Endsley, M. R., and Rodgers, M. D. (1998). Distribution of attention, situation awareness, and workload in a passive air traffic control task: Implications for operational errors and automation. Air Traffic Control Quarterly, 6(1), 21-44.

\

Distribution of Attention, Situation Awareness & Workload in a Passive Air Traffic Control Task:

Implications for Operational Errors and Automation

Mica R. Endsley,
SA Technologies,
Marietta, GA

Mark D. Rodgers,
Federal Aviation Administration,
Washington, D.C.

KEYWORDS: Attention, situation awareness, workload, human error, monitoring, automation, air traffic control

This research was conducted while Dr. Endsley was working at Texas Tech University and while Dr. Rodgers was working at the FAA's Civil Aeromedical Institute.

Abstract

A study was conducted to investigate factors underlying operational errors (OE) in en route air traffic control. Twenty active duty controllers watched re-creations of OEs and were asked to report on their situation awareness and workload at two stops during the re-creations. A total of 14 OEs were examined. Responses were analyzed to determine how subjects allocated their attention in viewing the scenarios. While observed patterns probably reflect necessary prioritization schemes, attention strategies identified in this study can be linked to data on factors underlying OEs. Both objective taskload, as indicated by the number of aircraft being controlled, and subjective workload were found to be related to controllers' ability to report situation awareness information. Workload was found to be higher at the time of the OE than at the other stop during the re-creation. During high workload, controllers appeared to reduce attention paid to certain aircraft and variables in order to maintain awareness of more important information. Implications of this research are drawn for potential problems in situation awareness under passive monitoring conditions that may be present if certain forms of automation are introduced in the future air traffic control system.

Introduction

In the history of the Federal Aviation Administration (FAA), no aircraft have collided while under positive control in en route airspace. However, aircraft have violated prescribed separation minima and approached in close proximity. This event can occur as a result of either a pilot deviation from clearances or an operational error (OE) on the part of the controller. An OE takes place when an air traffic controller allows less than applicable minimum separation between an aircraft and another aircraft or obstruction. Standards for separation minima are described in the Air Traffic Control (ATC) Handbook (FAA Order 7110.65) and supplemental instructions. While there is considerable complexity in those standards, the criteria are established as 2,000 ft of vertical separation or 5 miles of horizontal separation between aircraft operating at altitudes between 29,000 feet and 45,000 feet. For aircraft operating under instrument flight rules (IFR) at flight levels below 29,000 feet, a minimum of 1,000 feet of vertical separation or 5 miles of horizontal separation are required. These separation standards provide tolerance zones ensuring that aircraft pass well clear of one another.

A relatively small number of OEs occur nationwide each year. In 1993, 430 OEs were recorded at en route air traffic control centers in the U.S., as compared to 37,170,000 aircraft handled. In an effort to ensure flight safety, there is a desire to reduce the number of OEs that occur. Doing so requires an understanding of why these errors occur and the factors that are likely to increase the probability of an operational error.

Rodgers and Nye (1993) investigated causal factors associated with (minor and moderate) OEs occurring at en route ATC facilities over a three and one-half year period based on the FAA's Operational Error Data Base. This data base records circumstances associated with OEs as identified by quality assurance (QA) investigators following the OEs occurrence. They found that 36% of OEs involved problems with communications (including 20% that were specifically readback errors), 15% involved coordination problems, 3% involved deficiencies in position relief briefings, 13% were associated with problems in data posting, and 59% were related to the radar display (including 14% that involved misidentification of information and 47% that involved inappropriate use of displayed data). Some errors are attributed to multiple causal factors.

A number of research studies have sought to investigate the relationship between estimates of controller workload and the incidence of OEs. Operational errors have been found to occur under both high and low workload conditions, with more errors occurring under low and moderate levels of workload than under high levels of workload (Kinney, Spahn, & Amato, 1977; Schroeder, 1982; Stager & Hameluck, 1990) . It is unclear from these data, however, whether this reflects a decreased tendency to make errors under high workload conditions or a lower frequency of high workload conditions occurring overall.

A recent study by Schroeder and Nye (1993) found a positive correlation between the number of aircraft under the Air Traffic Control Specialist’s (ATCS) control (normalized for the average number of aircraft per ATCS in that center) and the occurrence of OEs involving data posting, position relief briefings and misuse of displayed radar data. They also found an association between OEs involving coordination problems and both a lower than average number of aircraft and a higher than average number of aircraft. There was no association found between number of aircraft and OEs involving communications problems.

It should be noted that these studies rely on estimates of controller workload which were made following the OE by FAA QA investigators who used a simple workload scale (1 to 5, non-anchored) that is not clearly defined. The investigators typically receive little or no training on what factors to include in their workload estimates. Thus non-standardization as well as potential inaccuracy of estimates made after the fact are limitations affecting the workload measures used in these studies.

There has also been an interest in determining to what degree OEs involve a problem with controller situation awareness (SA) — their mental picture of the constantly changing air traffic situation. Formally defined, SA is the “perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (Endsley, 1988) . It encompasses not only an awareness of specific key elements in the situation (Level 1 SA), but also an integration and comprehension of that information in light of operational goals (Level 2 SA), along with an ability to project future states of the system (Level 3 SA). These higher levels of SA (Levels 2 and 3) are felt to be particularly critical for effective functioning in complex environments such as air traffic control. In air traffic control, SA involves the continuous awareness of the location of each aircraft along with pertinent aircraft parameters (speed, heading, communications, etc.) and their projected future locations relative to each other, among many other pieces of information, in order to provide minimum separation and efficient aircraft movement. A complete delineation of SA information requirements for en route ATC is provided in Endsley and Rodgers (1994) .

Endsley (1995a) developed a taxonomy of SA errors, describing causal factors associated with the occurrence of SA errors. This study found that 88% of major air carrier accidents associated with pilot error involved a problem with situation awareness. Of these, 72% involved problems with Level 1 SA , 22% involved problems with Level 2 SA, and 6% involved problems with Level 3 SA.

