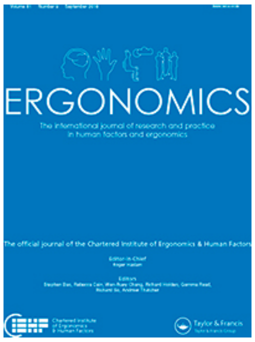

Ergonomics

ISSN: 0014-0139 (Print) 1366-5847 (Online) Journal homepage: http://www.tandfonline.com/loi/terg20

Level of automation effects on performance, situation awareness and workload in a dynamic control task

MICA R. ENDSLEY

Ergonomics, 1999, vol. 42, no. 3, 462 – 492

Level of automation effects on performance, situation awareness and workload in a dynamic control task

Mica R. Endsley† and David B. Kaber‡*

†SA Technologies, 4731 East Forest Peak, Marietta, GA 30066, USA

‡Department of Industrial Engineering, 125 McCain Building, Mississippi State University, Mississippi State, Mississippi 39762-9542, USA

Keywords: Level of automation (LOA); Human-centred automation; Situation awareness (SA); Workload; Out-of-the-loop performance

Various levels of automation (LOA) designating the degree of human operator and computer control were explored within the context of a dynamic control task as a means of improving overall human/machine performance. Automated systems have traditionally been explored as binary function allocations; either the human or the machine is assigned to a given task. More recently, intermediary levels of automation have been discussed as a means of maintaining operator involvement in system performance, leading to improvements in situation awareness and reductions in out-of-the-loop performance problems. A LOA taxonomy applicable to a wide range of psychomotor and cognitive tasks is presented here. The taxonomy comprises various schemes of generic control system function allocations. The functions allocated to a human operator and/or computer included *monitoring* displays, *generating* processing options, *selecting* an 'optimal' option and *implementing* that option. The impact of the LOA taxonomy was assessed within a dynamic and complex cognitive control task by measuring its effect on human/system performance, situation awareness and workload. Thirty subjects performed simulation trials involving various levels of automation. Several automation failures occurred and out-of-the-loop performance decrements were assessed. Results suggest that, in terms of performance, human operators benefit most from automation of the implementation portion of the task, but only under normal operating conditions; in contrast, removal of the operator from task implementation is detrimental to performance recovery if the automated system fails. Joint human/system option generation significantly degraded performance in comparison to human or automated option generation alone. Lower operator workload and higher situation awareness were observed under automation of the decision making portion of the task (i.e. selection of options), although human/system performance was only slightly improved. The implications of these findings for the design of automated systems are discussed.

1. Introduction

Various levels of automation (LOA) specifying the degree to which a task is automated are possible. The use of intermediate LOA may provide an approach to human-centred automation; automation that is designed and implemented to be compatible with human capabilities and capacities. Such approaches are currently being advocated in a wide variety of domains, in which human performance

*Author for correspondence.

problems have been noted when operating in conjunction with automated systems (Billings 1991). These problems have been associated with operator vigilance and complacency leading to loss of situation awareness (SA) and manual skill decay (Endsley and Kiris 1995). It has been hypothesized that by keeping the human involved in system operations, some intermediate LOA may provide better human/system performance and SA than that found with highly automated systems (Endsley 1987, Endsley and Kiris 1995).

Traditionally, automation design decisions have focused on optimizing the capabilities of the technology (technology-centred automation). Driven by a desire to reduce costs (through the reduction of human workload and thus human staffing requirements), such efforts usually assign a computer or mechanical controller to perform those tasks technically possible, and remove human operators from the control loop by placing them in the job of system monitor (Endsley 1995a). Unfortunately, monitoring is a role for which humans are generally ill-suited. This arrangement of functions between the human operator and the system has been found to be associated with operator performance problems in properly overseeing the automated system and assuming control when necessary (Wiener and Curry 1980, Moray 1986, Billings 1991, Wickens 1992). This out-of-the-loop performance problem has been attributed to numerous factors including vigilance decrements and complacency (Wiener 1988, Parasuraman *et al.* 1993), loss of operator SA (Endsley 1987, Carmody and Gluckman 1993, Endsley and Kiris 1995), poor feedback under automated conditions (Norman 1989) and manual skill decay (Wiener and Curry 1980, Shiff 1983).

Several LOA taxonomies have been proposed in the literature. Sheridan and Verplanck (1978) developed a LOA taxonomy, which incorporates ten levels comprising:

(1) human does the whole job up to the point of turning it over to the computer to implement;
(2) computer helps by determining the options;
(3) computer helps to determine options and suggests one, which human need not follow;
(4) computer selects action and human may or may not do it;
(5) computer selects action and implements it if human approves;
(6) computer selects action, informs human in plenty of time to stop it;
(7) computer does whole job and necessarily tells human what it did;
(8) computer does whole job and tells human what it did only if human explicitly asks;
(9) computer does whole job and decides what the human should be told; and
(10) computer does the whole job if it decides it should be done, and if so, tells human, if it decides that the human should be told.

This taxonomy incorporates issues of feedback (what the human should be told by the system), as well as relative sharing of functions determining options, selecting options and implementing. While this taxonomy can be applied in more general terms, it is instantiated in terms of which agent (the human or the computer) gets or requests options, selects actions, requests or approves selection of actions, starts actions, approves start of actions, or reports actions and has been framed in terms of the teleoperation environment.

Endsley (1987) developed a LOA hierarchy in the context of the use of expert systems to supplement human decision making. This hierarchy stipulated that a task could be performed using:

(1) manual control — with no assistance from the system;
(2) decision support — by the operator with input in the form of recommendations provided by the system;
(3) consensual artificial intelligence (AI) — by the system with the consent of the operator required to carry out actions;
(4) monitored AI — by the system to be automatically implemented unless vetoed by the operator; and
(5) full automation with no operator interaction.

The list is most applicable to cognitive tasks in which operator ability to respond to, and make decisions based on, system information (with expert system assistance) is critical to overall performance. Ntuen and Park (1988) have developed a similar 5-level LOA taxonomy within the context of a teleoperation system.

A 10-level taxonomy of LOA was developed here that is intended to have applicability to a wide array of cognitive and psychomotor tasks requiring real-time control within numerous domains including air traffic control, aircraft piloting, advanced manufacturing and teleoperations. All of these domains have many features in common, including: (1) multiple competing goals; (2) multiple tasks competing for an operator's attention, each with different relevance to system goals; and (3) high task demands under limited time resources. Four generic functions intrinsic to these domains were identified: (1) *monitoring* — scanning displays to perceive system status; (2) *generating* — formulating options or strategies for achieving goals; (3) *selecting* — deciding on a particular option or strategy; and (4) *implementing* — carrying out the chosen option. Ten levels of automation were then formulated by assigning these functions to the human or computer or a combination of the two, as shown in the taxonomy depicted in table 1.

(1) *Manual Control* (MC) — the human performs all tasks including monitoring the state of the system, generating performance options, selecting the option to perform (decision making) and physically implementing it.
(2) *Action Support* (AS) — at this level, the system assists the operator with performance of the selected action, although some human control actions are required. A teleoperation system involving manipulator slaving based on human master input is a common example.
(3) *Batch Processing* (BP) — although the human generates and selects the options to be performed, they then are turned over to the system to be carried out automatically. The automation is, therefore, primarily in terms of physical implementation of tasks. Many systems that operate at this fairly low level of automation exist, such as batch processing systems in manufacturing operations or cruise control on a car.
(4) *Shared Control* (SHC) — both the human and the computer generate possible decision options. The human still retains full control over the selection of which option to implement; however, carrying out the actions is shared between the human and the system.

(5) *Decision Support* (DS) — the computer generates a list of decision options that the human can select from or the operator may generate his or her own options. Once the human has selected an option, it is turned over to the computer to implement. This level is representative of many expert systems or decision support systems that provide option guidance, which the human operator may use or ignore in performing a task. This level is indicative of a decision support system that is capable of also carrying out tasks, while the previous level (shared control) is indicative of one that is not.
(6) *Blended Decision Making* (BDM) — at this level, the computer generates a list of decision options that it selects from and carries out if the human consents. The human may approve of the computer's selected option or select one from among those generated by the computer or the operator. The computer will then carry out the selected action. This level represents a higher level decision support system that is capable of selecting among alternatives as well as implementing the second option.
(7) *Rigid System* (RS) — this level is representative of a system that presents only a limited set of actions to the operator. The operator's role is to select from among this set. He or she may not generate any other options. This system is, therefore, fairly rigid in allowing the operator little discretion over options. It will fully implement the selected actions, however.
(8) *Automated Decision Making* (ADM) — at this level, the system selects the best option to implement and carry out that action, based upon a list of alternatives it generates (augmented by alternatives suggested by the human operator). This system, therefore, automates decision making in addition to the generation of options (as with decision support systems).
(9) *Supervisory Control* (SC) — at this level the system generates options, selects the option to implement and carries out that action. The human mainly monitors the system and intervenes if necessary. Intervention places the human in the role of making a different option selection (from those generated by the computer or one generated by the operator), thus, effectively shifting to the decision support LOA. This level is representative of a typical supervisory control system in which human monitoring and intervention, when needed, is expected in conjunction with a highly automated system.
(10) *Full Automation* (FA) — at this level, the system carries out all actions. The human is completely out of the control loop and cannot intervene. This level is representative of a fully automated system where human processing is not deemed to be necessary.

This LOA taxonomy provides several advantages in that it considers a wide range of options describing the way in which core functions can be divided between a human and a computer to achieve task performance. The functions it is based upon are generic enough to be applicable to a wide variety of domains and task types. The levels listed in the taxonomy represent a means of systematically examining the effect of automation, as implemented incrementally, on different aspects of a central task. It should be stated that the taxonomy shown represents a range of feasible assignments of the four functions of system(s) monitoring, and options generation, selection and implementation to human, computer and human/computer combinations. While it may be possible to conceive of certain combinations that are not

Table 1. Hierarchy of levels of automation applicable to dynamic-cognitive and psychomotor control task performance.

