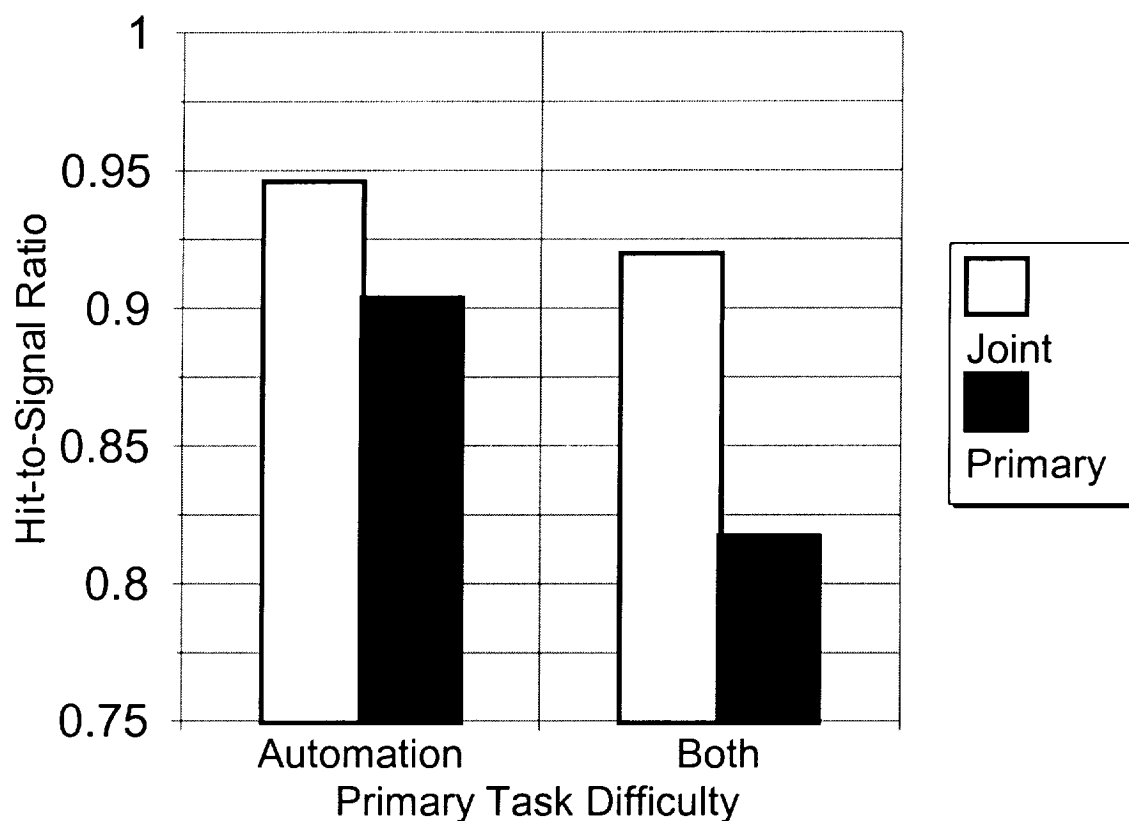


Figure 4. The mean Hit-to-Signal ratios for transitions to automation only (Automation) or transitions to both automation and manual control (Both) given the two performance assessment methods, Primary task performance alone or a Joint assessment of primary and secondary performance.

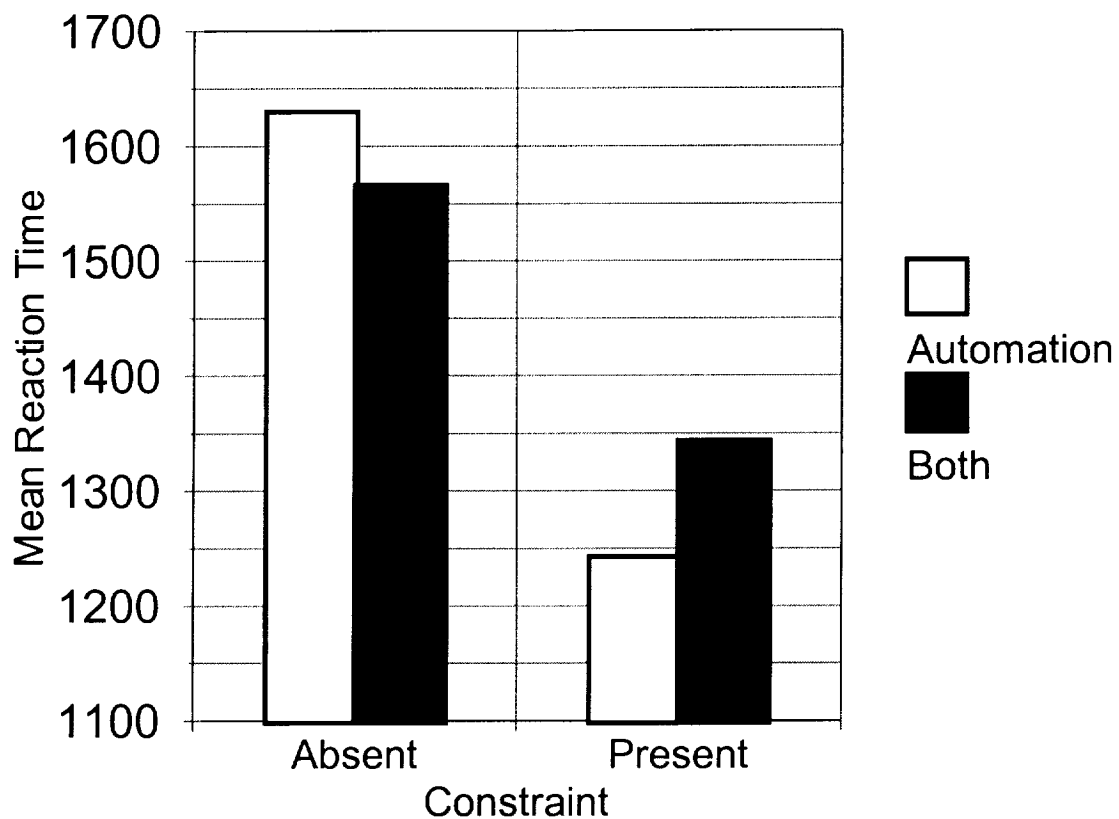


Figure 5. Mean reaction time measures under the absence and presence of a constraint on performance, for transitions to automation only (Automation) or transitions to both automation and manual control (Both).

Conclusions

As the load on the primary task increased performance on both primary task measures and the total points accumulated declined. As a result, the increase in the load resulted in more time spent in automation for the secondary gauge task. Correspondingly, this resulted in no direct impact of primary task difficulty on the secondary task measures (i.e., the hit-to-signal ratio remained high and mean reaction time remained relatively fast across the levels of primary task difficulty.) This suggests that as the load increased the secondary task was relinquished to the computer and the operator was able to direct more attention to the primary task.

