The Impact of Multi-layered Data-blocks on Controller Performance

M. L. Cummings, Chris G. Tsonis, and Andrew A. Rader

Abstract

As a consequence of the push to increase National Airspace System capacity, air traffic control displays will not only have to show the increasing number of aircraft, but also all the associated data such as airspeed and altitude. The representation of aircraft data and associated relational information, often superimposed on a map, leads to cluttered displays, which could negatively affect controller performance, especially as aircraft numbers increase. To investigate these issues further, an experiment was conducted that examined the effect of increasing data-block lines on controller performance in an aircraft vectoring task. Data-block design, the primary factor, varied in the number of lines displayed (2-5). In addition a data-block information priority factor was examined that addressed the frequency of information access across data-block lines. Results demonstrated that while task load, measured as an increasing number of planes under control, negatively influenced reaction times and task accuracy, the number of lines in a data block was not statistically significant. However there was a trend towards reduced performance when data-blocks exceeded more than three lines on a base layer. In addition, the data blocks that contained prioritized information across levels promoted faster reaction times, but at a cost of lower situation awareness. This research demonstrated that the design of data-blocks should consider the balance between reduction in data-block interaction time against the need to allow enough interaction time to build situation awareness.

INTRODUCTION

In human supervisory control tasks such as air traffic control (ATC), where humans supervise multiple entities depicted on a display, entity-specific information is often presented in the form of a data-block. Data-blocks can include alphanumerical and graphical information relevant to the target entity and the specific task. In the ATC domain, data-blocks include information pertinent to each aircraft displayed on a monitor. This information typically includes altitude, speed, heading, and flight and identification numbers, as well as possible additional information depending on the specific display type and ATC subtask. Two aircraft and accompanying data-blocks are shown in a representative ATC display in Figure 1.

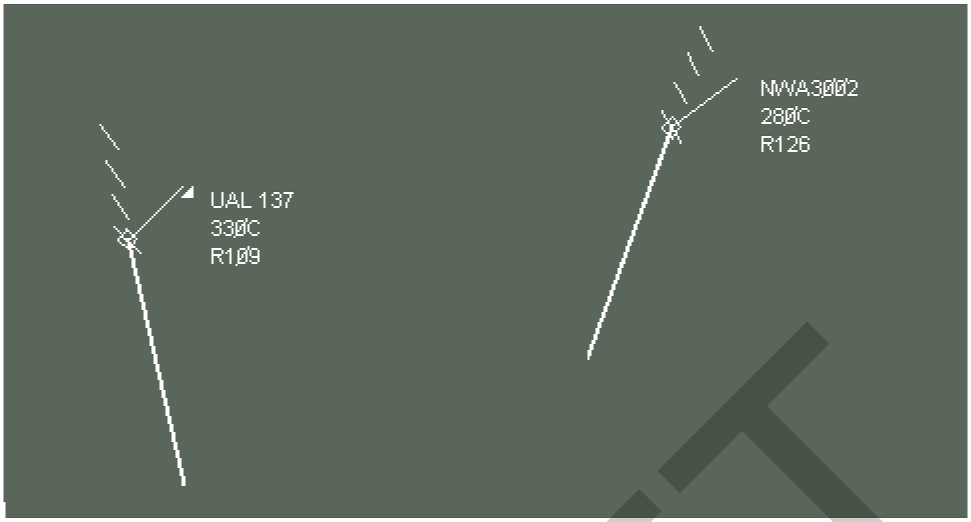

Figure 1. Representative en-route ATC display data-blocks containing three lines of information: flight number, altitude, and the computer identification number (CID), respectively.

It is essential that such a critical display element be carefully designed from a human factors perspective in order to ensure efficiency in the human-computer interaction process which should increase productivity and promote fewer errors, thus enhancing safety.

A cognitive task analysis, which included direct observation and interviews of controllers using en-route ATC interfaces at the FAA Technical Center in Atlantic City, NJ (detailed in (Tsonis, 2006)), yielded four important aircraft data-block design considerations:

1. Whether the information in data blocks should be embedded and layered such that portions could be accessed on demand, or presented in a single constant layer.
2. What types of actions in terms of clicks or cursor movements the human controller must perform in order to interact with the information contained in the data-blocks.
3. How to best graphically present symbols as well as past and future projections.
4. Where the data block should appear with respect to aircraft heading.

The experiment described in this article focused on the first design concern: the trade-offs, limits, and penalties of embedded, layered information in an ATC data-block.

MULTIPLE VERSUS SINGLE DATA-BLOCK LAYERS

The data-blocks represented in Figure 1 demonstrate the concept of a single layer data-block. There are three lines in the data-blocks, which are always displayed to the controller. We refer to this kind of design as a single layer data-block, in which all lines of a data-block are displayed, allowing for direct access to all information at all times. One alternate design strategy could be to layer data-block information in an on-demand, embedded format such that only when an action is intentionally communicated by the controller (e.g., through a button selection), additional data-block information is displayed. An example of this would be a data-block that always provided airspeed, altitude, and heading on the initial "base" layer, but when a control is activated (such as a mouse-click), additional data-block lines could give flight numbers, aircraft type, etc.