Level of automation	Roles: Monitoring	Generating	Selecting	Implementing
(1) Manual control (MC)	Human	Human	Human	Human
(2) Action support (AS)	Human/Computer	Human	Human	Human/Computer
(3) Batch processing (BP)	Human/Computer	Human	Human	Computer
(4) Shared control (SHC)	Human/Computer	Human/Computer	Human	Human/Computer
(5) Decision support (DS)	Human/Computer	Human/Computer	Human	Computer
(6) Blended decision making (BDM)	Human/Computer	Human/Computer	Human/Computer	Computer
(7) Rigid system (RS)	Human/Computer	Computer	Human	Computer
(8) Automated decision making (ADM)	Human/Computer	Human/Computer	Computer	Computer
(9) Supervisory control (SC)	Human/Computer	Computer	Computer	Computer
(10) Full automation (FA)	Computer	Computer	Computer	Computer

specifically listed here, these were not deemed to be either technically or practically feasible (e.g. it is difficult for either the human or machine to perform any task without directly monitoring either the state of the system or inputs from the other); however, other combinations cannot be completely ruled out. It should also be stated that although the levels in the taxonomy presented here have been preliminarily arranged in a linear fashion, this order is not necessarily ordinal and needs to be experimentaly examined on this basis.

Very little experimental work has been conducted to examine the benefit of applying intermediate LOA in complex tasks. Endsley and Kiris (1995) investigated people's ability to recover from automation failures when performing an automobile navigation task with the assistance of an expert system following Endsley's (1987) taxonomy. They found that out-of-the-loop performance decrements were decreased and SA increased when subjects used intermediate LOA as compared to full automation in performing this task. No experimental work on the other LOA taxonomies was found.

The purpose of the present research was to examine the impact of a wider range of LOAs within a complex, dynamic control task. An empirical investigation was therefore conducted to assess the effect of the LOA taxonomy in table 1 on human/system performance, operator SA and workload in a simulation of a dynamic control task. In particular, the impact of these LOAs on an operator's ability to assume manual control following automation failure was examined. The results of this research can be used to provide an indication of the relevant task aspects that may influence the success or failure of automation efforts that seek to create an effective human/machine system and provide a smooth transition during automation failures.

2. Method

2.1. *Task*

This study was conducted using a complex, computer-based dynamic control task called Multitask. The task was developed for this experiment based on an earlier simulation used by Tulga and Sheridan (1980). The task was designed to incorporate the common features found in many dynamic control jobs such as piloting, power systems operation, and air traffic control including: (1) object collision avoidance; (2) location and selection of objects; and (3) processing of tasks. These features require Multitask operators to develop complex strategies for optimizing system performance along multiple, simultaneous goals.

The Multitask simulation presents multiple tasks (targets) to an operator in the form of geometric abstractions. Specifically, the targets are represented by boxes travelling at different speeds towards a circular deadline at the centre of the display shown in figure 1. The operators' goal was to select and eliminate the targets (i.e. carry out the tasks) by collapsing their areas before they reached the deadline or collided with one another. The specific methods by which selection and elimination were accomplished through the interface are described below. For example, in manually controlling the simulation, operators pointed to a target with a cursor linked to a mouse controller and clicked on the target repeatedly in order to collapse it.

2.1.1. *Multitask point system and target characteristics*: Differing amounts of reward points were provided for collapsing each target (completing a task) and various penalty points were assessed for their expiration or collision with another target. A task expired if the target reached the deadline before being eliminated (missing a task). The basis on which reward and penalty points were assessed is as follows:

(1) The size (small, medium or large) and colour (red, blue or green) of each target represented various tradeoffs in rewards and penalties associated with the target. Small, medium and large targets occupied an area of 1.08, 2.15 and 4.3 cm^2 (0.342, 0.683 and 1.36° at a viewing distance of 71.12 cm).
(2) The exact rewards and penalties associated with each target were randomly generated as integers within the ranges shown in table 2 (depending on the size and colour of the target), and were displayed as data tags attached to the targets. These reward and penalty ranges were developed to simulate real-world tasks, which have varying levels of importance (both in terms of pay-off for completion and penalties for failure to complete). The most extreme reward and penalty ranges were subjectively associated with red and green targets, respectively. Red targets were the most important tasks, as reflected by higher levels of rewards and penalties (across all sizes). Green targets were correspondingly the least important tasks, as reflected by lower levels of reward and penalty points (across all sizes). In addition, greater reward ranges were associated with small targets because of an increased level of difficulty in acquiring these targets as compared to large ones. Large targets were assigned the most extreme penalty ranges because they were considered to be obvious threats to performance in terms of collisions.
(3) For each size/colour category, a 10-point range was used (with point values selected across a 100-point scale). Within these ranges, points were randomly assigned to each target to focus subject attention on the simulation and to promote dynamic decision making in selecting which targets to process first in order to maximize rewards and minimize penalties. By varying the point assignments, subjects were required to dynamically update their task processing strategies, which might otherwise have been fixed had target size and colour encoded specific point values.

The speed at which a target travelled and its distance from the centre deadline provided information on the time available for processing each task (i.e. collapsing the target). Targets travelled at one of 504 different rates (18 starting distances $\times$28 total travel times) randomly computed within the range of approximately 0.13 to 0.36 cm/s (0.102 to 0.286°/s). The distance to the deadline was covered within 28 to 57 s. Targets were allowed to pursue their trajectories until they collided with another target or reached the deadline. In all runs of the Multitask simulation, the maximum distance of any target to the deadline was approximately 10.16 cm and the minimum distance was 7.62 cm.

The size of a target indicated the time required to process it. Under automation of the Multitask, targets were collapsed at a constant rate of approximately 0.65 cm^2/s (0.204°/s) yielding processing times for small, medium and large targets of approximately 1, 3 and 6 s, respectively. (When processing targets manually with the mouse, 3, 4 and 5 button clicks were required to collapse small, medium and large targets, respectively.) (The time between a control action and collapsing of a target as well as the time-to-transfer control from one target to another was negligible under all levels of automation presented by Multitask.)

The minimum and maximum performance attainable in the simulation was computed. Given the target processing rates and the presence of five targets on the display at any given time, the minimum number of collapses in a 60-s period was 10. This would occur when all targets were large. The maximum number of collapses in a 60-s period was 60, which would occur if all of the targets were small. (This is a hypothetical range of target collapses. In reality, the size of targets was randomly determined throughout the simulation.)

Targets pursued one of eight approach paths from the edge of the display towards its centre causing convergent-type movement of targets, as shown in figure 1. Targets could collide on the same approach path if one was travelling faster than another, or they could collide on adjacent approach paths if they touched each other

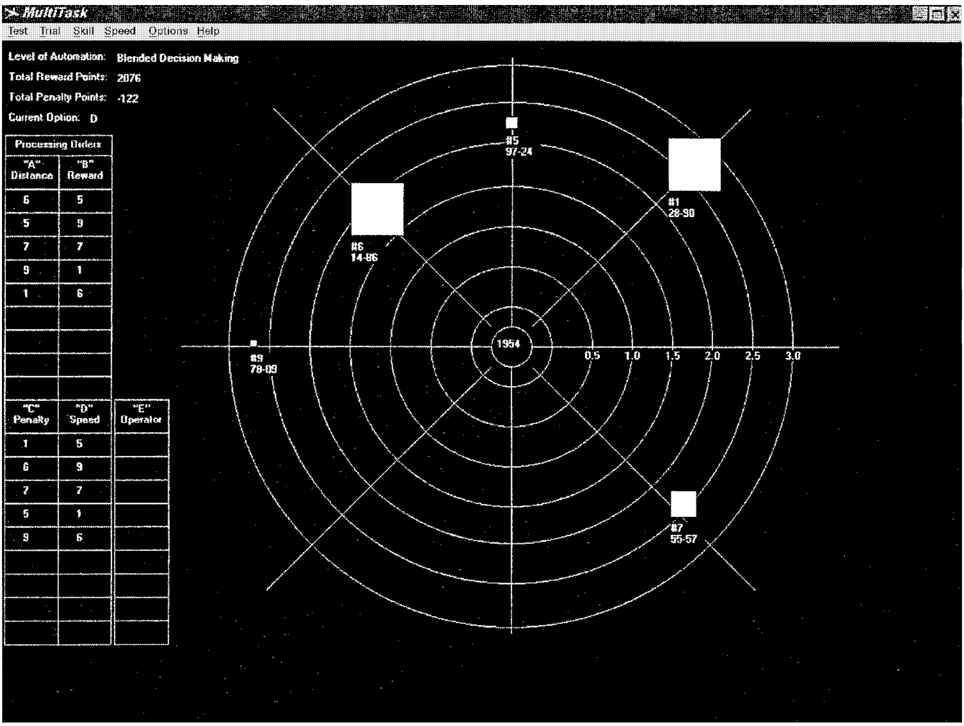

Figure 1. Multitask display.

Table 2. Reward and penalty point system used in Multitask simulation.

	Target colour					
	Red		Blue		Green	
Target size	Reward	Penalty	Reward	Penalty	Reward	Penalty
Small	90 – 100	20 – 30	80 – 90	10 – 20	70 – 80	0 – 10
Medium	60 – 70	60 – 70	50 – 60	50 – 60	40 – 50	40 – 50
Large	20 – 30	90 – 100	10 – 20	80 – 90	0 – 10	70 – 80

as they neared the centre of the display. This feature provided a certain level of complexity in the task owing to the interactive nature of different target characteristics (i.e. speed and size).

In order to optimize performance, subjects needed to develop a control strategy that took into account the various rewards associated with attending to each target, penalties associated with missing a target or allowing two targets to collide, assessments of the time available and required for processing each target, tradeoffs in attending to one target until completion versus switching to alternate targets (points for partial collapse of a target were not awarded), and tradeoffs in optimizing between different goals — maximization of reward versus minimization of penalty.