The SA error taxonomy was applied in a recent study of 146 incidents involving reported problems in SA among both pilots and controllers in NASA's voluntary Aviation Safety Reporting System (ASRS) (Jones & Endsley, 1996) (Only incidents involving air traffic control are discussed here). Of the 33 incidents involving air traffic controllers, 69% involved problems with Level 1 SA, 19% involved problems with Level 2 SA, and 12% involved problems with Level 3 SA. Of the Level 1 SA errors, the most common problem was a failure to monitor or observe data (51.5%). This was most frequently due to task distraction (53% of these cases), followed by problems with high workload (17.6%), vigilance (11.8%), and other miscellaneous causes (17.7%) such as a failure to visually scan the runway, failure to notice an aircraft overshoot and failure to notice traffic on the runway. Other causal factors were also related to Level 1 SA errors: 18.2% involved cases where needed data was not available, 18.2% involved cases where controllers forgot important information (frequently under high workload), 6.1% involved data that was hard to discriminate or detect, and 6.1% involved the misperception of information.

Level 2 SA errors were attributed to an incomplete or inaccurate mental model (22.2%), the use of an incorrect mental model (22.2%), over-reliance on default values (22.2%), and other miscellaneous factors (33.3%). Level 3 SA errors were attributed to over-projection of current trends (33.3%) and other miscellaneous factors (66.7%). There were no obvious cases of Level 3 errors due to poor mental models in this data.

It should be noted that the errors by air traffic controllers examined in the Jones and Endsley study involved voluntarily reported information from a variety of ATC facilities including Air Route Traffic Control Centers (ARTCC), TRACONs, and towers (both local and ground control). As such, while the data does provide some information about the types of errors that may occur across different types of ATC, this cannot be viewed as a truly representative, or completely unbiased, sample of controller errors. In addition, it is often difficult to ascertain exactly why some of the errors occurred from the limited information available in such reports.

Most information about errors is based on the analysis of available historical reports. These reports are often not developed with the objective of examining detailed causal factors, are usually based on after-the-fact interviews which may be incomplete or biased, and frequently suffer from problems of inconsistency as different people usually conduct each investigation. The objective of the present study was to collect more detailed data about OEs than what is available in such accounts, providing for a better understanding of factors that may contribute to their occurrence. This study focused on data gathered on OEs from the Atlanta Air Route Traffic Control Center.
To gain more insight into the nature of OEs, the Systematic Air Traffic Operations Research Initiative (SATORI) system was developed (Rodgers & Duke, 1993) .

SATORI graphically recreates a visual display of the radar data recorded during actual air traffic control (based on computer tapes routinely recorded at each air traffic control facility) synchronized with the recorded audio communications between controllers and between controllers and pilots. Prior to the development of the SATORI system, it was not possible for the FAA QA team investigating errors to review the control situation in a format like the one presented to the controller when the OE occurred. That is, the dynamics of the situation (the interaction between control actions and displayed data) were unavailable for review, not only by the QA team investigating the irregularity, but also by the controller who committed the error. This limited not only the extent to which reliable and accurate determinations of causal factors could be made for an error, but also the extent to which the effects of the dynamic situation on controller SA could be determined. The SATORI system allows QA specialists and controllers to view a high fidelity, dynamic representation of the ATC data associated with an OE in a format much like the one presented to the air traffic control specialist (ATCS).

In the present study, the recreation of OEs using SATORI was combined with a modification of the Situation Awareness Global Assessment Technique (SAGAT) (Endsley, 1988) . SAGAT is a technique used during simulations in which the simulation is frozen at random, unexpected intervals with all display screens blanked and the operator of the simulation is queried about the state of the current situation. The operator's perceptions are then compared to the actual state of the environment to provide an objective assessment of the operator's situation awareness. Although the participants in this study are all active controllers who are familiar with the airspace examined, it is possible that they may process the recreated scenarios differently than they would when actively controlling air traffic. Therefore, some caution is needed when interpreting the results. The combination of the two techniques, however, may provide information that would be difficult to obtain otherwise. The use of SAGAT to measure situation awareness in aircraft simulations has been extensively validated (Endsley, 1990a; 1990b) .

In this study, SAGAT was used with queries that pertain to major factors associated with SA in en route ATC, based on an analysis by Endsley and Rodgers (1994) . As a modification, the technique was employed in conjunction with SATORI which involves the passive viewing of a situation as opposed to an interactive simulation in which the subject is involved. While it is not clear how the SA of a passive observer differs from that of an active participant, this measure should still provide an indication of the way in which controllers distribute their attention to various factors involved in the scenarios. As this study involves currently certified controllers viewing re-creations of real OEs, the combined use of SATORI and SAGAT may provide unique insight into factors affecting OEs in operational settings. Due to their rare occurrence and the limited conditions usually involved in simulations, observing OEs in simulations can prove to be quite difficult.

In addition, it may be considered that SA under passive viewing conditions may be analogous to that expected if the ATC system ever becomes highly automated. Under conditions of high automation the role of the controller would become one of monitor of an air traffic situation that is controlled by an automated system. The scenarios recreated for subjects using SATORI similarly involve the passive monitoring of a situation that is actually controlled by another. The study may therefore provide some insights into SA with a hypothetical automated system. This is of concern as there is some indication that SA may be compromised under highly automated systems (Endsley & Kiris, 1995) .

Method

Subjects

Twenty volunteer subjects participated in the study. All subjects were experienced, full performance level (FPL) status Air Traffic Control Specialists at Atlanta ARTCC. The 20 subjects included four subjects from each of five areas of specialization in the facility. Area of specialization refers to the set of sectors and their associated configurations in which a controller is certified to work. All subjects were certified in the area of specialization for the re-created errors that they observed during the study. Subjects were relieved from their duties in the air traffic control room to participate in the study. Once subjects completed their participation, they returned to their assigned duties.

Procedure

Fifteen OEs that occurred in the Atlanta Air Route Traffic Control Center (ARTCC) in 1993 and 1994 were recreated using SATORI. These errors were selected from errors involving a single ATC sector, subject to the availability of complete and legible computer data tapes and audio recordings. Three errors in each of five areas of specialization of the center were selected. (One error was eliminated during testing due to problems with the data tapes, leaving 14 errors to be included in the analysis.)