Greater attention to the primary task when the secondary task was automated did not translate into improved control of the primary task, though. Instead, greater reliance on automation mainly affected performance on the secondary task measures (higher hit-to-signal ratios and faster reaction times). Moreover, the improved performance on the secondary tasks with an increased reliance on automation did not positively affect the total points accumulated, either. As a matter of fact, the effect of switching method on the total points accumulated, observed in the first analysis, showed the opposite trend. There were significantly more points accumulated for the Manual condition, which relied very little on automation, relative to the Automation-Joint method, which relied heavily on automation.

As suggested earlier, the high point accumulation probably reflects participants' efforts to accumulate as many points as possible. The point structure was set up such that the more time one spent in automation the fewer points he could accumulated from resetting the secondary gauge task. Thus, if the point structure were altered to remove the penalty and award points for all correct gauge resets (manual or automated), participants' points in the conditions that produced a greater reliance on automation should be higher. To test this idea, the total points were adjusted to award points for correct gauge resets during automation. The adjustment involved, first, estimating the number of critical gauge events reset by the machine by multiplying the number of trials in automation by the probability of a critical gauge event (.42). The estimated number of gauge resets were then multiplied by 20 and this product was added to the total points in all conditions for each participant.

Table 3. Total points accumulated and the Adjusted points accumulated.

Switching Method	Total Points		Adjusted Points	
	<u>M</u>	<u>SE</u>	<u>M</u>	<u>SE</u>
Manual	1713.92	210.17	1871.95	186.39
Automation				
Joint	1367.36	193.35	1868.95	159.55
Primary	1569.95	209.35	1933.35	164.14
Both				
Joint	1543.60	193.48	1943.81	157.31
Primary	1632.86	235.74	1777.17	211.05

The total points accumulated and the adjusted values are listed in Table 3. As evidenced in the table, the adjustment to the points produced a different pattern of results. The initial

assessment of the point accumulation showed no significant differences among the machine-based transfers. (Although, there was a tendency to gain more points when primary assessment was used and transitions were to both automation and manual control.) The analysis of the adjusted points, however, showed a main effect of difficulty, $F(2,32) = 90.21$, $p < .001$, as found before. In addition, a significant interaction emerged for the switching method and the assessment method, $F(1,16) = 5.75$, $p < .03$. As evidenced in Table 3, there was no difference between the two assessment methods for the transition to automation only. However, analytic tests indicated that there was a difference in the adjusted points accumulated between the primary and joint assessment methods given transitions in both directions ($p < .01$). Finally, the Manual condition is no longer the leader in total points accumulated.

Altogether, the machine-based transfers that tend to produce the higher adjusted point values are the ones that tend to place greater reliance on automation and, thus, better performance on the secondary task measures. This tendency for better performance given greater reliance on automation is consistent with previous pilot data (Montgomery & Gronert, 1999). Moreover, similar to the pilot data switching methods producing transfers to automation only tended to produce high performance on variety measures, but unlike previous pilot data there were not consistent advantages for the joint assessment method over the primary task assessment method.

Transfers to both automation and manual control did, however, show a dependence on the type of assessment in the predicted direction. Transitions in both directions that use a joint assessment method (Both-Joint) produce a higher reliance on automation, higher hit-to-signal ratios, and higher adjusted points relative to transitions in both directions using primary task assessment. Moreover, despite fewer trials in automation given the Both-Joint method relative to conditions with transitions to automation only (see Figure 3), this condition produced the highest adjusted point value (see Table 3) and a relatively high hit-to-signal ratio (see Figure 4), providing some support for the idea that there may be advantages to assessing both overload and underload in the operator.

Finally, there was a possible repercussion for relying on automation of the secondary gauge task. The more time one spent in automation the greater the likelihood of an accidental response while in automation (i.e., a mode error). This effect emerged in the first analysis when the manual condition was compared to the other methods despite the adjustment for reliance on automation. However, in the second analysis comparing the machine-based methods, the opposite pattern emerged. When there were transition in both directions there were significantly fewer trials in automation, but significantly more mode errors relative to transitions to automation only.

This same effect does not emerge, however, for the performance assessment methods. Even though participants spend more time in automation given a joint assessment, this did not lead to a corresponding increase in the number of mode errors. Thus, the difference observed given transitions in both directions may have something to do with the greater cycling between modes that would occur when transfer occur in both directions. However, since primary task assessments should also show greater cycling than the joint assessment method, greater cycling

may not be the cause of the difference. Instead, the fact that participants were simply informed of shifts to automation rather than being directly involved in the transfers would account for increased confusion about the current mode, leading to accidental responses in automation.

Experiment 2

In the second experiment, operators performed the same tasks described in Experiment 1 and similar conditions were included to test the experimental reliability of effects found in the first experiment. More importantly, the goals of the second experiment were (1) to assess the effect of the level of operator involvement on system performance and situation awareness and (2) to compare the effectiveness of two performance criteria used for machine-initiated mode transfers on system performance. Among the performance criteria tested, an absolute threshold method was compared to a method that required a change in operator performance.

Similar to the first experiment, mode transfers were either machine-initiated or human-initiated. Moreover, machine-initiated transfers depended on operator performance on both the primary and secondary tasks (i.e., the joint performance assessment method). The joint method of assessment was used in the second experiment since this approach yielded consistently high performance on the secondary task measures in the first experiment whether the transfer was to automation only or to both automation and manual control. The primary task performance assessment did not produce the same consistent results.

Besides the machine- and human-initiated transfers an additional hybrid method was tested in the second experiment. This hybrid method used the same performance threshold criteria as the machine-initiated method to signal the operator when a transfer should occur, but the actual transfer was manually controlled. Moreover, for the hybrid method a signal indicated the appropriate time to transfer to automation as well as when to return to manual control. Thus, this condition was similar to the Both-Joint condition found in the first experiment, except the human initiated the transfers after being signaled rather than being simply informed of a change.

Based on the results from the previous experiment, it was expected that when mode transfer were completely human-initiated operators would probably attempt to maintain control of the secondary task. Thus, the likelihood of a mode error would be minimal and operator performance would suffer on the secondary task measures. Similarly, machine-initiated mode transfers were expected to produce greater reliance on automation and improved performance on the secondary tasks, at the expense of increased mode errors. Finally, it was expected that the hybrid method would produce similar results as found under the Both-Joint condition in the first experiment (i.e., relatively good performance on the secondary task measures). However, it was also expected that mode awareness would be improved since the operator was more actively involved in mode transfers than found in the previous experiment.

Previous research indicates that situation awareness improves with increased operator involvement in the tasks. For example, Endsley and Kiris (1995) assessed operators' situation awareness after automation failure under five different levels of operator involvement in

automobile navigation task. The operators either performed the task with no assistance from an expert system or with some level of decision support (ranging from recommendations provided by an expert system to full automation where the system did everything and the operator monitored events). At some point during the task, the expert system failed and the operator had to perform the tasks manually during subsequent scenarios. They found that when the operator was at least partially involved in the decision making prior to the failure, situation awareness remained relatively high. Thus, similar to the Endsley and Kiris (1995) study, it was expected that greater operator involvement in the hybrid conditions would reduce the likelihood of mode errors (reflecting greater situation awareness).