2.1.2. *Performance of multitask at levels of automation*: Multitask allowed for targets to be collapsed by different methods conforming to each of the LOAs in table 1. In all 10 conditions, the Multitask display, current LOA, total reward points, and total penalty points accumulated were displayed. In all levels except manual control, the status of the system was jointly monitored by the human and the computer for current target information and status, and for target selection information. The operator interface and the responsibilities of the human and computer at the various LOAs are described below.

(1) *Manual control* — the operator was required to: (a) continually monitor the display to take in information on the status of the competing targets and their relevant attributes; (b) generate a strategy (processing order) for eliminating targets; (c) select targets for elimination accordingly by pointing to them with the mouse; and (d) implement their strategy (process the tasks) by continually depressing the mouse button over the selected target until it disappeared.
(2) *Action support* — the operator was required to: (a) generate a target processing order; (b) select targets to be collapsed using the mouse; and (c) implement their strategy (process the targets) by clicking the mouse button once over the desired target to activate automated processing. This LOA, therefore, provided joint human/machine control in carrying out the task implementation.
(3) *Batch processing* — a target processing order (which was input by the operator) was also shown in the upper left corner of the display. The operator was required to: (a) generate a strategy for processing targets; and (b) select targets to be added to the processing order list by depressing the numeric keys 1 to 9 on a keyboard corresponding to numbers tagged to the displayed targets. (In all conditions requiring subjects to use numeric keys on the keyboard to select targets, only one keystroke was required to initiate automated computer processing of a target.) The computer implemented the operator's processing order by automatically collapsing each target in the processing list. This LOA, therefore, provided full automation of the implementation portion of the task.
(4) *Shared control* — four processing orders were generated by the computer (based on target distance, reward, penalty and speed), and were displayed to assist the operator in target selection. Additional guidance was offered by the computer in the form of a magenta dot tagged to the target that was currently the 'best' choice in terms of all variables based on an optimization

algorithm that considered distance, reward, penalty and speed. The operator was required to: (a) generate a processing strategy, which could be his or her own or could be based on the computer guidance; (b) select targets to be collapsed using the mouse; and (c) implement this strategy (process the targets) by clicking the mouse button once over the desired target to activate automated processing. This LOA provided joint human/computer generation of decision options (strategies) and joint implementation of the human selected decision.

(5) *Decision support* — the same information as at the shared control level was presented along with a column that allowed for a processing order to be entered by the operator (as under batch processing). The operator and computer both generated strategies for eliminating targets, as above. The operator was required to select one of the computer's four processing orders by depressing the keys A to D on the keyboard, or to enter his or her own processing order by depressing the keys 1 to 9 corresponding to the numbers tagged to the individual targets. (The selected strategy (processing order) could be changed at any time; in all conditions requiring subjects to use alpha character keys on the keyboard to select a computer-generated processing strategy only one keystroke was required to initiate automated processing of the strategy.) The computer implemented the selected processing order by automatically collapsing targets on the selected list one at a time. Decision support, therefore, provided a LOA analogous to that provided with many expert systems whereby the computer provides recommended actions (which the human can use or ignore) and performs as directed by the human operator.

(6) *Blended decision making* — the same display as in the decision support level was presented. The operator and computer both generated strategies for eliminating targets (as above); however, the computer selected the processing order to be implemented. The order selected by the automation could be approved by the operator or overridden (if he/she did not agree with the computer's choice) by depressing the keys A to E corresponding to the desired order. The computer implemented the selected processing order by automatically collapsing targets on the selected list one at a time. The blended decision making LOA provides a higher level of automation by incorporating computer selection with human veto power.

(7) *Rigid system* — the same information as in the shared control level was presented. At this LOA, however, the operator was not allowed to generate his or her own strategy; only computer formulated processing orders were available. The operator was required to select from the computer's orders by depressing the keys A to D corresponding to the reward, penalty, distance and speed processing orders. The computer implemented the selected processing order by automatically collapsing the targets on the list one at a time. A rigid system LOA represents one in which the human is limited in options and discretion by the system.

(8) *Automated decision making* — the same information as in the blended decision making level was presented. The operator and the computer both generated options, as in the blended decision making level. At this LOA the computer selected (in real-time) a processing order from those that it had formulated or the one entered by the operator. The computer implemented

the selected order by automatically collapsing the targets. Automated decision making, therefore, provides automated selection and implementation of options that are either computer- or operator-generated.

(9) *Supervisory control*—this mode offered automation of all functions with human override capability. Therefore, the computer: (a) generated a processing strategy by taking into account all target variables; (b) selected targets for elimination; and (c) implemented the strategy by automatically collapsing targets one at a time. The operator could intervene in the control process, if he/she thought that the computer was not efficiently eliminating targets. Operator intervention was accomplished by depressing a key that temporarily shifted (for 1 min) the LOA to decision support. The operator could return to automation of all functions before the end of the temporary shift in LOA by depressing a second key. This LOA is, therefore, representative of many supervisory control systems in which the system is mostly automated, but human monitoring and intervention is expected.

(10) *Full automation*—in this mode all functions comprising: (a) processing order generation; (b) target selection; (c) strategy implementation (target elimination); and (d) system monitoring were performed by the computer. Operator intervention was not permitted. Therefore, under full automation the operator could only observe system performance.

It should be noted that the performance capability of the human-machine system across the various LOAs of the Multitask simulation was not perfect in that the computer algorithm used in the simulation was unable to recognize potential target collisions (task conflicts). This algorithm was programmed to encourage operator task involvement through intervention in the control loop when necessary. (It was also representative of automation in many systems that rarely functions perfectly because it only considers particular aspects of a task in its algorithm.) Full automation of the Multitask simulation permits, on average, less than 1 target collision per minute (mean = 0.85). This is greater than the number of collisions potentially realizable at all other levels of automation as human interaction tends to overcome the algorithm problem (Kaber 1996). The minimum number of target collisions (task conflicts) attainable in the simulation with human intervention is zero.

The effect of collisions on overall task performance was minimal, however, in comparison to the effect of target selection strategy. Therefore, Multitask allows the benefits of human-automation interaction across LOAs to be evaluated under circumstances that are representative of real-world automation.

2.2. *Experimental design*

The independent variable for the experiment was the LOA provided for performing the task, i.e. (1) manual control; (2) action support; (3) batch processing; (4) shared control; (5) decision support; (6) blended decision making; (7) rigid system; (8) automated decision making; (9) supervisory control; and (10) full automation.

The dependent variables for the experiment were:

(1) Task performance during normal operation quantified in terms of the number of target collapses (also represented as total reward points), expirations and collisions (also represented as total penalty points).

(2) Performance during automation failure measured in terms of time-to-recovery (the time elapsed between an automation failure and a subsequent appropriate operator control action), and manual task performance during failure (as measured by the performance variables above) — with respect to the former measure of time-to-system recovery, return of the Multitask simulation to normal functioning required operators to assume manual control of the simulation and use the mouse controller to point and click on a target on the task display.

(3) Operator SA captured using Situation Awareness Global Assessment Technique (SAGAT) queries (Endsley 1988) — SAGAT was used in this study as an objective means by which to quantify SA. Subjective measures of SA were not used because of limitations in the veracity of self-ratings and observer ratings of SA. Self-ratings of SA may not be truly representative of actual SA because operators are limited to their own perceptions of the task environment and may not have an accurate picture of reality by which to judge the completeness or correctness of their perceptions (Endsley 1995b). Observer ratings of SA are limited by the fact that trained observers often have information about the simulation and reality, but may have only limited knowledge of an operator's concept of a situation (Endsley 1995b).

(4) Operator-perceived workload quantified in terms of NASA-TLX scores (Hart and Staveland 1988) — this subjective measure of workload was selected because of its demonstrated reliability and sensitivity as an overall workload measure in empirical investigation (Vidulich 1989). It was chosen over other potential objective secondary task and physiological measures of workload because of concerns regarding the obtrusiveness (primary task interference) and sensitivity of these measures. In addition, workload effects must be indirectly inferred from differences revealed through physiological measures (Wickens 1992: 391 – 397).

Thirty subjects participated in the study, each performing the task under two levels of automation. Therefore, each LOA was experienced by a total of six subjects. That is, the full range of LOAs was covered across the subjects.

2.3. *Apparatus*

The study was conducted by running a simulated task on a Gateway 66-MHz Pentium® based personal computer (PC) linked to a high resolution, 15-in graphics monitor, a standard keyboard, and mouse. The monitor operated at 60 Hz under 1024×768 resolution with a refresh rate of 30 frames/s.

2.4. *Subjects*

Thirty undergraduate students comprising nine females and 21 males having normal or corrected to 20/20 visual acuity participated for monetary compensation on a voluntary basis. Subjects ranged in age from 20 to 43 years. The majority of participants (26) were right-hand dominant.

2.5. *Procedure*

All subjects participated in a 5-min familiarization period at the onset of the experimental session. The period was designed to acquaint subjects with the procedures and equipment used in the experiment including: (1) the SAGAT procedure and specific queries that would be administered to assess SA in this task; (2) the NASA-TLX workload battery; and (3) the mouse and graphics display used for the Multitask simulation. Ranking of NASA-TLX components was also completed. The familiarization period was followed by a training session during which each subject was required to perform the Multitask using the mouse at a manual LOA with no computer assistance for 20 min. The training session was followed by a 5-min rest period.

All subjects were required to complete two trials, each under a different LOA. Prior to each trial, subjects received a 10-min practice period in which they performed the task at the LOA to be tested. This training period was followed by a 20-min trial for data collection. Both trials were separated by a 5-min break.

During data collection trials, three automation failures were simulated within the first 10 min (except in those trials involving manual control of the simulation during normal operating conditions). A failure constituted a shift in the LOA to manual control. In trials implementing manual control under normal conditions, there was no 'automation' to fail and manual control was maintained throughout testing. All automation failures were made salient to subjects by an audio-tone of 530 Hz, approximating the musical note of high C (Goldstein 1989: 388, Figure 11.5) for 1.5 s and display of the message 'Failure' in the LOA data field on the Multitask display. The automation failures occurred at random intervals ranging from 2 to 3 min. Subjects were required to assume control manually following each failure for 1 min before automation was restored to the LOA under study in that condition.