Subjects were shown scenarios involving three errors from sectors in the area of specialization on which they were certified. Each scenario was viewed by four subjects. The subjects were instructed to observe the scenarios as if they were the controller working the sector. Each scenario consisted of a recreation of the ten minutes immediately prior to the occurrence of the OE. SATORI presented all aircraft required to be displayed to the controller, including limited datablocks for aircraft in airspace near the sector boundaries (i.e., airspace adjacent to, above, and below the error sector). Extraneous information not required for display was not included. The number of aircraft in each scenario varied over the ten minute recreations, which is reflective of typical traffic variation. Twice during each scenario, the recreation was halted and the screen blanked. The first freeze occurred two minutes prior to the occurrence of the error and the second freeze occurred at the time of the OE in each scenario. Although subjects were informed that freezes would occur, they were not informed as to the timing of the freezes or the occurrence of the error. While subjects may have figured out there might be an operational error in the scenarios, they did not know when it would occur or which aircraft it may involve.

During each freeze subjects were provided with a map of the sector. Sector boundaries, navigation aids, airways and intersection markings were shown on the map, however, no aircraft were included. Subjects were asked to indicate the location of all known aircraft on the map, and, for each aircraft, to indicate or make a judgment of:

(1) if the aircraft was:

(a) in the displayed sector's control,

(b) other aircraft in the sector not under sector control, or

(c) would be in the sector's control in the next two minutes,

(2) aircraft call sign, (3) aircraft altitude, (4) aircraft groundspeed, (5) aircraft heading, (6) the next sector the aircraft would transition to, (7) whether the aircraft was climbing, descending or level, (8) whether the aircraft was in a right turn, left turn or straight, (9) which pairs of aircraft had lost or would lose separation if they stayed on their current (assigned) courses, (10) which aircraft would be leaving the sector in the next two minutes, (11) which aircraft had received clearances that had not been completed, and, for those, whether the aircraft received its clearance correctly and whether the aircraft was conforming to its clearance, and (12) which aircraft were currently being impacted by weather or would be impacted in the next five minutes. Of these, queries 1, 2, 3, 4, 5, 7, and 8 can be regarded as pertaining to Level 1 SA, and queries 6, 9, 10, 11, and 12 can be regarded as pertaining to SA Levels 2 and 3.

Following the completion of the questionnaire, each subject completed a NASA-TLX subjective workload rating (Hart & Staveland, 1988) indicating the amount of workload they felt they would be under if they were controlling the traffic in the scenario presented. Following completion of the NASA-TLX questionnaire, the scenario was resumed until the second freeze. At that time, the SA queries and NASA-TLX questionnaire were again presented in the same order, following which the scenario was terminated and the next scenario presented. After all three scenarios had been presented, subjects completed a NASA-TLX workload paired comparison ranking form, allowing each subject's ratings on each NASA-TLX subdimension to be weighted based on the subjective importance of the subdimension to each subject.

Apparatus

SATORI recreations were presented on a DEC 3000-300 Alpha computer system using dual Sony 19-inch high-resolution (1280 x 1024) color monitors. NASA-TLX ratings were obtained using Hypercard on a Macintosh Powerbook.

Results

SA Questionnaire

Subjects' responses to each question were scored for accuracy based on computer data for each aircraft at the time of each freeze. Subjects' indications of each aircraft's location on the map were matched to the closest aircraft actually present in the sector at the time of the freeze and the distance error recorded. The percentage of aircraft present that were reported by the subject was calculated. Following that, scoring for each subsequent question was calculated as the number of correct responses compared to the number of aircraft that the subject reported knowing about (e.g. percent correct for altitude was calculated as the number of correct aircraft altitudes reported divided by the total number of aircraft reported.) Subjects' responses for each question were scored as either correct or incorrect based on operationally determined tolerance intervals (as listed in Table 1). Missing responses were scored as incorrect. Although mean scores are rather low on some items, this is most likely reflective of the large amount of information (many aircraft each with many parameters) that must be attended to. An information sampling strategy would result in a lack of SA on lower priority, infrequently scanned information. This would most likely be accompanied by a fairly high standard deviation as was observed for some variables.

The frequency of correct responses on each variable provides some insight into the tradeoffs that controllers make in allocating their limited attention across multiple aircraft and pieces of information that compete for that attention. This analysis is not meant to be critical regarding the information controllers did not attend to or retain in working memory. Attention allocation strategies, such as those indicated here, are needed and are effective the vast majority of the time in dealing with the demands of controlling air traffic as can be demonstrated by the effective daily performance of controllers and relatively low error rates nationwide. A point that can be made, however, is that these strategies may lead to a lack of situation awareness that occasionally (due to a probabilistic link between SA and performance (Endsley, 1995b)) results in errors. This point is reinforced in that the patterns of attention demonstrated here can be correlated with certain systematic characteristics of OEs.

Means, standard deviations, and significance for subject response accuracy are shown in Table 1. On average, 12.8 aircraft were actually present at the time of the freezes (range 4 to 23) across scenarios. Of these, subjects on average reported 8.0 aircraft or 67.1% of the aircraft present. Mean distance error was 9.6 miles (.68 inches) from the aircraft's reported location to their actual location. These results indicate that subjects may have sacrificed keeping up with aircraft dynamics (at least for many aircraft), as their attention was directed to other aspects of the situation.