Among the hybrid and machine-initiated methods, two criteria for initiating mode transfers were also tested. The absolute performance threshold criterion used in Experiment 1 was compared to a method that required a continued change in performance beyond the threshold level. With the former criterion, the operator was transferred to automation or signaled to transfer to automation as soon as performance on the various sub-tasks performed by the operator passed specified threshold values. Transfers back to manual control were either human-initiated or occurred when performance returned to acceptable levels, similar to the Automation-Joint and Both-Joint conditions found in experiment 1, respectively. The second criterion was similar to the absolute threshold method, but required a continued change in performance after surpassing the absolute threshold value before automation was invoked.

Freeman, Mikulka, Scerbo, and Hadley (1998) compared an absolute measure of arousal derived from participants' electroencephalogram (EEG) measures with a slope method that evaluated changes in arousal for their effectiveness as triggers for adaptive automation. It was expected that a method that maintains optimal participant arousal, by placing a tracking task in manual control or automation at the appropriate time, would produce better tracking performance under manual control. They found the participants' tracking performance was better under the slope method than under an absolute threshold method.

If the change criterion helps to maintain optimal levels of operator involvement in the current experiment, then system performance should be better given the change criterion relative to the absolute threshold criterion, as found in the Freeman et al. (1998) study. However, in this experiment the change in performance is relative to a threshold point. Thus, mode transfers will be less frequent given the change criterion than the absolute threshold criterion. If this reduces reliance on automation, then there may be deleterious effects on the system performance, but improved situation awareness (i.e., fewer mode errors) relative to the absolute threshold method, as found in experiment 1.

Method

Participants Twenty-two individuals were tested on a preliminary set of trials to identify those who were able to meet performance criteria described in the first experiment. Eighteen individuals (13 females and 5 males) were selected. Again, all participants were undergraduate students enrolled at Bradley University and they were paid minimum wage for their participation.

Procedure: The apparatus and experimental tasks were identical to the first experiment. The procedure was very similar, also, however some changes were made. First, primary task difficulty was manipulated again, but only medium and high levels of task difficulty were tested. Moreover, for medium task difficulty 3, 6, or 9 units were drawn and for high task difficulty the units were 5, 10, or 15.

Second, transitions between manual control and automation of the secondary gauge task were either human-initiated or machine-initiated (as found in the previous experiment). An additional hybrid condition was included as well. As described earlier, human-initiated mode transfers were implemented when the operator presses a key labeled “AU/M.” The machine-initiated transfers were controlled by the computer and operators were simply informed of the current mode, via a tone and a message on the display. For the hybrid conditions, however, the operator initiated the transfer by pressing the “AU/M” key, but the computer signaled the operator when to transfer modes.

A tone was used to signal the operator when a mode transfer should occur and the criteria used to signal the participant for the hybrid method were the same as those used for machine-initiated transfers, as described below. If the operator was signaled to switch modes, but chose to keep the task in the current mode despite the computer’s recommendations, subsequent signals to the operator were coupled with changes in the color of the visual display indicating the current mode. Initially, the display was white. If the operator did not respond to the signal and performance remained beyond the criteria for the current mode, the display turned blue. Subsequently, a lack of response to the computer’s recommendations led to the following color changes: green, yellow, and red.

Finally, the last change relative to the first experiment involved varying the criterion used for producing or signaling a mode transfer. For the machine-initiated and hybrid transfers the criteria were based on either an absolute threshold value or a continued change in performance. The absolute threshold criteria were the same criteria used for the joint assessment method found in the first experiment. That is, the secondary gauge task was automated when the reservoir levels exceeded the critical level indicators, error rates on the secondary gauge task were greater than 10%, and reaction times exceeded $M_{RT} + SD_{RT}$ for four consecutive display updates.

For the machine-initiated method transfers were only to automation. The participants could return to manual control at their discretion by pressing the key labeled “AU/M”. Thus, this condition was similar to the Automation-Joint method used in the previous experiment. For the hybrid method transfer were to both automation and back to manual control. Thus, this condition was similar to the Both-Joint method used in the previous experiment. The return to manual control occurred when performance returned to acceptable levels.

The second criterion tested required a performance change beyond the threshold value just described. Thus, transfers were initiated or observers were signaled to change modes only when the operator’s performance continued to decline beyond the threshold point for two

consecutive display updates. For example, if the participant allowed the reservoir levels to exceed the critical level for two consecutive display updates the gauge task would be automated.

Altogether, there were five methods for transferring modes, again. There was one manual method, where mode transfers were completely initiated by the participant. There were also four machine-initiated transfers, two used the absolute *Threshold* criterion and two used the *Change* in performance criterion. Two of the four machine-initiated transfers produced transfers to *Automation* only (i.e., the operator controlled transfers back to manual control) and two produced transfers to both automation and manual control (*Hybrid*).

As in experiment 1, all individuals who were interested in participating in the experiment began by completing a preliminary test set. Those who met the selection criteria were scheduled for five subsequent sessions. During the first session, participants learned how to control the three sub-tasks and they completed a training period for each switching method: Manual, Automation-Threshold, Automation-Change, Hybrid-Threshold, and Hybrid-Change. The order of presentation was varied, again, based on a Latin square.

During the remaining four sessions the participants completed monitoring periods, composed of a factorial combination of the three independent variables: task difficulty (medium and high), method (Manual, Automation-Threshold, Automation-Change, Hybrid-Threshold, and Hybrid-Change), and time constraint (present or absent). As in the previous experiment, the presentation order for the response constraint was counterbalance and the presentation order for the other conditions was randomized.

Results and Discussion

Similar to the first experiment, there was an incomplete factorial combination between switching method (Manual, Automation, and Hybrid) and performance criterion (Threshold and Change). Thus, ANOVAs were first performed comparing five switching methods: Manual, Automation-Threshold, Automation-Change, Hybrid-Threshold, and Hybrid-Change. Subsequently, the Manual condition was removed and ANOVAs performed, treating switching method and criterion as two separate variables. Analogous to the first experiment, most significant results found in the first analysis, with the Manual condition included, were replicated in the second analysis. Thus, the results from the second analysis are reported and when there is an exception the results of the first analysis are reported, as well.

Primary Task and Total System Performance

The means for the primary task performance measures and total points accumulated are reported in Table 4 for the two levels of primary task difficulty. There was a significant main effect of task difficulty on all three measures, $p < .001$. (The F values are also listed in Table 4.) Consistent with the first experiment, the number of times operators exceeded an unsafe level and the number of emergency resets were higher for greater task difficulty. Similarly, the total accumulated points were lower with higher task difficulty.

Table 4. Mean performance measures for the primary task and total system performance by task difficulty.