Subjects were informed in advance that automation failures might occur during the test periods. This was necessary to ensure that they had knowledge of how to appropriately interact with the system in the event of a failure. (The procedure was in-line with the experimental objective of assessing the effect of the LOA on operator failure mode performance versus assessing, for example, operator failure diagnosis ability.) Without some training and instruction on what to do under conditions of a failure, subjects might have been unable to continue the experiment. This is also reflective of real-world scenarios where operators of automated systems are instructed as to their responsibilities when intervening in a control loop during an automation failure. No information concerning the number, or timing, of failures that a subject might expect within a trial was provided in order to prevent advanced preparation. The subjects were informed that a failure would constitute a shift in the LOA to manual control, the level of functioning they had experienced during the training session.

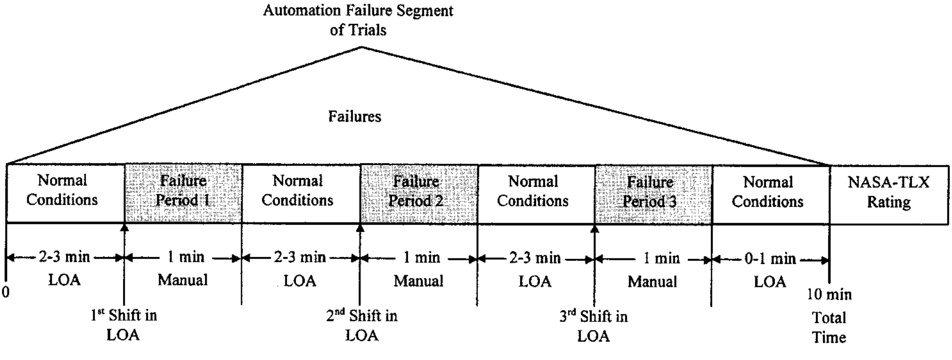

Figure 2. Schedule for automation failure segment of trials.

Figure 2 shows a typical schedule for the first 10 min of the test trials during which automation failures occurred. The first failure occurred after 2 to 3 min of performance time had elapsed.

Three simulation freezes were dispersed throughout the remaining 10 min of the trial to adminster SAGAT queries. The SAGAT data collection freezes were intentionally distributed in a separate time period from the automation failures to avoid possible confusion for the subject and confounding of these measures. (This procedure was in-line with the experimental objective of assessing the effect of the different LOAs on SA, and not assessing, for example, system failure effects on SA. Subjects were informed of this procedure in advance of testing.)

Simulation freezes occurred at random intervals ranging from 2 to 3 min. When a freeze happened, the Multitask display was blanked and subjects responded to an electronic form of the SAGAT queries also presented on the Pentium® PC. Each freeze lasted until subjects completed the queries which asked for: (1) target colour and size identifications for each target (level 1 SA); (2) four questions concerning the reward, penalty, distance and speed of targets (level 2 SA); and (3) a single question concerning when a target would reach the deadline at the centre of the display (level 3 SA). The latter question was posed on a relative basis. Therefore, the question required identification of which target, out of all the targets on the display at the time of the freeze, would be next to reach the deadline.

After a freeze, the task was resumed until trial completion. No advanced knowledge of the number of freezes or inter-freeeze interval times was given to subjects at the onset of the experiment to prevent advanced preparation. Although during testing under the second LOA subjects might have guessed at the number of freezes and inter-freeze intervals that would occur, subjects could not be certain of this as they did not know that the number of freezes would be the same in both trials. In addition, the timing of each freeze was randomly determined within the range given, making prediction of the freeze times difficult.

Figure 3 shows a typical schedule for the latter 10 min of the test trials during which simulation freezes occurred to administer the SAGAT queries. The first freeze occurred between 2 and 3 min after the close of the automation failure segment of the trials.

At the end of each 10-min segment of the trial, participants rated workload on the NASA-TLX dimensions comprising: (1) mental demand; (2) physical demand;

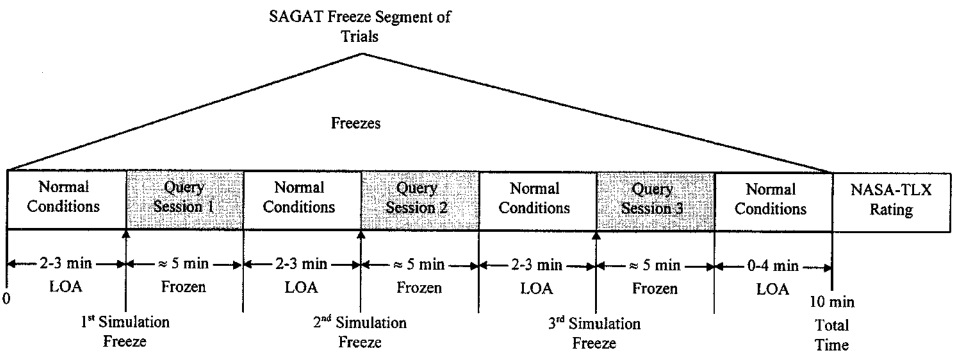

Figure 3. Schedule for SAGAT freeze segment of trials.

(3) temporal demand; (4) performance; (5) frustration; and (6) effort. All trials were performed in an environmental chamber adjusted to normal room conditions (26°C) to block-out extraneous distractions. In total, each experimental session required approximately 2 h per subject.

3. Results and discussion

Performance measures observed during the experiment comprised the number of targets collapsed, expired, and collided in each trial. These variables also determined the total target reward and penalty points assessed. The measures were averaged at 1-min intervals across the 20-min testing period during all 60 trials (10 LOAs $\times$ 6 subjects within LOA) yielding 1200 observations per response. The performance data were divided into two subsets for analysis: performance under normal operating conditions and performance during automation failure. The former set contained all observations of normal functioning at the various LOAs and the latter contained points collected during shifts to manual control following an automation failure. This allowed for an examination of the effects of the LOAs on normal human/system operating performance and of human performance during automation failures. No attempt was made to compare automation, in general, to human performance as this would be dependent upon system implementation and beyond the goals of the study. For the purposes of comparison, three 1-min intervals of manual control (level 1) performance were ascribed to the second data set. These periods would have been coded by the computer as failure modes, had such a condition been possible in trials involving manual control (level 1) under normal operating circumstances.

The time-to-recover from an automation failure was recorded for each of the three simulated automation failures in all 60 trials producing 180 data points. Situation awareness was recorded as the percentage of correct responses for each of the SAGAT queries for a total of 180 responses. Two workload observations were also obtained for each of the 60 trials.

All data sets were analysed through one-way analyses of variance (ANOVAs) with LOA as a between-subject variable. An arcsine transform was applied to the percentage of correct responses to the SAGAT data set to ensure that all assumptions of the ANOVA were upheld.

The findings of ANOVAs on all response measures collected during subject functioning at each LOA under normal operating conditions and during the manual control (level 1) periods (during automation failures) are summarized in table 3. The table reveals significant changes in the majority of the responses, including the number of target collapses, collisions and expirations, operator levels 2 and 3 SA, and NASA-TLX overall workload scores, attributable to the LOA manipulations.

3.1. *Performance under normal operating conditions*

Analysis of variance results on the number of target collapses, $F(9,20) = 23.75$, $p = 0.0001$; expirations, $F(9,20) = 17.81$, $p = 0.0001$; and collisions, $F(9,20) = 16.52$, $p = 0.0001$, revealed a main effect of LOA. The one-way model on total reward points, $F(9,20) = 12.28$, $p = 0.0001$, and penalty points, $F(9,20) = 23.56$, $p = 0.0001$, revealed the same finding. Tukey's honestly significant difference (HSD) test was used to further investigate the effect of LOA. The results of the multiple comparison procedure on the significant LOA effect on target collapses, expirations and collisions are shown in table 4. (Only the significant differences revealed by the *post hoc* analysis are discussed in this section.)

Table 3. Summary of ANOVA results on automation performance under normal operating conditions, manual control performance during automation failure, operator SA and workload data.

	Normal conditions		Automation failure	
Response measures	*F*-statistic	*p*-value	*F*-statistic	*p*-value
Target collapses	$F(9,20)= 23.75$	$p= 0.0001$**	$F(9,20)= 5.10$	$p= 0.0012$**
Target expirations	$F(9,20)= 17.81$	$p= 0.0001$**	$F(9,20)= 3.13$	$p= 0.0162$*
Target collisions	$F(9,20)= 16.52$	$p= 0.0001$**	$F(9,20)= 0.62$	$p= 0.7693$
Total target reward points	$F(9,20)= 12.28$	$p= 0.0001$**	$F(9,20)= 0.34$	$p= 0.9490$
Total target penalty points	$F(9,20)= 23.56$	$p= 0.0001$**	$F(9,20)= 9.9$	$p= 0.0001$**
Time-to-recover			$F(9,20)= 8.18$	$p= 0.0001$**
Level 1 SA	$F(9,20)= 0.81$	$p= 0.6161$		
Level 2 SA	$F(9,20)= 4.3$	$p= 0.0032$**		
Level 3 SA	$F(9,20)= 2.58$	$p= 0.0369$*		
NASA-TLX	$F(9,20)= 4.48$		$p= 0.0025$**	

Independent variables—level of automation.
* Significant at the $\alpha= 0.05$ level.
** Significant at the $\alpha= 0.01$ level.

The mean number of target collapses, expirations, and collisions of targets across subjects, as a function of LOA, are shown in figure 4. The number of target collapses (tasks processed) peaked under batch processing (level 3) and action support (level 2). The next highest number of collapses occurred under supervisory control (level 9), full automation (level 10), and rigid system (level 7). The fewest collapses occurred at manual control (level 1).