Variable	Mean	Std. Dev.	df	F
Actual aircraft present (number)	12.9	5.6		
Aircraft reported (%)	67.1	18.0	1,110	54.409 **
Distance error (miles)	9.6	4.5	1,110	6.091 *
Control level (% correct)	73.8	17.3	1,110	13.324 **
Call sign: alphabetic (% correct)	79.9	23.4	1,110	8.945 **
Call sign: numeric (% correct)	38.4	32.0	1,110	12.619 **
Altitude (+/- 300 feet) (% correct)	59.7	22.1	1,110	9.556 **
Change in altitude (Climbing, Descending) (% correct)	66.4	25.6	1,110	6.660 *
Speed (+/- 10 knots) (% correct)	28.0	25.6	1,110	19.400 **
Heading (+/- 15 degrees) (% correct)	48.4	30.6	1,110	4.187 *
Turn (% correct)	35.1	40.2	1,110	.695
Separation problems (% correct)	86.2	32.3	1,110	.386
Transition to next sector (% correct)	63.5	45.1	1,110	16.200 **
Assigned clearances complete (% correct)	23.2	22.9	1,110	8.479 **
Assigned clearance correct (% correct)	74.4	43.9	1,80	.003
Assigned clearance conformance (% correct)	82.9	37.9	1,80	2.789
Weather impact (% correct)	60.7	49.1	1,110	8.898 **

\- *all measures expressed in percentages were subjected to an arcsine transformation prior to analysis*

* *significant at a= .05 level*

** *significant at a = .01 level*

Table 1. Awareness of Situation Across all Subjects and Scenarios

For the aircraft reported, the correctness of subject responses on the remaining questions was calculated. Subjects correctly identified the control level of the aircraft (in sector control, other aircraft in sector, will be in sector control in the next 2 minutes) for 73.8% of the aircraft reported. Aircraft callsigns were often incomplete. The initial alphabetical part of the callsign (indicating airline company, military or civil aircraft designation) was reported correctly 73.8% of the time. The numerical part of the callsign (the aircraft identification number) was reported correctly for only 38.4% of the aircraft. It should be noted that other studies have found that in general 4% of OEs involve readback errors associated with aircraft identification (Rodgers & Nye, 1993) . The low level of accuracy in recall knowledge of aircraft callsigns is probably highly indicative of these readback errors, as it indicates that controllers may not attend to or retain much information on aircraft callsign in working memory, particularly the identification number.

Aircraft altitude was correctly reported (+/- 300 ft) for 59.7% of the aircraft (mean error of 655 ft). The aircraft were correctly identified as ascending, descending or level 66.4% of the time. Correct groundspeed (+/- 10 knots) was reported for only 28.0% of the aircraft (mean error 21.8 knots). Correct aircraft heading (+/- 15 degrees) was reported for 48.4% of the aircraft (mean error 15.6 degrees). Only 35.1% of the aircraft were correctly identified as being in a left turn, right turn or proceeding straight ahead. These results indicate that subjects were fairly poor at reporting with the dynamics of the aircraft in the scenario, at least for many of the aircraft. While flight strips are normally present during ATC operations, thus reducing the demand on working memory, the results revealed by the SAGAT may provide an indication of the controllers' internal mental model of the air traffic situation. This internal model is needed in order to understand what is happening and formulate an air traffic control plan. Thus, it is reflective of the controller's internalization of the traffic scenarios, including the way in which they allocated their attention across information and their higher level assessments of the situation.

An argument can be made that perhaps subjects simply did not retain this type of detailed information about each aircraft (Level 1 SA), instead maintaining awareness of higher level situation comprehension and projection issues (e.g. aircraft separation and future projections of actions). Previous research, however, indicates that people do maintain task relevant information about Level 1 SA elements that can be reliably recalled under SAGAT testing when actively performing in a simulation (Endsley, 1990a) and when passively monitoring automated systems (Endsley & Kiris, 1995; Endsley & Kaber, in review). There is also evidence that this measurement technique is reflective of subject attention allocation across sources of information (Fracker, 1990) . It is more likely, therefore, that these measures do provide some indication of the ways in which subjects in this study were deploying their attention across displayed information, at least on a relative basis.

The subjects' higher level of understanding of the scenarios was also evaluated. The aircraft pairs that the subjects identified as having "lost or will lose separation if they stay on their current (assigned) courses" was compared to those aircraft that actually had lost or would lose separation (in the following 2 minutes) at the time of the freeze. Subjects correctly identified 86.2% of these aircraft pairs. (Aircraft pairs that the subject identified as having potential separation problems, but did not, were not scored.) Subjects correctly identified 63.5% of aircraft that would be leaving the sector in the next two minutes. Thus, they did not appear to be fully aware of upcoming sector transitions.

Subjects correctly identified only 23.2% of aircraft that had not yet completed control assignments. Of those that they identified as not having completed an assignment, subjects were correct in 74.4% of the cases in their identification of whether the aircraft had correctly received its assignment, and in 82.9% of the cases in their identification of whether the aircraft was conforming to its assignment. Overall, subjects did not attend well to an aircraft after a clearance was given in terms of monitoring for compliance or progress in completing the control action, most likely because they were concentrating on other traffic present. They may have monitored for compliance more actively if they had issued the assignment themselves.

Subjects were incorrect in identifying weather as a current impact (or impact in the next 5 minutes) in 39.3% of the scenarios. This is perplexing in that even though light and heavy weather symbols were displayed in some scenarios, poor weather did not actually impact traffic in any of the scenarios presented. This finding most likely indicates that controllers have difficulty estimating the impact of weather on air traffic based on available data, an issue which has previously been raised by controllers. Subjects were most likely guessing due to their lack of information regarding weather.

It should also be noted that a fairly high degree of variability was present on many of the variables, across aircraft, subjects, freezes and scenarios. Possible sources of these variations will be examined more closely in the following sections.

Analysis of Freeze Number

An analysis was conducted to ascertain whether there was a difference between the first and second freezes in the subjects' ability to correctly identify dynamic elements in the scenarios. The first freeze always occurred two minutes before the OE and the second freeze always occurred at the time of the error. A multivariate test was performed on the accuracy of subjects' responses across the queries to examine differences between these two freeze times. (All measures expressed as percent correct were subjected to an arcsine transformation prior to analysis in order to meet the conditions of ANOVA.) The MANOVA was not significant, $F(13,98) = 1.554$, $p>.05$. Therefore, subjects' recall of the situation was not significantly different as a function of the presence of an OE.

An analysis was also conducted to determine whether subjects reported a different level of subjective workload between the two freezes, as there is a concern that higher workload may be associated with the occurrence of OEs. The NASA-TLX ratings were weighted based on each subject's rankings and a combined workload score was calculated. An ANOVA was conducted on the combined workload ratings to test for differences between the two freezes. The overall NASA-TLX workload rating was significantly higher at the time of the second freeze (during the OE), $F(1,109) = 24.08$, $p<.001$.