Primary Task Difficulty	Number of Critical Reservoir Events		Number of Emergency Resets		Total Points	
	<u>M</u>	<u>SE</u>	<u>M</u>	<u>SE</u>	<u>M</u>	<u>SE</u>
Medium	10.60	1.68	0.49	0.15	1699.75	142.94
High	15.70	1.59	1.18	0.21	1254.30	134.80
E(1,17)	78.70		45.31		120.75	

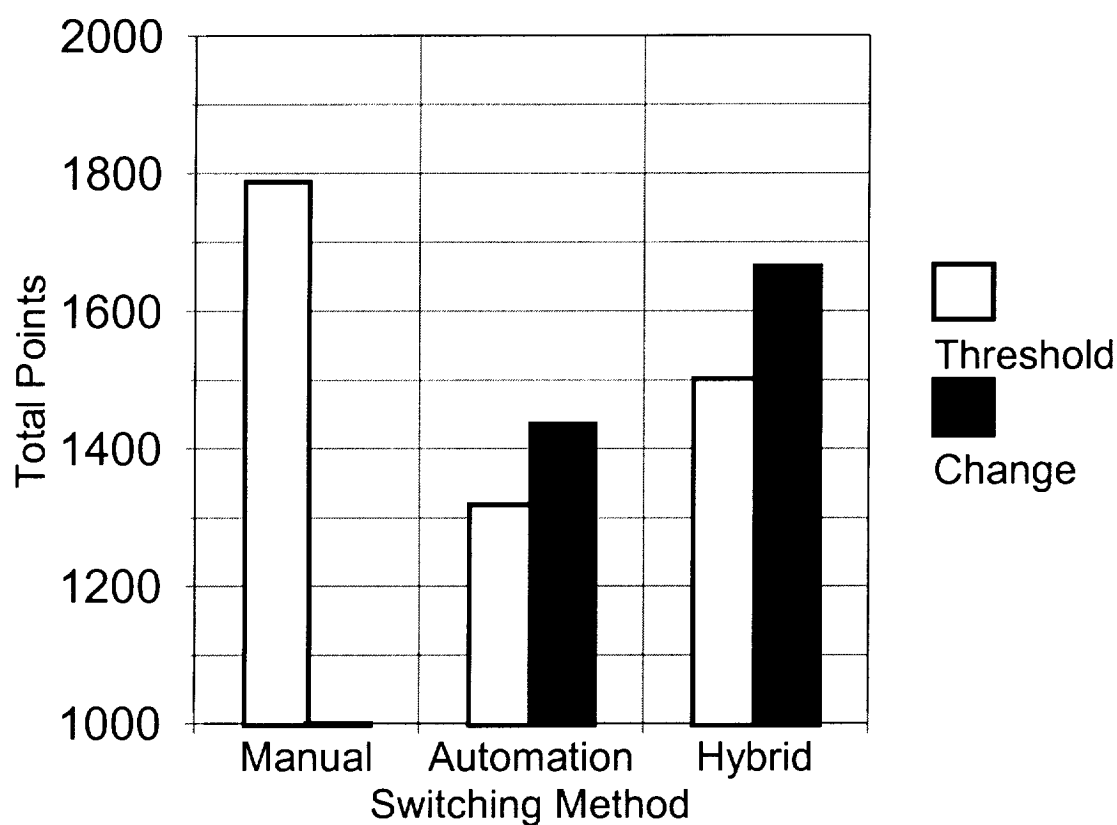


Figure 6.

Total points accumulated for transitions that were completely human-initiated (Manual), machine-initiated transfers to automation with a human-initiated return to manual control (Automation), or human-initiated transfers to both automation and manual control after being signaled by the computer (Hybrid). The gray bars represent conditions where the change in performance criterion was used and the white bars represent conditions using the absolute threshold criterion, except for the Manual condition where the criterion was not relevant.

There were no other significant main effects for the number of critical events or the number of emergency resets. However, there was a significant main effect of the criterion for the total points accumulated, $F(1,17) = 12.88$, $p = .002$. The total points accumulated was higher given the criterion that required a change in performance beyond threshold ($M = 1576.76$, $SE = 126.26$) relative to the threshold criterion ($M = 1377.28$, $SE = 152.92$). Moreover, there was also a main effect of switching method, $F(4,68) = 6.78$, $p < .00$, in the first ANOVA which included the Manual condition. The means for the five switching methods are reported in Figure 6. Analytic t-tests, with a Bonferroni adjustment to the error rate, indicated that the Manual and the Hybrid-Change switching methods produced significantly more points compared to the machine-initiated methods including the threshold criterion ($p \leq .003$).