Performance measured in terms of the number of target expirations (missed tasks) decreased as a function of LOA with the highest number of expirations occurring under manual control (level 1). A slight increase in missed tasks occurred at the higher LOAs (levels 9 and 10). The number of target collisions was slightly higher for manual control (level 1), action support (level 2) and batch processing (level 3) (mean collissions equalled 2.9, 2.6 and 2.9, respectively). All three of these levels yielded more collisions than the other higher LOAs including shared control (level 4), decision support (level 5), blended decision making (level 6), rigid system (level 7), automated decision making (level 8), supervisory control (level 9) and full automation (level 10) (mean collisions equalled 1.5, 1.5, 1.5, 1.4, 1.6, 1.4 and 1.2, respectively). These two groups of LOAs were significantly different ($p< 0.05$) because the variance in the response measure attributable to subjects was very small compared to the variance attributable to the LOA (Mean Square $(\mathrm{MS})_{\mathrm{LOA}} = 35.88$ versus $\mathrm{MS}_{\mathrm{Subject}} = 2.17$) and, consequently, an extremely significant *F*-value was produced. The insignificant performance variance ($F(20,1019) = 1.22$, $p = 0.2287$) among subjects, measured in terms of the number of target collisions overlooked during testing, may be explained by subject experience in the Multitask attained through the training period.

The mean total reward points, as a function of LOA, revealed the same pattern of performance as the number of target collapses. The total penalty points produced the same pattern as the number of target expiration, with a peak occurring under manual control (level 1). Since the general trends of observations on the reward and

Table 4. Tukey's HSD test results on performance measure means across LOAs, as recorded under normal operating conditions.

	Level of automation									
Response measure	Manual control (1)	Action support (2)	Batch processing (3)	Shared control (4)	Decision support (5)	Blended decision making (6)	Rigid system (7)	Automated decision making (8)	Supervisory control (9)	Full automation (10)
Target collapses		A	A							
							B		B	B
				C	C	C		C		
	D									
Target expirations	A									
		B	B							
									C	C
						D			D	
				E	E	E	E	E		
Target collisions	A	A	A							
				B	B	B	B	B	B	B

Levels with the same letter are not significantly different ($\alpha = 0.05$).

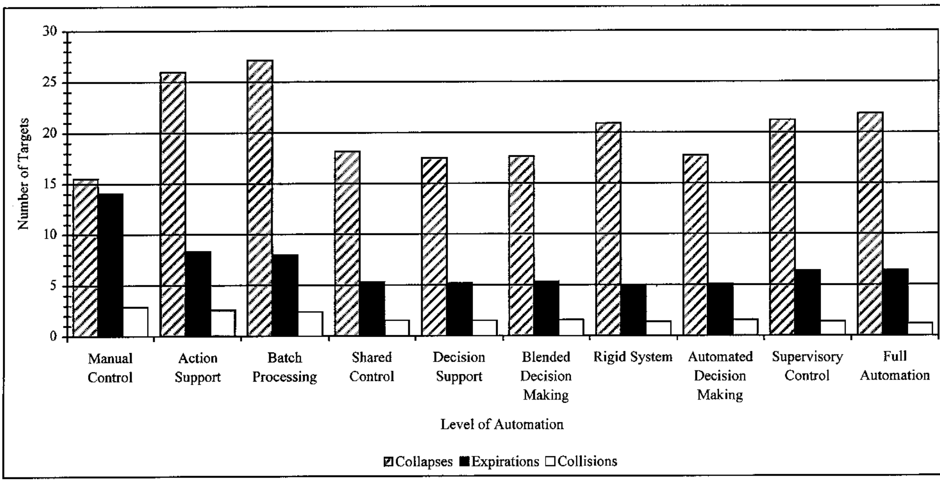

Figure 4. Mean number of target collapses, expirations and collisions under normal operating conditions for each LOA.

penalty point measures were in agreement with those developed for the target measures (collapses, expirations and collisions), and the ANOVA results on reward and penalty points corroborated those on target collapses, expirations and collisions, summary of the performance data was limited to the target response measures. These variables provided for a finer analysis of human-machine system functioning than the reward and penalty points. In particular, the ability of the human/computer to address the different negative events in the simulation (i.e. target expirations and collisions) could be examined in detail. Penalty points did not delineate these events.

With reference to the LOA taxonomy presented in table 1, these findings indicate that LOAs involving computer aiding or computer assumption of the implementation aspect of a task allowed for significant improvements in overall operator/system functioning. Specifically, mean target collapses under action support (level 2) and batch processing (level 3) were significantly greater ($p < 0.05$) than manual control (level 1), and better at higher LOA that also added automation to other task aspects. The number of target expirations also significantly decreased ($p < 0.05$) at these levels relative to manual control (level 1), but were lowest at shared control (level 4) and automated decision making (level 8) among other intermediate LOAs. Whether tasks had to be selected one at a time (action support (level 2)) or could be queued for processing (batch processing (level 3)), did not significantly affect performance ($p > 0.05$).

LOA requiring joint human-computer generation of options produced a decrease in performance compared to purely human generation in terms of collapses, but not expirations or collisions. This can be seen in comparing mean target collapses under action support (level 2) and batch processing (level 3) to that under shared control (level 4), decision support (level 5), blended decision making (level 6) and automated decision making (level 8). The mean number of missed tasks (target expirations) was significantly lower ($p < 0.05$) at the latter four levels involving joint human-computer generation of options. Although the way in which generation of options was conducted varied considerably between these four levels (in terms of the respective roles of the human and computer in the sharing of the generation option), this did not appear to have a substantial impact on performance. Comparing joint human-computer generation of options (levels 4, 5, 6 and 8) to purely computer generated options (rigid system (level 7), supervisory control (level 9) and full automation (level 10)) reveals slightly better performance at the latter levels (in terms of target collapses but not in terms of target expirations). Performance was still not as good, however, as with the purely human generation of options (action support (level 2) and batch processing (level 3)).

Based on statistical comparisons, LOA involving joint human-computer selection (blended decision making (level 6)) had no significant impact ($p < 0.05$) on performance, as compared to purely human selection (decision support (level 5)), when all other role allocations were held constant. Likewise, statistical comparisons revealed computer selection (automated decision making (level 8)) to be no different than joint human-computer selection (blended decision making (level 6)) or human selection (decision support (level 5)), where all other function allocations were equivalent. This finding is also present in comparing computer selection under supervisory control (level 9) to human selection under rigid system (level 7) in terms of processing and overlooking tasks, where the generation role is maintained by the computer in both cases.

The peak in performance at the rigid system (level 7), almost equal to the level of performance under supervisory control (level 9), may be attributed to the human not being required to perform the options generation role jointly with the computer at these two LOAs. Joint generation of options proved to be difficult across shared control (level 4), decision support (level 5), blended decision making (level 6) and automated decision making (level 8). The time required for operators to interact with the computer in the generation role might have detracted from their devotion of resources to the other system functions of monitoring and selecting (except at level 8), potentially inhibiting performance. Contrary to expectation, the assignment of human versus computer control, as part of LOAs, made a significant difference in the implementation and option generation functions, but not in the actual decision making (option selection) portion of the task.

3.2. *Manual performance during automation failure*

Results of the ANOVA indicated a significant main effect on target collapses, $F(9,20) = 5.1$, $p = 0.0012$, and expirations, $F(9,20) = 3.13$, $p = 0.0162$, but not for collisions, $F(9,20) = 0.62$, $p = 0.7693$. An ANOVA on reward points, $F(9,20) = 0.34$, $p = 0.9490$, was not supportive of the finding on collapses, however, total penalties, $F(9,20) = 9.9$, $p = 0.0001$, did corroborate the result regarding expirations. Further investigation of the LOA effect on manual performance during automation failures was conducted using Tukey's HSD test. Table 5 presents the results of the multiple comparison procedure for the number of target collapses and expirations recording during automation failures and serves as a basis for the comparisons made in the following discussion.

Figure 5 shows the mean target collapses, expirations and collisions attained manually during the period immediately following automation failures under each LOA. Human performance (in terms of the number of targets collapsed) in the period immediatley following an automation failure from all conditions except batch processing (level 3) and automated decision making (level 8) was not significantly

Table 5. Tukey's HSD test results on manual control performance and time-to-recovery means across LOAs, as recorded during automation failures.

	Level of automation									
Response measure	Manual control (1)	Action support (2)	Batch processing (3)	Shared control (4)	Decision support (5)	Blended decision making (6)	Rigid system (7)	Automated decision making (8)	Supervisory control (9)	Full automation (10)
Target collapses	A	A		A	A	A	A		A	A
			B		B	B	B	B	B	
Target expirations	A	A	A	A	A		A	A		
	B	B		B	B	B	B			B
	C	C		C		C			C	C
Time-to-recovery			A					A		
						B		B		
				C	C	C	C		C	C
	D			D	D		D		D	D
	E	E		E	E				E	E

Levels with the same letter are not significantly different ($\alpha = 0.05$).

different ($p > 0.05$) than human performance when the task had been performed manually all along (level 1). The number of target expirations (tasks missed) was also not significantly different ($p > 0.05$) from complete manual control (level 1) across LOAs during this time period, except for blended decision making (level 6), supervisory control (level 9) and full automation (level 10), which showed a decrease in the number of tasks missed.

Even though the number of target collapses was different across LOAs, a means plot of the reward points result across all LOAs revealed a near-straight line with a marginal decrease in total reward under blended decision making (level 6). The lack of significant differences in reward points obtained during manual functioning in the failure periods could possibly be attributed to differences in operator strategies as influenced by the preceding LOA. For example, although significantly more targets were addressed during failures in action support (level 2), as compared to batch processing (level 3), the target collapsing strategy employed during failures in action support (level 2) may have dictated selection of targets carrying comparatively lower reward points. Whereas, few, small targets carrying high rewards may have been selected during failures in batch processing (level 3). Differences in the strategies used during failure modes may have been affected by the different strategies used under the LOAs. Consequently, average reward points across these and other levels would not be reflective of the number of target collapses.