To further investigate, ANOVAs were performed on each of the subscale ratings (performance, temporal demand, frustration, mental demand, effort, and physical demand), revealing that ratings were significantly higher at the second freeze for all of the subscales ($p<.05$) except for physical demand, which is as would be expected. This supports the contention that higher workload is associated with OEs, however, it is unclear whether higher workload caused the error, or whether the higher workload ratings were the result of the error.

Analysis of Workload Impact

Subjective workload. Since subjective workload was higher at the time of the OE, further analysis was conducted to determine if there was a direct relationship between the NASA-TLX combined rating and subjects' accuracy in their awareness of the situation. A MANOVA showed a significant relationship between situation awareness accuracy and the NASA-TLX score, F (14,96) = 1.93, p <.05. (Assignment correct and assignment conformance were not included in the multivariate analysis as they would significantly reduce the degrees of freedom in the test. These two questions had a lower sample size resulting from the fact that they were only asked for aircraft which had yet not completed their assigned clearances). Regressions were performed on the subjects' scores on each query (in terms of percentage correct) in comparison to their workload ratings at the time. Those variables scored as correct or incorrect were subjected to an analysis of variance to determine whether workload ratings were associated with accuracy of recall.

The regressions revealed that subjects reported a significantly lower proportion of the aircraft present, F(1,109)=7.418, p=.008, and identified aircraft heading correctly significantly less often for those aircraft, F(1,109)=5.484, p=.021, as subjective workload level increased, as shown in Figures 1 and 2 (each data point represents each freeze for each controller/scenario). In addition, with increasing levels of workload, subjects were significantly less likely to correctly identify whether aircraft had completed their assigned clearances. F(1,109)=7.573, p=.007, and were significantly less likely to identify whether an aircraft had received its assigned clearance correctly, F(1,79)=5.705, p=.019. No other SA measures were significantly affected by workload.

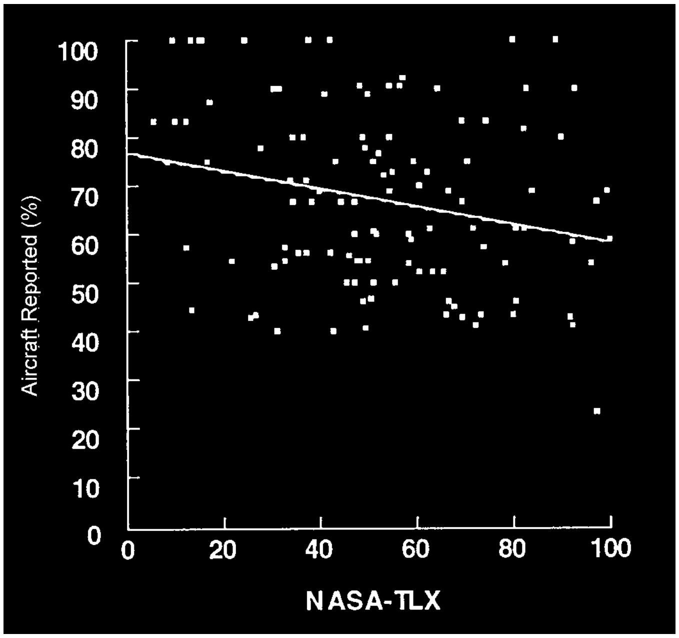

Figure 1. Impact of Subjective Workload on Aircraft Reported

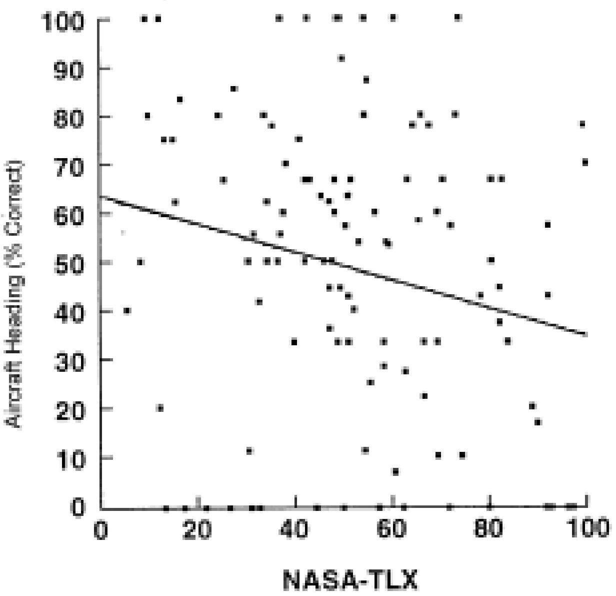

Figure 2. Impact of Subjective Workload on Awareness of Aircraft Heading

Number of aircraft. An objective measure of taskload was also examined to determine its impact on subject awareness and perceived workload. The number of aircraft present at the time of each freeze was calculated. A MANOVA showed a significant relationship between subject situation awareness and the number of aircraft present at the time of each freeze, F (14,97) = 9.50, p <.001. Regressions were performed to examine the relationship between number of aircraft present and accuracy on each question (in terms of percentage correct). Those variables scored as correct or incorrect were subjected to an analysis of variance to determine whether number of aircraft was related to accuracy.

As the number of aircraft increased, the percentage of aircraft present that subjects reported significantly decreased, as shown in Figure 3 (each data point represents each freeze for each controller/scenario). For those aircraft that subjects did report, subjects were also significantly less accurate on most of the other factors about those aircraft as the number of aircraft present increased. They were significantly more erroneous in their awareness of the location of the aircraft, and correct less frequently regarding the aircraft's control level, both the alphabetic portion and numeric portion of the call sign, altitude, change in altitude, airspeed, and heading as the number of aircraft increased. They were also correct significantly less frequently in their awareness of which aircraft would transition out of the sector in the next two minutes, which aircraft had completed their assigned clearances, and if weather was an impact as the number of aircraft increased. Interestingly, the number of aircraft present did not impact subjects' accuracy in reporting which aircraft had a potential or current separation problem.