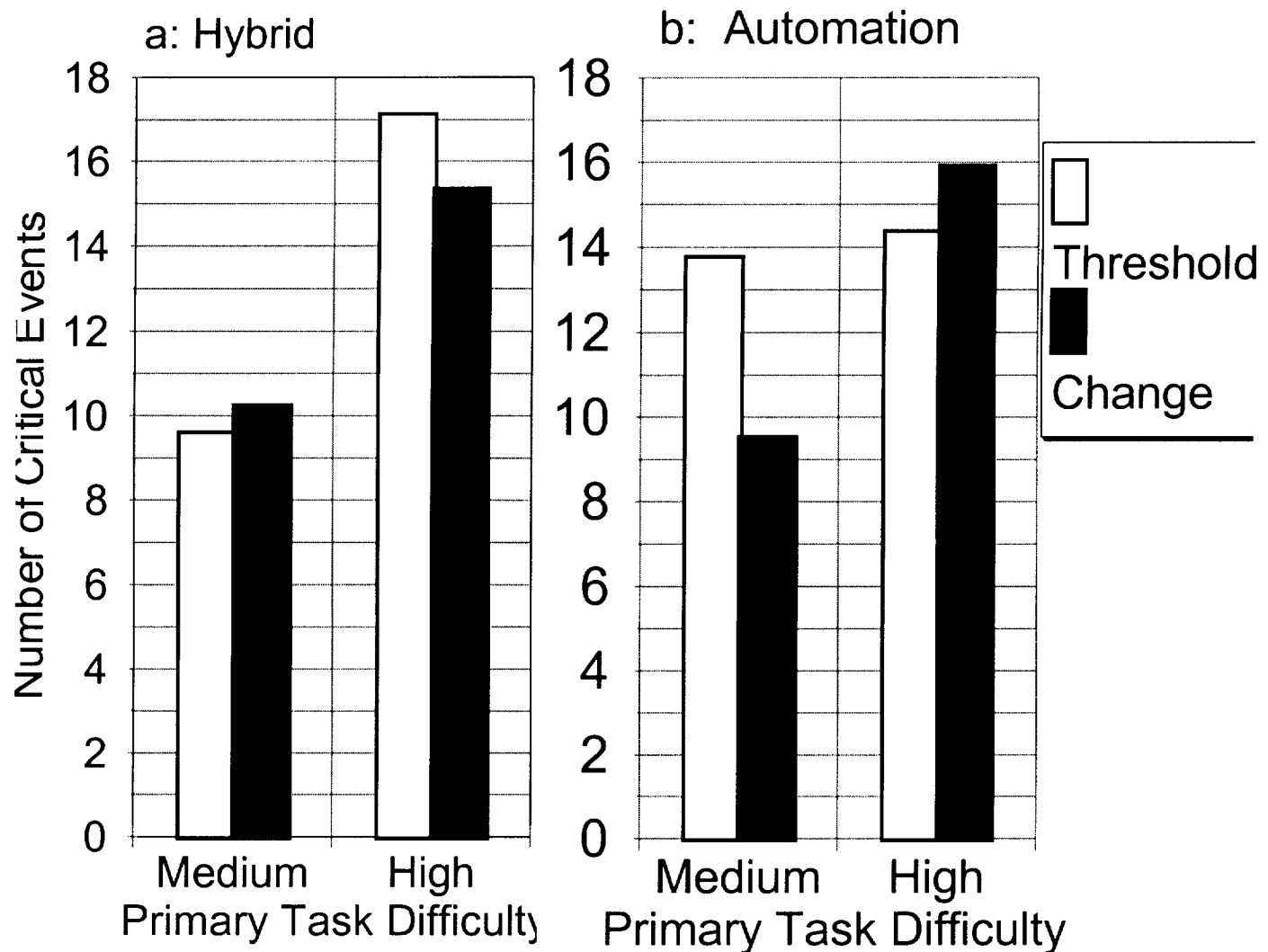


Figure 7. Number of critical events for transitions to both automation and manual control (panel a) and to automation only (panel b) given primary task difficulty and the two performance criteria.

Finally, among the primary task measures there were some significant interactions involving primary task difficulty. First, for the number of critical events there was a significant three-way interaction between the switching method, the criterion, and primary task difficulty, $F(1,17) = 6.80$, $p = .018$. Figure 7 depicts the mean number of critical events for the two switching methods (Hybrid and Automation in panels a and b, respectively) and the two assessment methods for the two levels of task difficulty. Analytic tests indicated that there was a significant ($p \leq .003$) increase in the number of critical events for all conditions, except for the Automation-Threshold condition in which the number of critical events is relatively high at both levels of task difficulty.

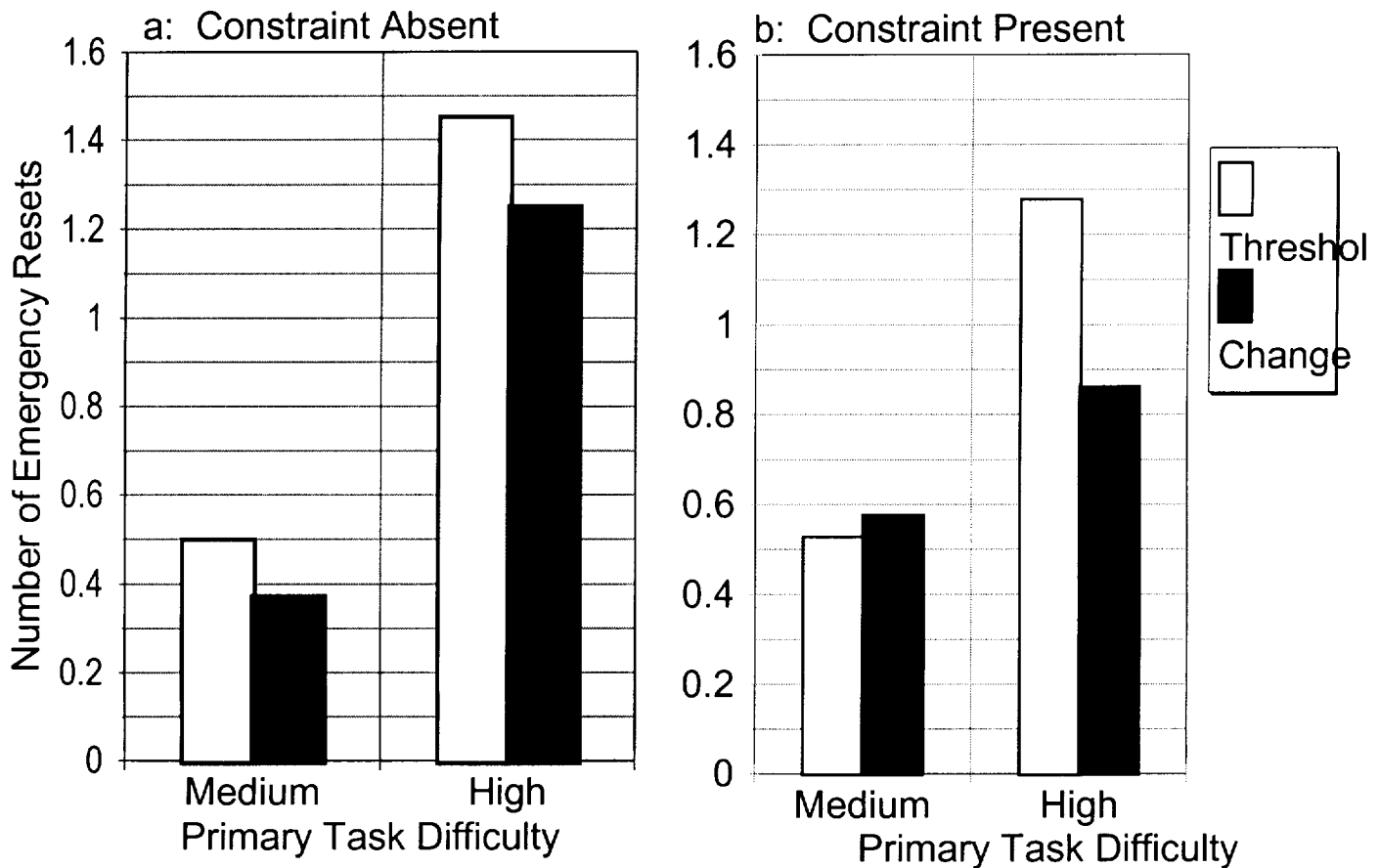


Figure 8. Mean number of emergency resets when the constraint was either present (panel a) or absent for the threshold and change in performance criteria as a function of primary task difficulty.

There was also a significant three-way interaction between primary task difficulty and the constraint and criterion variables, $F(1,17) = 5.77$, $p = .028$, for the number of emergency resets. Subsequent analytic tests indicated that the interaction between the criterion and task difficulty variables was not significant in the absence of a constraint. (See Figure 8a.) However, there was a significant interaction in the presence of a constraint, $F(1,17) = 6.69$, $p = .019$. As shown in Figure 8b, there was a significant increase ($p < .001$) in the number of emergency resets as

difficulty increased for the threshold criterion. However, the effect of difficulty was much weaker (statistically not present) for the change criterion.