An ANOVA on time-to-recover from automation failure, $F(9,20) = 8.18$, $p = 0.0001$, revealed a significant effect of LOA. Tukey's HSD test was used to further investigate the LOA differences. The results of the multiple comparison procedure are also shown in table 5 and indicate that recovery time was significantly greater ($p < 0.05$) for automation failures that occurred during batch processing (level 3) and automated decision making (level 8), and was lowest for manual control (level 1) (providing an indication of time between control actions where no failure was present) and action support (level 2). The mean operator time-to-recover from automation failure at each LOA is shown in figure 6.

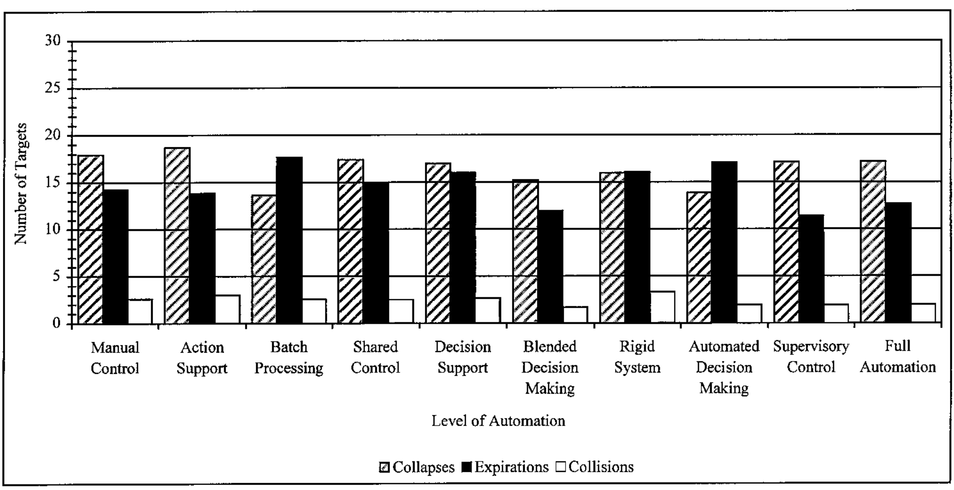

Figure 5. Mean number of target collapses, expirations and collisions during automation failures for each LOA.

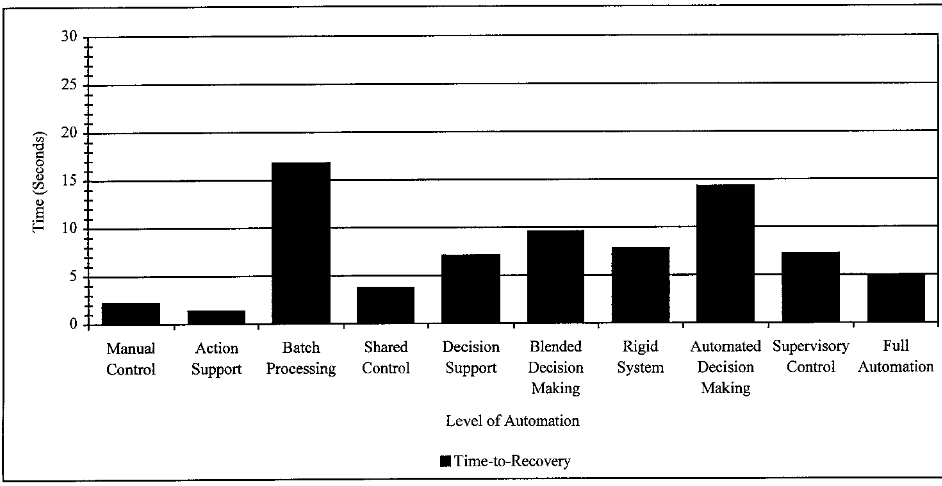

Figure 6. Mean time-to-recover from automation failures at each LOA.

These findings indicate that operator ability to recover from automation failures was substantially improved with lower LOA requiring some operator interaction (either the human alone or joint human-computer interaction) in the implementation role. In particular, action support (level 2) yielded significantly shorter ($p < 0.05$) failure recovery times than all other levels (equal to that under manual control). This level required subjects to select targets one at a time for elimination in order to implement the processing strategy, versus those that allowed them to queue targets for automated processing by the computer (e.g. batch processing (level 3)) or to simply choose a computer-generated processing order (e.g. decision support (level 5)). It is interesting that peaks in time-to-recover occurred at batch processing (level 3) and automated decision making (level 8) — those levels that also produced the worst manual performance during failure periods, as measured in terms of target collapses.

Following automation failures, time-to-recovery of task control was substantially worse when the preceding LOA allowed for advanced queuing of targets. A greater number of target expirations also occurred at these levels ($F(9,20) = 3.13$, $p = 0.0162$). However, whether advanced target queuing was performed by the human, or the computer, did not cause batch processing (level 3) and automated decision making (level 8) to significantly differ ($p < 0.05$) in terms of collapses and expirations, or in terms of their effect on operator recovery time from the automation breakdowns. When automation failures occurred during batch processing, subjects remained oblivious to the events for extended periods of time (while they performed other aspects of the task). Most likely they did not remain aware of which target was being processed and that the target was not disappearing from the display.

3.3. *Situation awareness data*

ANOVA results on operator SA indicated that LOA was not significant in influencing the percentage of correct responses to SAGAT queries involving operators' awareness of target features (level 1 SA questions). To assess level 1 SA, SAGAT queries were posed to operators concerning both the colours and sizes of targets on the Multitask display prior to a freeze and blanking of the PC screen. Graphical analysis (second-order regression trend) on the mean percentage of correct responses to level 1 SA queries on target colour revealed SA to improve marginally from low to upper-intermediate LOAs with manual control (level 1) producing the poorest average SAGAT score (mean = 48%) and automated decision making (level 8) yielding the greatest SA (mean = 71%). However, the impact of LOA on level 1 SA of target colours could not be separated from chance.

A similar analysis was conducted on the average percentage correct responses to level 1 SA queries regarding target sizes. A second-order regression trend on the observations revealed level 1 SA to increase slightly from low to intermediate LOAs and to marginally decrease from intermediate to high levels. Specifically, manual control (level 1) produced the worst average number of correct target size identifications (mean = 50%) with blended decision making (level 6) causing the greatest size awareness (mean = 73%). Operator SA under full automation (level 10) (mean = 58%) was less than that observed under the intermediate levels ranging from shared control (level 4) to automated decision making (level 8) and approached the percentage of correct target identifications for manual control (level 1). Again, however, the behaviour of these non-significant trends could be explained by chance.

Percentage correct for queries that required operators to comprehend the meaning of target characteristics in terms of their operational goals (maximizing task reward and minimizing penalty (level 2 SA questions)), was significantly different between LOA, $F(9,20) = 4.3$, $p = 0.0032$. Comparison of the mean percentages of correct responses to the level 2 SA queries was conducted using Duncan's Multiple Range test at the $\alpha = 0.05$ significance level to emphasize power in revealing truly significant differences. The results of the *post hoc* comparisons are shown in table 6.

Mean percentage of correct responses to level 2 SA questions, as a function of LOA, are shown in figure 7. Peaks occurred under blended decision making (level 6) and automated decision making (level 8), supervisory control (level 9), and full automation (level 10), while all other levels showed significantly lower ($p < 0.05$) level 2 SA. (Mean percentage correct for level 2 SA did not fall below chance, which was 11% for each question, in any of the conditions.)

These findings indicate that higher LOAs that do not require the human to perform selection allow for improved SA in terms of task/system understanding (level 2 SA). However, whether strategy selection was only aided or completely controlled by the computer did not result in blended decision making (level 6) significantly differing ($p > 0.05$) from automated decision making (level 8), supervisory control (level 9) or full automation (level 10) in terms of operator comprehension. Lower LOA corresponded to poorer levels of operator comprehension of target priorities and completion status. Lower operator SA under these LOAs may be attributed to the added burdens of the selection and monitoring roles along with responsibility for strategy generation in situations in which there was limited time available for perceiving target characteristics.

Only one SAGAT query required operators to project the state of the system in the future (level 3 SA). Although the percentage of correct responses was significantly influenced by LOA, $F(9,20) = 2.58$, $p = 0.0369$, the trend of the data could not be meaningfully interpreted in terms of the LOA taxonomy or the functions assigned to either the human and/or computer across LOA.

Table 6. Tukey's HSD test results on mean percentage of correct responses to level 2 and 3 SA queries across LOAs along with comparisons on NASA-TLX overall workload means across conditions.

Response measure	Level of automation: Manual control (1)	Action support (2)	Batch processing (3)	Shared control (4)	Decision support (5)	Blended decision making (6)	Rigid system (7)	Automated decision making (8)	Supervisory control (9)	Full automation (10)
Level 2 SA						A		A	A	A
	B	B	B	B	B		B			
Level 3 SA			A	A			A			A
			B	B						B
	C	C	C		C	C		C	C	C
NASA-TLX	A	A	A	A	A		A			
				B	B	B	B	B	B	B

Levels with the same letter are not significantly different ($\alpha = 0.05$).