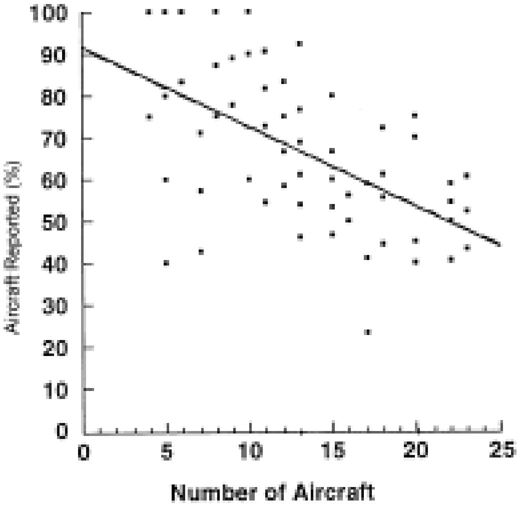

Figure 3. Impact of Number of Aircraft on Aircraft Reported

Correlation between measures. An analysis was also conducted to determine the relationship between the number of aircraft present, as an objective measure of taskload, and subject's reported subjective workload. A regression showed a significant relationship between these two measures of load, F(1,109) = 6.45, p = .013. It appears that the number of aircraft present in a scenario is a significant driver of perceived workload that negatively impacts subjects' ability to keep an accurate mental picture of the situation, independent of any other factors that might also drive subjective workload. An increase in the number of aircraft present not only was related to a tendency to attend to fewer aircraft, but also to the tendency to know significantly less about these aircraft. This load related to shedding of information and appears to reflect some prioritization of tasks, as awareness of those aircraft with current or potential separation problems was not significantly impacted by workload.

Analysis of Operational Errors

A closer examination of the nature of the OEs included in this study was made. Table 2 provides a summary of each error. Each error was classified in terms of the SA Error Taxonomy (Endsley, 1995a). Classifications were made by two independent raters based on a description of each OE contained in the Final Operational Error/Deviation Report completed by internal FAA QA investigators after the OE, an analysis of each OE from the SATORI recreation, and verbal comments made by the subjects in this study.

Scenario	Description of Error	SA Error Type	Additional Information
1	Clearance to wrong aircraft	Level 1 - misperception Level 1 - failure to monitor (task distraction)	Confused two aircraft Attending to other aircraft
2	Descended aircraft into other aircraft	Level 2 - inadequate mental model Level 3 - lack of projection	
3	difference in climb/closure rates of two aircraft	Level 3 - lack of projection	
4	difference in climb/closure rates of two aircraft	Level 3 - lack of projection	
5	Expected aircraft to descend faster than it did	Level 2 - over-reliance on defaults	
6	Readback error	Level 1 - misperception (expectations)	
7	Delay in turning aircraft to accommodate slow descent of other aircraft	Level 3 - lack of projection	Didn't turn soon enough
8	Forgot aircraft - provided inadequate clearance	Level 1 - memory loss task distraction	
9	Did not separate aircraft	Level 1 - failure to monitor task distraction	Other radio calls
10	Readback error - aircraft at different altitude	Level 1 - misperception task distraction	
11	Gave wrong heading command Didn't monitor compliance	Level 1 - failure to monitor	
12	Descended aircraft into other aircraft	Level 1 - failure to monitor Level 2 - over-reliance on defaults	
13	Climbed aircraft into other aircraft didn't judge separation	Level 1 - task distraction	
14	Issued clearance to aircraft off frequency Lost track of slow aircraft climb rate	Level 1 - memory loss - workload Level 1 - failure to monitor- workload	

Table 2. Summary of Operational Errors Investigated

Of the 14 errors investigated, five clearly involved task distractions in which the controller was distracted by the need to attend to other aircraft in the sector. Three involved the misperception of some information (due to expectations, workload or task distraction). Two OEs involved memory loss (associated with task distraction and high workload) in which the controller forgot about an aircraft or a previous action. Two OEs involved an over-reliance on defaults, expecting aircraft to behave as they usually do. Four involved problems with inadequate projection of the dynamics of the aircraft to anticipate separation problems. High workload was specifically cited as a problem in one OE.

In view of the fact that controllers routinely must cope with competing task demands and high workload situations, an analysis was made to determine whether the situations in which these OEs occurred were intrinsically likely to induce errors. Each of the investigated errors is listed in Table 3, along with the number of aircraft in the sector at the time of the OE and a rating of the complexity of the scenario (as reported in the Final Operational Error/Deviation Report). The measure of air traffic complexity was a rating made by the quality assurance specialist of the difficulty of the job tasks, based on factors such as weather, airspace restrictions, and variety of duties. The complexity rating is made on a 5-point scale with anchors 1 = "easy," 3 = "average," and 5 = "complex." (This information was not included in the reports for Scenarios 6, 12 and 13.) Scenarios 1, 2, and 14 had the highest level of complexity and highest number of aircraft in the sector at the time of the error, followed by scenarios 9 and 11. Scenarios 3, 4, 7 and 10 had lower than average complexity and number of aircraft at the time of the error. This agrees with previous findings that many OEs occur during low workload situations.

Scenario	Complexity Rating	# of A/C in Sector	Awareness of Error @ Occurrence	Subjects Aware of Involved AC @ Stop1 (%)	Subjects Aware of Involved AC @ Stop2 (%)	Subjects Aware of error @ Stop 1 (%)	Subjects Aware of error @ Stop 2 (%)
1	5	14	no	75	87	-	75
2	5	12	no	87	87	-	75
3	5	12	yes	75	100	25	100
4	4	10	yes	87	100	50	100
5	4	8	no	100	87	-	75
6	3	5	yes	100	87	-	75
7	3	-	no	75	87	-	75
8	2	7	no	87	100	25	75
9	2	5	no	100	100	-	100
10	1	2	yes	100	100	-	100
11	1	3	no	100	100	50	100
12	-	-	no	100	100	-	100
13	-	-	-	87	100	-	75
14	-	-	-	67	100	-	50

Table 3. Subject Awareness of Error

Table 3 also lists whether the controller making the OE was aware that the OE was occurring at the time based on the Final Operational Error/Deviation Report. (This information was not available for two of the errors.) In four out of the remaining 12 OEs, the controller was aware that an error was building, but was not able to avoid it. In the remaining eight, the controller was not aware that an OE had occurred.