In summary, analogous to the first experiment, there is an effect of primary task difficulty on the primary task measures in the expected direction (i.e., more critical events and emergency resets and fewer points with increased task difficulty). In addition, more points were accumulated given a Manual switching method relative to a machine-initiated switch to automation only, given performance surpassing threshold values on either the primary or secondary tasks (i.e., Automated-Threshold condition). However, in this experiment there is an advantage for the Manual method relative to the hybrid-threshold method, as well.

Moreover, there appears to be an advantage for a criterion that requires changes in performance beyond the threshold value. First, more points were accumulated under a criterion that required a continued change in performance relative to surpassing a threshold only. Similarly, given the appropriate circumstances there are fewer critical events and emergency resets and when the criterion that required a continued change in performance was used rather than the threshold criterion.

In the first experiment, conditions that yielded higher total point accumulations tended to be conditions that also showed relatively low reliance on automation. Performance on the primary task was comparable for all conditions in Experiment 1, but less reliance on automation resulted in lower hit-to-signal ratios and slower reaction times. A similar pattern appears to be present in this experiment as well, as found among the secondary task measures.

Secondary Gauge and Reset Tasks First, the means for the secondary task measures (hit-to-signal ratio, number of trials in automation, and mean reaction time) for the assessment criteria, the computer based switching methods, and the presence or absence of a constraint are reported in Table 5. In general, the more participants relied on automation, the higher their hit-to-signal ratios and the lower the mean reaction times for all three variables. There was one exception, though. In the absence of a constraint, the hit-to-signal ratio was higher even though participants spent fewer trials in automation than when a constraint was present. ANOVAs performed on the secondary task measures indicated significant main effects of criterion, switching method, and constraint (F values are reported in the table) for all three measures, with one exception. The effect of criterion on the mean reaction time measure was only marginally significant.

Analyses performed with the Manual condition included produced similar results. There was a main effect of constraint on all three secondary task measures and the pattern among the means is the same as reported in Table 5. Similarly, there was a significant main effect ($p < .001$) of switching method for all three measures, $F(4,68) = 10.67$, $F(4,68) = 29.59$, and $F(4,68) = 9.12$, respectively. In general, analytic analyses with a Bonferroni adjustment to the error rate indicated that the manual switching method produced significantly ($p < .001$) fewer trials in automation ($M = 20.46$, $SE = 4.99$) than the other conditions and significantly ($p \leq .003$) lower hit-to-signal ratios ($M = .86$, $SE = .02$) than the Automated-only switching methods. Moreover, the manual method produced significantly ($p < .005$) greater reaction times ($M = 1393.5$, $SE = 110$) than the Automation-only switching methods.

Table 5. Mean performance on the secondary task measures.

	Hit-to-Signal Ratio		Number of Trials in Automation		Reaction Time	
	<u>M</u>	<u>SE</u>	<u>M</u>	<u>SE</u>	<u>M</u>	<u>SE</u>
Constraint						
Present	.914	.01	58.47	4.55	1126.12	65.98
Absent	.932	.008	43.75	6.14	1376.58	125.06
$F(1,17)$	5.12 ($p = .037$)		10.81 ($p = .004$)		6.90 ($p = .018$)	
Criterion						
Threshold	.931	.01	57.26	4.89	1231.15	86.21
Change	.915	.008	44.96	5.09	1271.55	90.62
$F(1,17)$	9.04 ($p = .008$)		51.85 ($p < .001$)		4.16 ($p = .057$)	
Switching Method						
Automated	.951	.004	62.80	4.58	1189.0	83.76
Hybrid	.895	.017	39.42	5.96	1313.24	96.0
$F(1,17)$	11.02 ($p = .004$)		33.85 ($p < .001$)		9.78 ($p = .006$)	

Finally, there was one additional main effect among the secondary task measures that involved effects of primary task difficulty. As in the first experiment, there was a significant effect of primary task difficulty, $F(1,17) = 5.62$, $p = .03$, on the number of trials in automation. The number of trials in automation increased as task difficulty increased from medium ($M = 48.76$, $SE = 5.24$) to high ($M = 53.46$, $SE = 4.82$) difficulty. There were no other significant main effects. However, there was one three-way interaction among constraint, switching method, and difficulty for the mean reaction time measure, $F(1,17) = 5.83$, $p < .03$.

The mean reaction time values are depicted in Figure 9. As suggested by the figure, analytic tests indicated that the interaction between switching method and difficulty was not statistically significant when the constraint was present. However, there was a significant effect in the absence of a constraint, $F(1,17) = 4.95$, $p = .04$. For the automated switching method there was no difference between participants reaction times at the two levels of task difficulty for transfers to automation only, but there was a statistically significant increase in reaction time as primary task difficulty increased for the hybrid method, ($p < .02$).