Figure 8 shows the mean percentage of correct responses to the level 3 SA question, as a function of LOA. Operator ability to project future expirations appeared to significantly increase ($p < 0.05$) from low-level automation (manual control (level 1) and action support (level 2)) to low-intermediate levels (batch processing (level 3) and shared control (level 4)). It is possible that this was due to substantial reductions in subjective workload with increasing LOA, including operator removal from the implementation aspect of the task under batch processing (level 3), freeing-up operator cognitive resources for predictions. However, at mid- and upper-intermediate LOAs, including decision support (level 5), blended decision making (level 6), automated decision making (level 8) and supervisory control (level 9), level 3 SA was equivalent to the lower operator prediction capability observed at manual control (level 1) and action support (level 2). General consistencies did not exist in the way in which functions were allocated to the human and/or computer

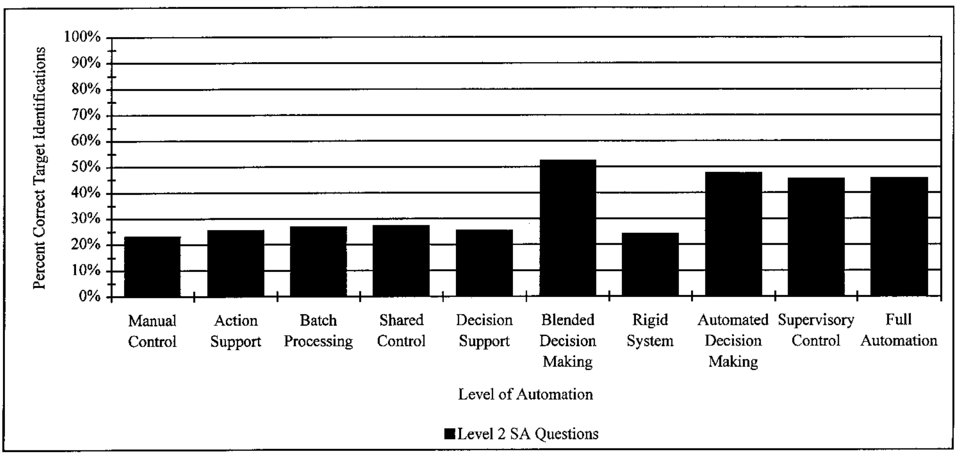

Figure 7. Level 2 SA mean percentage correct responses for each LOA.

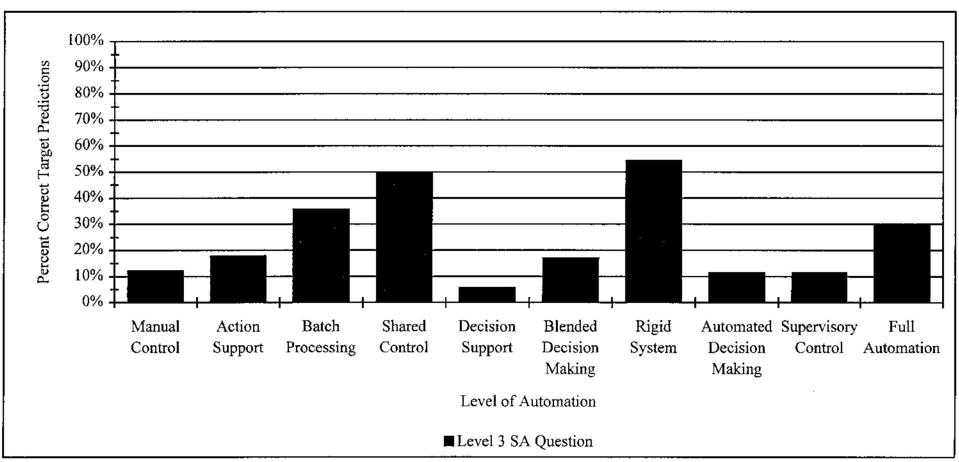

Figure 8. Level 3 SA mean percentage correct responsese for each LOA.

across these levels preventing meaningful inferences on the results. Peaks in level 3 SA were also observed for rigid system (level 7) and full automation (level 10) performance, but were not interpretable, as well.

The *post-hoc* analysis for the level 3 SA response is presented in table 6 (along with that for the level 2 SA measure) and supports the graphical analysis. In general, these results most likely reflect a low level of stability in this measure, as only one query involved level 3 SA or the significance of the effect could be an artefact.

As an aside, to assess whether the simulation freezes for administering SAGAT questionnaires were intrusive in system functioning, an ANOVA was conducted on performance, as measured in terms of the number of target collapses. Target collapses were selected for this analysis because the primary goal of subjects in the simulation was to maximize reward points through target eliminations; therefore, it was expected that this goal and its relevant subtasks would be most significantly affected by any interruptions in, or disturbances to, task performance. Comparison was made of functioning during normal operations in the first 10-min block of each 20-min trial and performance in the remaining time period during which the simulation freezes for SAGAT occurred. There was no significant difference in target collapses, $F(1,1198) = 1.45$, $p = 0.2303$, between these two periods suggesting that the freezes were not intrusive or as equally intrusive as the automation failures.

An alternate interpretation of this result is that the automation failures did not have an impact on automated performance under normal operating conditions following a failure period (compared to periods in which no failure occurred). Therefore, there was no statistically apparent carry-over effect of manual control (level 1) performance during a failure on subsequent LOA performance. It is important that this inference be clearly separated from the highly significant effect of LOA on failure mode performance measured in terms of target collapses, expirations and time-to-recover (figures 5 and 6).

3.4. *Workload data*

ANOVA results for the NASA-TLX data revealed LOA to be significant, $F(9,20) = 4.48$, $p = 0.0025$. Comparison of the workload means across LOAs was conducted using Duncan's Multiple Range test at the $\alpha = 0.05$ level of significance to emphasize power in revealing truly significant differences. The results of the *post-hoc* comparisons are also presented in table 6. They reveal blended decision making (level 6), automated decision making (level 8), supervisory control (level 9) and full automation (level 10) to produce significantly lower ($p < 0.05$) subjective workload.

The mean NASA-TLX workload score across subjects for each LOA is shown in figure 9. Interestingly, the plot revealed a significant negative correlation, $r = -0.9021$, $p = 0.0004$, between the NASA-TLX means and the average percentage correct for level 2 SA questions. That is, those LOAs promoting improved operator understanding of target priorities and completion status (blended decision making (level 6), automated decision making (level 8), supervisory control (level 9) and full automation (level 10)) produced the lowest workload ratings, while all other levels yielded higher workload ratings.

Separate analyses were also carried out on the rating data for each of the NASA-TLX demand components, revealing the trends on mental, physical and temporal demands, as well as frustration and effort to mimic the pattern of overall workload across LOA shown in figure 9. Correlation analyses revealed mental demand ($r = 0.676$, $p = 0.0001$), physical demand ($r = 0.545$, $p = 0.0001$), temporal demand

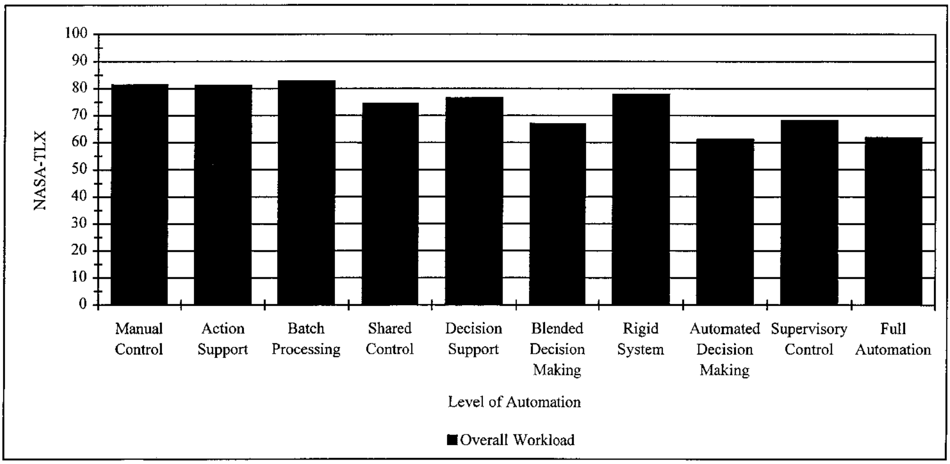

Figure 9. Mean NASA-TLX perceived workload for each LOA.

($r = 0.759$, $p = 0.0001$), frustration ($r = 0.448$, $p = 0.0001$) and effort ($r = 0.838$, $p = 0.0001$) to all share a significant positive relation with the overall NASA-TLX score.

A trend developed on the performance-rating dimension of the subjective workload measure revealed ratings of operator perceived successfulness in the task to increase with increasing LOA. A Pearson product-moment correlation coefficient developed on the performance ratings and overall workload scores demonstrated a significant negative relation ($r = -0.294$, $p = 0.0012$).

Since inconsistencies did not exist in the LOA effect on the dimensions of the NASA-TLX, use of the overall workload score in statistically evaluating changes in workload associated with the various LOAs was considered to be a reliable indicator.

These findings indicate that LOAs allocate the selection role to the computer or human-computer jointly provided reductions in NASA-TLX ratings. However, as with the level 2 SA response, whether the selection role was merely aided by, or completely allocated to, the computer did not result in perceived workload differences. Reductions in workload under LOA involving human-computer or purely computer strategy selection is most likely the factor allowing for the improvements in operator SA in this task.

4. Conclusions

Although care must be taken to generalize the results of human/system performance in this task (as it can be influenced by the relative capabilities of the human operators and computer algorithms that exist for a particular system), several key points can be made. Results suggest that LOAs distributing the roles of option generation and implementation between human and/or computer servers have a significant impact on automated system performance. Specifically, LOAs that combine human generation of options with computer implementation produce superior overall performance during normal operations, as compared to purely manual control and to higher LOA involving computer generation of options.

While computer selection (decision making) produced slightly better performance at the higher levels of rigid system (level 7), supervisory control (level 9) and full automation (level 10), this can be seen to occur only when coupled with computer generation of options and not when coupled with joint human-computer generation of options (automated decision making (level 8)). Computer selection was never better than when both human generation and selection of options were present. Clearly people benefited most from physical implementation assistance and were actually somewhat hindered when assistance was provided with higher level cognitive functions.

Under LOAs that provided computer guidance to subjects in regard to option selection, operators appeared to become distracted from task performance by new information or became doubtful of their own choices (although possibly optimal in nature). The fact that the joint human-machine generation of options produced worse performance than generation by either the human or machine component alone is significant. Most expert system and decision support systems being currently developed are directed at this type of interaction. Yet, significant difficulties were found when operators had system information available to use in conjunction with their own. This finding is in agreement with recent research by Selcon (1990), Endsley and Kiris (1995), Kibbe and McDowell (1995), and Smith *et al.* (1995) who have all noted difficulties in performance when humans are acting with the assistance of these types of aids.