As a comparison, the responses of the subjects in this study were examined to determine whether they were aware that separation errors were developing. The percentage of the four subjects viewing each scenario who reported the existence of the two aircraft involved in the OE at each of the two freezes is listed in Table 3. While for many of the scenarios both aircraft were reported by all four subjects, at least one of the involved aircraft was not reported at all by at least one subject in approximately one-half of the scenarios at the first freeze and approximately one-third of the scenarios at the second freeze.

In their response to the question regarding which aircraft had lost or would lose separation in the next two minutes, in only four of the scenarios did one or more of the four subjects viewing the recreation list the two aircraft involved in the OE at the freeze which occurred two minutes before the OE. At the time of the second freeze, when the OE occurred, at least one of the four subjects did not identify the separation error in eight of the fourteen scenarios. In three of these cases they reported both aircraft, but did not indicate a separation problem. In the remaining five scenarios, they also did not report at least one of the involved aircraft, indicating it was outside of their focus of attention.

Based on this analysis, it would appear that the factors associated with many of these OEs were significant enough to be problematic for other trained controllers. This result should be viewed with caution, however, as the method employed in this study involved passive viewing of the scenarios instead of actual operational control.

Discussion

This study reveals many interesting findings on the role of situation awareness and workload in operational errors. Significant deficiencies in the ongoing situation awareness of the subjects were present in this study. They had a fairly low ability to report on the existence of many aircraft, or accurately recall their location or many of their parameters. Their accuracy was significantly impacted by the number of aircraft present in the scenario and, to a lesser degree, by perceived workload. After the number of aircraft present exceeded approximately eight to ten, the ability of subjects to report on each aircraft declined quite rapidly. This finding is consistent with the classic study by Sperandio (1971) that showed controllers handled an unexpected increase in traffic load adaptively by decreasing the amount of time they spent processing each aircraft. Even for those aircraft they did report on, their awareness of the relevant parameters for these aircraft also declined when there were more than approximately ten aircraft present. In the face of a high number of aircraft, subjects tended to attempt to maintain their awareness of aircraft separation, however this was still less than perfect (86.2% correct on average). Other tasks, such as follow through on clearances given to aircraft, also appeared to suffer under increases in the number of aircraft and perceived workload.

While it is difficult to say that controllers need to be able to remember aircraft parameters as long as they know about aircraft separation, the pattern of attention represented by the accuracy scores in this study are indicative of many of the OEs that occur. Readback errors, for instance, may be directly related to the tendency for subjects in this study to have a fairly low awareness of whether aircraft given clearances had received the clearance correctly. (Subjects were only correct in 23.2% of the cases in knowing if an aircraft had completed its assigned clearance, and, of these, in only 74.4% of the cases in knowing if the aircraft had received its assigned clearance correctly.) Problems in reporting the numeric portion of callsigns is also reflective of OE error patterns. Many of the OEs included in this study involved over-reliance on expectations about how fast aircraft would travel or should ascend or descend. Low accuracy in reporting on the speed and heading of aircraft and being able to report whether aircraft were turning, ascending or descending are most likely indicative of this type of error.

An important issue is why these SA problems occur. It is a mistake to infer that these subjects (or the controllers involved in the errors) were simply inattentive. All people have a limited amount of attention to distribute in any situation (Wickens, 1992) . In complex activities, such as air traffic control, there is a great deal of information to process. Controllers do not fully commit all details to memory such that they can recall them verbatim (or to the accuracy level required by SAGAT), several seconds (or even a few minutes) later (i.e. the delay characteristics of the SAGAT retrieval). During normal ATC operations there is little functional reason for a controller to commit to memory these changing parameters, since memory would constantly have to be refreshed. Perceptual information is continuously available on the PVD (barring the very infrequent failures), and is also more permanently available on flight progress strips should it be needed in response to PVD failure. Hence, perhaps the most important functional knowledge that should be committed to memory (and does not change as rapidly) is the general knowledge ("gist") of the status of pending conflicts.

The pattern of errors in this study suggests that even if subjects made as much use of their attention as possible in keeping up with the scenarios, they may have had to limit attention to some information in order to keep up with the need to ensure that all aircraft were separated. Probabilistically, this strategy is effective much of the time, but is likely to produce occasional errors of the type described here.

Information regarding the role of workload in the occurrence of OEs is also present here. Subjective workload as measured by NASA-TLX was significantly higher at the time of the operational error than it was two minutes prior in the scenarios. (The number of aircraft was not higher, however.) While the NASA-TLX score was correlated with the number of aircraft present, the number of aircraft present was more closely related to decreases in subjects' awareness of the situation (beyond a certain number of aircraft). What other factors are included by subjects in their subjective workload ratings is unknown.

It is important to note, however, that low and moderate levels of SA were also demonstrated by many variables even when the number of aircraft was relatively low. This finding concurs with other studies that have found that SA and workload can operate independently for various reasons (Endsley, 1993) . OEs were also found to occur under high, moderate and low load conditions (as indicated by the number of aircraft present). Thus, while high workload and in many of these cases, momentary task distraction, can be associated with loss of SA and OEs, errors can also occur for other reasons at lower levels of workload.

A possible source of caution in generalizing the results of this study is that it involved subjects who, although current, experienced controllers, passively viewed the ATC scenario recreations. While this procedure provided insight into factors affecting actual OEs (which can be difficult to produce under simulated conditions), one can speculate as to whether their situation recall accuracy and subjective workload ratings are the same as they would be for controllers actively working the same scenarios. For instance, the absence of flight progress strips in this study may have caused the subjects to rely on their working memory for specific information which active controllers have routinely available on flight progress strips, consequently lowering the subjects accuracy scores. In a strip-less system, controllers would be faced with a similar situation as subjects in this study, and thus we might expect an increase in recall errors for callsign and other aircraft specific information. In addition, the subjects in this study did not have to generate a plan, but rather monitor someone else's plan, potentially contributing to the low accuracy scores. This finding might generalize to a controller out of the loop situation where the controller is relegated to monitoring computer generated plans. While it is difficult to stipulate that the absolute levels of SA and workload reported would be the same during active control, the general patterns presented here are probably valid as reflections of subjects' attention distribution.