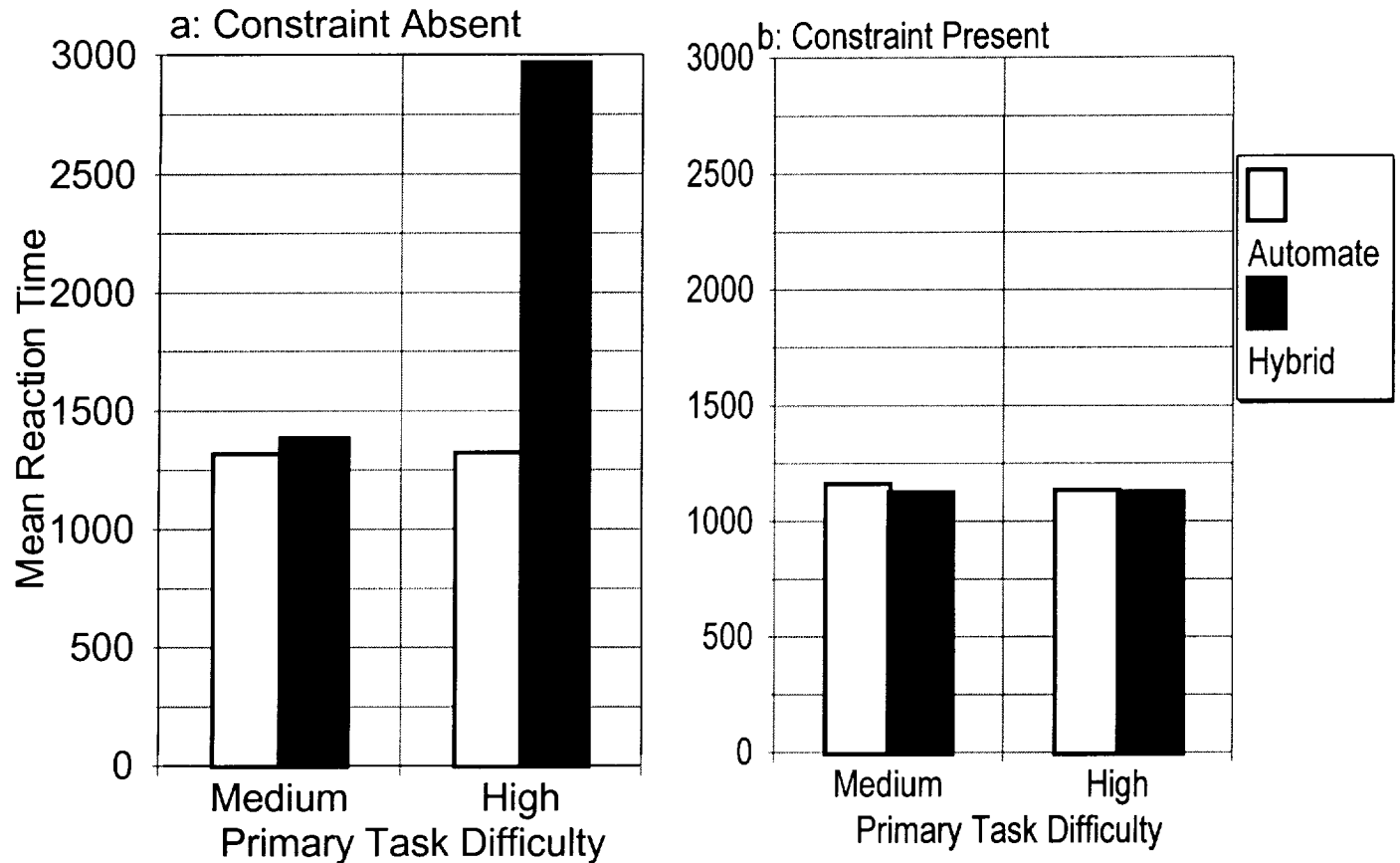


Figure 9. Mean reaction time when the constraint was either absent (panel a) or present (panel b) for human-initiated transfers to both automation and manual control (Hybrid) or automation only as a function of primary task difficulty.

In summary, consistent with experiment 1, the conditions yielding the higher point accumulation corresponded with the conditions with the lower reliance on automation. Moreover, those conditions relying on automation less tended to produce lower hit-to-signal ratios and higher reaction times (i.e., manual transitions, transitions based on changes in performance beyond threshold, and hybrid methods). Likewise, as in the first experiment, there may be a lower proportion of accidental responses in automation with lower reliance on automation as well.

Once more, to test the impact of automation on mode errors the number of accidental responses while in automation were, first, converted to a proportion of the total trials in automation. Consistent with the results from experiment 1, the Manual condition, which produced the lowest number of trials in automation, produced a lower proportion of accidental responses while in automation ($M = .011$, $SE = .006$) than any of the other conditions. However, the difference among the switching methods was only significant under high primary task difficulty (i.e., a significant interaction between switching method and primary task difficulty, $F(4,68) = 3.17$, $p < .02$).

However, inconsistent with the first analysis, the second ANOVA, with the Manual condition removed, did not show a significant difference in the mode errors between the conditions with transitions to automation only and the conditions with transitions to automation and manual control. In experiment 1, transitions to automation only produced a higher reliance on automation, but fewer mode errors, than transitions to both automation and manual control. It was suggested that transitions in both directions produced greater confusion about the current mode and, thus, a higher number of accidental responses while in automation. Since there is no difference between these conditions in this study, signaling the operators to initiate the transitions rather than simply informing them of the transition may have reduced the errors.

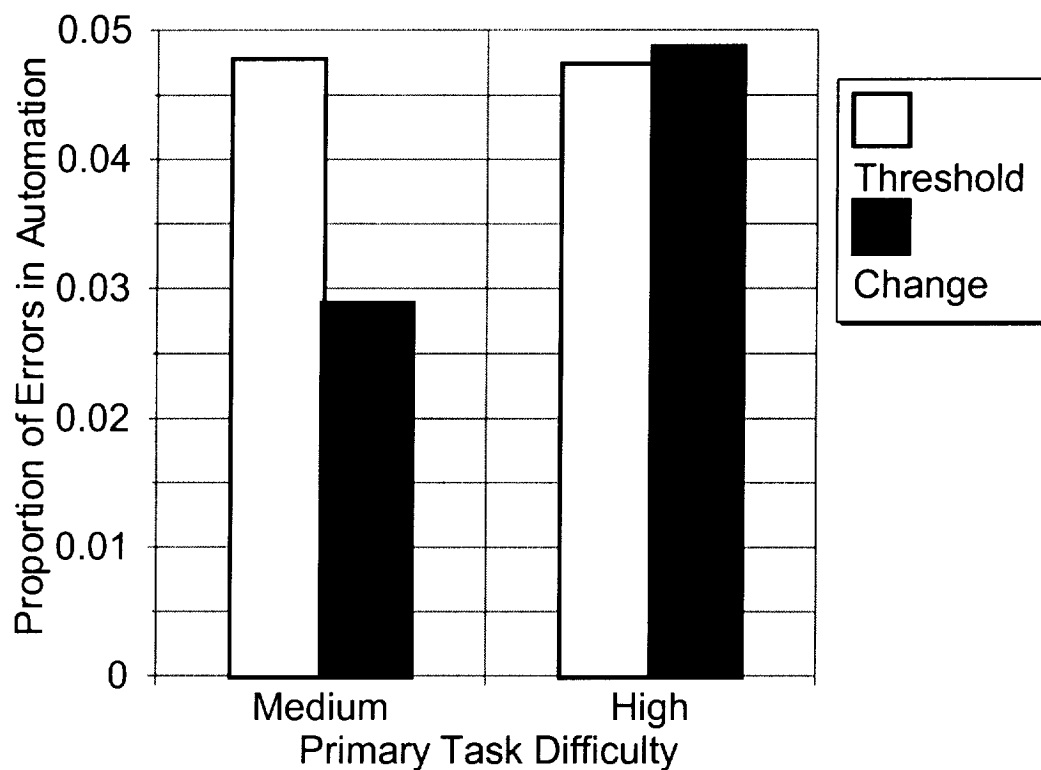


Figure 10. The proportion of accidental responses to the gauge task while the task was automated for the Threshold and Change in performance criteria as a function of primary task difficulty.

In both analyses, there were no other significant main effects, only one significant interaction involving primary task difficulty. For the second analysis, there was statistically significant interaction between criterion and primary task difficulty, $F(1,17) = 6.64$, $p = .02$. The mean proportion of accidental responses while in automation for the two criteria are reported in Figure 10 for the two levels of primary task difficulty. As suggested by the figure, analytic tests indicated that there was no effect of task difficulty for the threshold criterion. However, there were significantly fewer errors ($p < .005$) for the medium task difficulty when the change

criterion was used. Fewer mode errors given the change criterion at medium task difficulty is probably linked to a lower reliance on automation, since lower task difficulty and use of the change criterion both tend to produce less reliance on automation.

Conclusions

Resembling the first experiment, the number of errors on the primary task measures increased and the total points accumulated decreased as primary task difficulty increased. Similarly, as the load on the primary task increased there was a greater reliance on automation, eliminating the effect of primary task difficulty on the secondary task measures. Moreover, there was a tendency for a higher reliance on automation to yield better performance on the secondary task measures (i.e., the hit-to-signal ratios and mean reaction times) as in experiment 1.