Beyond the performance capabilities of the human and computer in the Multitask simulation, as dictated by the control allocations at each LOA, it is important to consider any potential effects of operator practice on the response measures. In this study, the subject population was comprised of naïve operators (university students) and not, for example, trained pilots or air traffic controllers. Subjects were trained in manual control (level 1) and in automated performance of the Multitask. The duration of the manual practice sessions (20 min) exceeded the time required for subjects to achieve peak performance (mean = 14 min), based on graphical analysis of training data. Further, the duration of automated performance practice sessions (10 min) appeared to be sufficient to promote asymptotic performance at the various LOAs, again, according to training data plots. Furthermore, since performance (measured in terms of target collapses) in the first 10 min of the task was not significantly different ($p < 0.20$) than in the second 10 min of the task, there is no evidence to suspect that training or practice had a major effect on this data.

Irrespective of the relative capabilities of the human and computer components, when human operators must take control in the event of an automation failure they are affected by the LOA they have been operating under prior to the failure. While automated implementation was related to improved performance under normal operating conditions, this was not the case for performance following an automation failure. This effect did not occur in a direct way, however. Primarily, implementation strategies that allowed for advanced process planning (through batch processing) were the most disruptive. It is most likely that this capability allowed operators to become distracted from the current task status (by focusing on future actions), thus producing decrements in response times and manual performance during automation failures. This may also be indicative of a lower level of direct feedback experienced when not actually performing task implementation (Kessel and Wickens 1982, Norman 1989) even though visual display feedback was equivalent in all conditions. Implementation strategies that provide assistance with the manual workload associated with a task while still keeping the operator involved in current operations appears to be optimal.

The pattern of SA and workload observed in this study is inconsistent with prior research by Endsley and Kiris (1995) who found better SA at intermediate LOA and poorer SA at full automation. They, however, found no difference in perceived workload across LOA. Explaining this difference requires an analysis of the way in which the task used here was implemented. This finding may well be an artefact of the relatively short task duration (20 min) involved in that there was not sufficient time for vigilance effects to take place and subjects were able to keep up reasonably well with monitoring the display for this duration. In addition, it should be noted that only the Multitask task was present. Parasuraman *et al.* (1993) found that vigilance effects in monitoring automation were only present when subjects were required to attend to three tasks, one of which was automated. When a single-task was present, as in this study, Thackray and Touchstone (1989) did not find out-of-the-loop performance decrements. In the present study, subjects were able to be reliant on the system for the short test period and were able to use the workload reduction provided by the higher LOA to maintain high SA. It is possible that this result would not be repeated in situations where: (1) subjects are required to attend to other tasks that are not automated; (2) subjects are required to perform the task over extended time periods; or (3) SAGAT queries are better integrated with regular system performance instead of being administered in a separate session. It is also possible that this result reflects a measurement effect. That is, subjects were better able to prepare for the queries due to freed-up resources under higher LOAs. This possibility needs to be further examined by future research that employs real-world tasks and experienced operators.

5. Future research

This paper has presented an empirical comparison of various level of automation options. The need exists for further research into how complex system performance is affected by LOA as a human-centred approach to automation. This type of research needs to be extended in the context of automation efforts in a variety of domains to determine its generalizability in realistic task settings. It does, however, provide a contribution to these efforts by offering insight into which factors are likely to make a difference (e.g. automation of task implementation) and which factors are not (e.g. allocation of responsibility within joint human-computer selection of options (decision making)). This type of guidance is particularly important when conducting high fidelity research in arenas such as air traffic control, piloting or advanced manufacturing where simulation costs are high and time consuming and access to expert operators is frequently limited.

These findings should also be further extended to dual task situations, which involve longer periods of performance. The Multitask simulation provides a controlled environment that is ideal for examining issues of vigilance, trust and complacency in automated systems. A challenging and worthwhile direction of future study involves scheduling different LOA across time to maximize human-computer functioning. A recent experiment conducted by Parasuraman (1993) achieved improvements in human monitoring of an automated system by scheduling manual performance at periodic intervals (every 10 min for a duration of 10 min) during fully automated operation. Further investigations are needed to assess whether adaptive automation techniques such as this can be used to allocate low and intermediate LOA over time to achieve improvements in performance. In addition, methods need to be developed for determining when such allocations should or should not occur.

In conclusion, the present research expands upon prior research that found a benefit for intermediate LOAs (Endsley and Kiris 1995) by examining their effect in the context of a dynamic control task. It also proposes an LOA taxonomy that appears to be viable in terms of its ability to explain differences in human/system performance in a systematic fashion. Further research is greatly needed in this area in order to provide useful design guidance for the developers of systems who are grappling with a wide variety of automation options, or who, worse yet, are not even considering the options available, but are assuming that full automation with human monitoring is the *de facto* choice.

References

Billings, C. E. 1991, Human-centered aircraft automation: a concept and guidelines, NASA technical memorandum 103885, NASA Ames Research Center, Moffet Field, CA.

Carmody, M. A. and Gluckman, J. P. 1993, Task specific effects of automation and automation failure on performance, workload and situational awareness, in R. S. Jensen and D. Neumeister (eds), *Proceedings of the 7th International Symposium on Aviation Psychology*, Ohio State University, Columbus, OH 167 – 171.

Endsley, M. 1987, The application of human factors to the development of expert systems for advanced cockpits, *Proceedings of the Human Factors Society 31st Annual Meeting* (Santa Monica, CA: Human Factors Society), 1388 – 1392.

Endsley, M. R. 1988, Design and evaluation for situation awareness enhancement, in *Proceedings of the Human Factors Society 32nd Annual Meeting* (Santa Monica, CA: Human Factors and Ergonomics Society), 97 – 101.

Endsley, M. R. 1995a, Towards a new paradigm for automation: designing for situation awareness, paper presented at the 6th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design and Evaluationa of Man-Machine Systems, MIT, Cambridge, MA, 421 – 426.

Endsley, M. R. 1995b, Measurement of situation awareness in dynamic systems, *Human Factors*, **37,** 65 – 84.

Endsley, M. R. and Kiris, E. O. 1995, The out-of-the-loop performance problem and level of control in automation, *Human Factors*, **37,** 381 – 394.

Goldstein, E. B. 1989, *Sensation & Perception*, 3rd edn (Pacific Grove, CA: Brooks/Cole).

Hart, S. G. and Staveland, L. E. 1988, Development of NASA-TLX (Task Load Index): results of empirical and theoretical research, in P. A. Hancock and N. Meshkati (eds), *Human Mental Workload* (Amsterdam: Elsevier Science/North-Holland), 139 – 183.

Kaber, D. B. 1996, The effect of level of automation and adaptive automation on performance in dynamic control environments, doctoral dissertation, Texas Tech University, Lubbock, TX.

Kessel, C. and Wickens, C. D. 1982, The transfer of failure detection skills between monitoring and controlling dynamic systems, *Human Factors*, **24,** 49 – 60.

Kibbe, M. and McDowell, E. D. 1995, Operator decision making: information on demand, in R. Fuller, N. Johnston and N. McDonald (eds), *Human Factors in Aviation Operations* (Aldershot: Avebury, Aviation, Ashgate Publishing), 43 – 48.

Moray, N. 1986, Monitoring behavior and supervisory control, in K. R. Boff, L. Kaufmann and J. P. Thomas (eds), *Handbook of Perception and Human Performance. Volume II: Cognitive Processes and Performance* (New York: Wiley).

Norman, D. A. 1989, The problem of automation: Inappropriate feedback and interaction not overautomation, ICS Report 8904, Institute for Cognitive Science, University of California at San Diego, La Jolla, CA.

Ntuen, C. A. and Park, E. H. 1988, Human factors issues in teleoperated systems, in W. Karwowski, H. R. Parsaei and M. R. Wilhelm (eds), *Ergonomics of Hybrid Automated Systems I* (Amsterdam: Elsevier), 230 – 210.

PARASURAMAN, R. 1993, Effects of adaptive function allocation on human performance, in D. J. Garland and J. A. Wise (eds), *Human Factors and Advanced Aviation Technologies* (Daytona Beach, FL: Embry-Riddle Aeronautical University Press), 147 – 158.

PARASURAMAN, R., MOLLOY, R. and SINGH, I. L. 1993, Performance consequences of automation induced complacency, *International Journal of Aviation Psychology*, **3**, 1 – 23.

SELCON, S. J. 1990, Decision support in the cockpit: probably a good thing?, in *Proceedings of the Human Factors Society 34th Annual Meeting* (Santa Monica, CA: Human Factors Society), 46 – 50.

SHERIDAN, T. B. and VERPLANCK, W. L. 1978, Human and computer control of undersea teleoperators, Technical Report, MIT Man-Machine Laboratory, Cambridge, MA.

SHIFF, B. 1983, An experimental study of the human-computer interface in process control, unpublished thesis, University of Toronto, Toronto.

SMITH, P. J. MCCOY, E., ORANASU, J., DENNING, R., VAN HORN, A. and BILLINGS, C. 1995, Cooperative problem solving in the interactions of airline operations control centers with the national aviation system, Technical Report, Ohio State University, Columbus, OH.

THACKRAY, R. I. and TOUCHSTONE, R. M. 1989, Detection efficiency on an air traffic control monitoring task with and without computer aiding, *Aviation, Space, and Environmental Medicine*, **60,** 744-748.

TULGA, M. K. and SHERIDAN, T. B. 1980, Dynamic decisions and work load in Multitask supervisory control, *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC-10**, 217 – 232.

VIDULICH, M. A. 1989, The use of judgment matrices in subjective workload assessment: the subjective WORkload Dominance (SWORD) technique, in *Proceedings of the Human Factors Society 33rd Annual Meeting* (Santa Monica, CA: Human Factors Society), 1406 – 1410.

WICKENS, C. D. 1992, *Engineering Psychology and Human Performance*, 4th edn (New York: Harper Collins).

WIENER, E. L. 1988, Cockpit automation, in E. L. Wiener and D. C. Nagel (eds), *Human Factors in Aviation* (San Diego, CA: Academic Press), 433 – 459.

WIENER, E. L. and CURRY, R. E. 1980, Flight-deck automation: promises and problems, *Ergonomics*, **23,** 995 – 1011.