It is somewhat likely, however, that levels of SA may be lower under passive viewing conditions rather than active decision making, such as has been demonstrated under higher levels of automation in recent research (Endsley &

Kiris, 1995) . In this light, difficulties in accurately identifying aircraft separation problems shown by subjects in this study may be at least partially reflective of the difficulties associated with passive monitoring. This possibility needs to be seriously investigated with regard to systems being developed for automating future air traffic control.

As a basis of comparison, Mogford and Tansley (1991) , using a procedure similar to SAGAT during actual simulations of air traffic control with controller trainees, found that subjects were able to report aircraft position with 86% accuracy, heading 82%, altitude 73%, callsign 55% and speed 53%. In comparison with the present data obtained during passive viewing, the attention devoted to each type of information is similarly distributed; however the Mogford and Tansley study showed much higher levels of SA, even though they studied controller trainees who would be theorized to have lower SA than Full Performance Level controllers. This supports the contention that SA may be compromised under passive viewing, which should be a significant concern for automation systems designed to place the controller in the role of passive monitor.

In conclusion, this study may indicate several sources leading to operational errors. The degree to which these results are generalizable to controllers involved in actively controlling air traffic needs to be investigated. Comparable data need to be examined during simulations of air traffic control to verify similar attention allocation to situation variables and responses to workload. At the very least, issues regarding SA demonstrated in this study may be indicative of problems that can be expected if air traffic control ever becomes highly automated, relegating the controller to a monitoring role. Alternate automation designs that keep the controller in the active decision making loop need to be explored to prevent such an outcome.

Acknowledgments

The authors would like to thank the controllers and management at the Atlanta Air Route Traffic Control Center for their participation and support of this study, in particular Al Guice for his coordination efforts. In addition, our appreciation is extended to Dennis Rester of the FAA for his programming efforts, and to Cindy Fox of the FAA and Ashok Sundaram at Texas Tech University for their assistance in data analysis.

References

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

Endsley, M. R. (1990a). A methodology for the objective measurement of situation awareness. In Situational Awareness in Aerospace Operations (AGARD-CP-478) (pp. 1/1 - 1/9). Neuilly Sur Seine, France: NATO - AGARD.

Endsley, M. R. (1990b). Predictive utility of an objective measure of situation awareness. In Proceedings of the Human Factors Society 34th Annual Meeting (pp. 41-45). Santa Monica, CA: Human Factors Society.

Endsley, M. R. (1993). Situation awareness and workload: Flip sides of the same coin. In R. S. Jensen & D. Neumeister (Eds.), Proceedings of the Seventh International Symposium on Aviation Psychology (pp. 906-911). Columbus, OH: Department of Aviation, The Ohio State University.

Endsley, M. R. (1995a). A taxonomy of situation awareness errors. In R. Fuller, N. Johnston, & N. McDonald (Eds.), Human Factors in Aviation Operations (pp. 287-292). Aldershot, England: Avebury Aviation, Ashgate Publishing Ltd.

Endsley, M. R. (1995b). Toward a theory of situation awareness. Human Factors, 37(1), 32-64.

Endsley, M. R., & Kaber, D.B. (in review) Level of Automation Effects on Performance, Situation Awareness and Workload in a Dynamic Task.

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. Human Factors, 37(2), 381-394.

Endsley, M. R., & Rodgers, M. D. (1994). Situation awareness information requirements for en route air traffic control (DOT/FAA/AM-94/27). Washington, D.C.: Federal Aviation Administration Office of Aviation Medicine.

Federal Aviation Administration. (1989). *Air Traffic Control* (DOT/FAA/Order 7110.65). Washington, DC: Federal Aviation Administration.

Fracker, M. L. (1990). Attention gradients in situation awareness. In Situational Awareness in Aerospace Operations (AGARD-CP-478) (Conference Proceedings #478) (pp. 6/1-6/10). Neuilly Sur Seine, France: NATO - AGARD.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), Human mental workload (pp. 139-183). Amsterdam: North-Holland.

Jones, D. G., & Endsley, M. R. (1996). Sources of situation awareness errors in aviation. Aviation, Space and Environmental Medicine, 67(6), 507-512.

Kinney, G. C., Spahn, M. J., & Amato, R. A. (1977). The human element in air traffic control: Observations and analyses of the performance of controllers and supervisors in providing ATC separation services (MTR-7655). McLean, VA: METREK Division, MITRE Corporation.

Mogford, R. H., & Tansley, B. W. (1991). The importance of the air traffic controller's mental model. Paper presented at the Human Factors Society of Canada Annual Meeting, Canada.

Rodgers, M. D., & Duke, D. A. (1993). SATORI: Situation assessment through the re-creation of incidents. Journal of Air Traffic Control, 35(4), 10-14.

Rodgers, M. D., & Nye, L. G. (1993). Factors associated with the severity of operational errors at Air Route Traffic Control Centers. In M. D. Rodgers (Ed.) An analysis of the operational error database for Air Route Traffic Control Centers (DOT/FAA/AM-93/22). Oklahoma City, OK: Human Factors Research Laboratory, Civil Aeromedical Institute, Federal Aviation Administration.

Schroeder, D. J. (1982). The loss of prescribed separation between aircraft: How does it occur? In Proceedings of the Behavioral Objectives in Aviation Automated Systems (pp. 257-269). Anaheim, CA: Society of Automotive Engineers (SAE).

Schroeder, D. J., & Nye, L. G. (1993). An examination of the workload conditions associated with operational errors/deviations at air route traffic control centers. In M. D. Rodgers (Ed.) An analysis of the operational error database for Air Route Traffic Control Centers (DOT/FAA/AM-93/22). Oklahoma City, OK: Human Factors Research Laboratory, Civil Aeromedical Institute, Federal Aviation Administration.

Sperandio, J.C. (1971) Variation of operator's strategies and regulating effects on workload. *Ergonomics* 14:571-577.

Stager, P., & Hameluck, D. (1990). Ergonomics in air traffic control. Ergonomics, 33(4), 493-499.

Wickens, C. D. (1992). Engineering Psychology and Human Performance (2nd ed.). New York: Harper Collins.

List of Figures