Also similar to experiment 1, there were significantly more points accumulated under conditions where participants attempted to maintain control of the secondary gauge task (i.e., the Manual condition and those conditions including assessment of changes in performance). An adjustment was made to award points for gauge resets during automation in the second experiment similar to that described earlier for experiment 1. This adjustment eliminated any significant effects of switching method and criterion on points accumulated. However, there was still a tendency for higher point accumulation under the Manual condition and those using a criterion that required a change in performance relative to the conditions using a threshold criterion only.

Those conditions requiring a change in performance not only showed a tendency for higher point accumulation, but other advantages in terms of the primary task measures. Under the appropriate conditions, there were fewer emergency resets and fewer critical events in controlling the primary task when the criterion required a change in performance beyond threshold. These advantages of the change in performance criterion on the primary task measures are consistent with the Freeman et al. (1998) results. Their participants' performance on a primary tracking task in an adaptive system was better with a slope method (assessing changes in arousal) than given absolute threshold method. The hit-to-signal ratios in the current study did, however, reflect lower secondary task performance for the criterion involving changes in performance relative to the threshold criterion, but the actual difference was fairly low (.016) and the difference in the reaction times between the two assessment methods was not statistically significant.

Finally in the first experiment, there was an advantage for transitions to automation only rather than transitions in both directions (i.e., the hybrid method in Experiment 2) for the hit-to-signal ratios and reaction time measures. This same effect was present in the second experiment and dependent on a heavy reliance on automation. Moreover, there were no particular performance advantages for the Hybrid-Threshold method as observed under the comparable Both-Joint condition in Experiment 1. Thus, advantages related to assessment of both overload and underload were not present in Experiment 2. Also, unlike experiment 1 the difference in the number of mode errors between transitions to automation only and transitions in both directions

was not present. It was suggested that signaling the operators to initiate the transitions rather than simply informing them of the transition could have reduced the errors.

Finally, there was a significantly lower proportion of mode errors when the criterion required a change in performance for medium primary task difficulty compared to high primary task difficulty. The differences observed here probably have little to do with cycling between modes, since the threshold criterion would be more likely to produce greater cycling than the change criterion which has more strict requirements for a mode transition. Instead, differences observed here probably reflect differences in time spent in automation. Participants tend to rely on automation less when the change in performance criterion is used and when the primary task difficulty is relatively low.

General Conclusions

First of all, there was substantial evidence that the primary task load and the time constraint had an impact on performance in both experiments. As the load on the primary task increased performance on the primary task measures and the total points accumulated declined, and more time was spent in automation for the secondary gauge task. Similarly, a time constraint on performance reduced participants' reaction times, increased their reliance on automation, and in the second experiment resulted in a lower hit-to-signal ratio. More importantly, though, there was evidence that some approaches for adaptive task transfers were more effective than other methods.

From the first experiment, the evidence indicated that machine-initiated transfers to automation with a human-initiated return to manual control produced better performance on the secondary task measures relative to machine-initiated transitions to both automation and manual control. In addition, a machine-initiated transition that considered both primary and secondary task performances yielded better operator performance on the secondary task measures and higher adjusted points relative to transition based on primary task performance alone. These gains tended to results from greater reliance on automation, though. Finally, despite the higher reliance on automation for machine-initiated transfers to automation only, this switching method produced a significantly lower proportion of mode errors compared to machine-initiated transitions in both directions.

In the second experiment, experimental reliability was demonstrated in that transfers to automation-only produced better performance on the secondary task measures than machine-initiated (by computer signal) transfers to both automation and manual control. However, involving the operator in making mode transfers after being signaled by the computer, rather than simply signaling the operator after the change, appears to have eliminated the advantage of the transfers to automation-only relative to transfers to both automation and manual control for the proportion of mode errors found in Experiment 1.

Furthermore, in experiment 2 there is, again, evidence that certain approaches to adaptive task transfers provide benefits relative to other methods. Under the appropriate conditions the

criterion that required changes in performance yielded advantages on the primary task measures, a higher total point accumulation, and a lower proportion of mode errors than a criterion based on an absolute threshold value. Moreover, the advantages for the change in performance criterion were gained without a heavy reliance on automation or a substantial loss on the secondary task measures.

Finally, there are some issues that should receive further attention in future research. First of all, there was essentially no evidence supporting performance advantages for a method that assessed both overload and underload. This contradicts evidence from vigilance studies (e.g., Parasuraman, Mouloua, & Molloy, 1996) which demonstrate the advantages of returning control to the operator given vigilance related declines in performance. In the current experiments, however, the monitoring periods were probably too brief (less than 10 minutes) to produce vigilance related declines in performance. Thus, other factors probably need to be considered regarding underload on the operator within the context of the tasks performed in these experiments.

In addition, differences among the switching methods for the proportion of mode errors found in experiment 1 and performance differences among the threshold criteria in experiment 2 could be linked to differences in the amount cycling between modes. Since heavy cycling can have deleterious effects on performance, future studies should address whether differences observed among the switching methods tested in these experiments are related to high reactivity related to the methods tested. Finally, in experiment 2 both criteria were relative to a threshold point. In one case a mode transfer depended on an absolute threshold value and in the other case transfers depended on changes in performance beyond this absolute threshold point. These criteria should be compared to a criterion based on changes in performance, alone, that are not dependent on a threshold point in the future as well.

References

- Byrne, E. A. & Parasuraman, R. (1996). Psychophysiology and adaptive automation. Biological Psychology, 42, 249-26
- Endsley, M. R. & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. Human Factors, 37, 381-394.
- Freeman, F. G., Mikulka, P. J., Scerbo, M. and Hadley, G. (1998). Comparison of two methods for use in adaptive automation. Perceptual and Motor Skills, 86, 1185-1186.
- Kaber, D. B. & Riley, J. M. (1999). Adaptive automation of a dynamic control task based on workload assessment through a secondary monitoring task. In Scerbo, M. W. and Mouloua, M. (Eds.) Automation technology and human performance: Current research and trends. (pp. 129-133). Mahwah NJ: Erlbaum.
- Montgomery, D. A. & Gronert, M. E. (1999) Effects of selected task performance criteria at initiating adaptive task reallocations. In Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting. (pp. 1389) Houston, TX
- Parasuarman, R., Mouloua, M. and Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. Human Factors, 38, 665-679.
- Rencken, W.D. & Durrant-Whyte, H. F. (1993). A quantitative model for adaptive task allocation in human-computer interfaces. IEEE Transactions on Systems, Man, and Cybernetics, 23, 1072-1090.
- Scerbo, M. W. (1996). Theoretical perspective on adaptive automation. In R. Parasuraman and M. Mouloua (Eds.) Automation and human performance: Theory and applications (pp. 37-63). Mahwah NJ: Erlbaum.
- Wickens, C. D., Gordon, S. E., & Liu, Y. (1998). An introduction to human factors engineering. New York: Addison-Wesley Longman. (pp. 392-395)