This kind of data-block dynamically expands to meet additional controller information requirements, and collapses back to its original base layer once no longer needed.

There are several advantages and disadvantages of single-layered data-blocks versus ones that are interactive, multi-layered, and expandable. A single layer data-block, in which all lines of a data-block are always shown on the display, allows for direct access to all information, which reduces operator reliance on working memory. While air traffic controllers can store and access large amounts of information in long term working memory (Ericsson & Kintsch, 1995), direct perception-interaction of display elements has been shown to improve reaction time in supervisory control tasks since operators employ more efficient perception processes rather than the cognitively demanding processes involved when relying on memory (Bennett, 1992; Rasmussen & Vicente, 1989).

However, as the number of aircraft on a screen increases, clutter inevitably increases to the point where it could increase search time. While the number of aircraft can contribute to display clutter, data-block design has also been shown to be a significant source of display clutter (Abbott et al., 1980). Clutter, those display elements that mask the visual perception of a stimulus in the presence of surrounding stimuli, can reduce controller performance through increased search time and reduced text readability, particularly in air traffic control displays (Xing, 2004).

Previous research has shown that the negative effects of display clutter can be reduced by reducing the amount of text in a display (Xing & Heeger, 2001). Highlighting the cost-benefit role of clutter, in one study the integration of display elements has been shown to reduce scan time in traffic display monitoring for student pilots, despite increased display clutter (Kroft & Wickens, 2002). In addition, map overlay clutter has been shown to impose more significant response times than additional aircraft clutter, suggesting that all display clutter is not equal (Wickens, Nunes, Alexander, & Steelman, 2005). In terms of interactive, layered display elements, previous research has demonstrated that presenting less information on a display does not outweigh the cost of retrieving information (Yeh & Wickens, 2001).

Thus, in this research effort, we hypothesize that data-blocks with lines that are only revealed on user action allows for collapsed data-blocks. This reduces display clutter and possibly improves controller performance, as measured by reaction time and errors. However, we recognize that this same display intervention meant to improve controller performance could also incur longer reaction times due to the activation of a control to access additional information.

In addition to the importance of the number of lines and an interactive component in a data-block, we also hypothesize that data-block information content should be linked to the information priority. Those pieces of information used most often and are the most critical (such as airspeed, altitude, and heading) should always be accessible through the base layer. Those pieces of information that are used more infrequently should be placed on “lower level” data-block lines, both to convey priority as well as to reduce search time. We propose that the data-block design configuration that yields the best performance (in terms of reaction times and accuracy of task) depends not only on the number and visibility of layers, but also on the data-block information priority (i.e. level of interaction required to find a required piece of information). This paper describes an experiment that was conducted in order to investigate the impact of number of and access to data-block lines on ATC controller performance, as well as the role of data-block information priority.

METHOD

In order to systematically analyze the previously discussed data-block design issues, an experiment was conducted. Using a MATLAB® ATC simulation program developed for this experiment, the primary task of controllers consisted of basic en-route aircraft separation and vectoring. The primary objectives were to determine if interactive multi-layered data-blocks provided any benefit over the single layered data-blocks, and if so, how many data-block lines should be accessible upon user request. Moreover, the impact of varying information priorities across the data-block was also a primary research question.

Participants

The subject controller population consisted of 22 U.S. Navy air traffic controllers (17 male, 5 female) with a mean age = 22.9 years, and standard deviation (SD) = 5.7 years. Seventeen of these controllers were trainees (4 female) and five were instructors (1 female). The novices had less than a year of experience and the instructors had more than five years experience. Navy air traffic controllers perform functions very similar to civilian air traffic controllers in that they qualify in air traffic control towers and radar air traffic control facilities. However, they also support offshore fleet area control and surveillance facilities, as well as distinctly military operations such as aircraft carrier air traffic control.

Apparatus

A human-in-the-loop ATC simulation interface known as the Reprogrammable Air Traffic Experimental Test bed (RATE) was programmed in MATLAB® to support this experiment. RATE was intended to emulate a Display System Replacement (DSR) type interface used in en-route ATC operations. An example of the RATE interface is shown in Figure 2.

The large diamond in the center of the display outlines the airspace sector for which a subject operator is formally responsible, with four surrounding “zones” representing adjoining airspace. The bottom left of the display contains a data-link interface where questions from a “supervisor” are answered (in this simulation, the supervisor was automated). This data-link interface simulates communication demands and acts as a secondary workload device by measuring the response time and accuracy of questions. The alert confirmation button, located on the bottom center, is pressed when operators first notice an alert that appears beside an aircraft data-block, in the form of a dot (as seen in Figure 2). The aircraft command panel, in the lower right portion of the screen, appears when an aircraft is selected, and is used to issue altitude, speed or heading commands. Selected aircraft were displayed as green on the display and unselected aircraft were yellow.

Training and testing were conducted on three identical Fujitsu tablet PC computers connected to external Dell 19 inch monitors (resolution 1024x768, 16 bit color), and user inputs were achieved through an external keyboard and mouse.

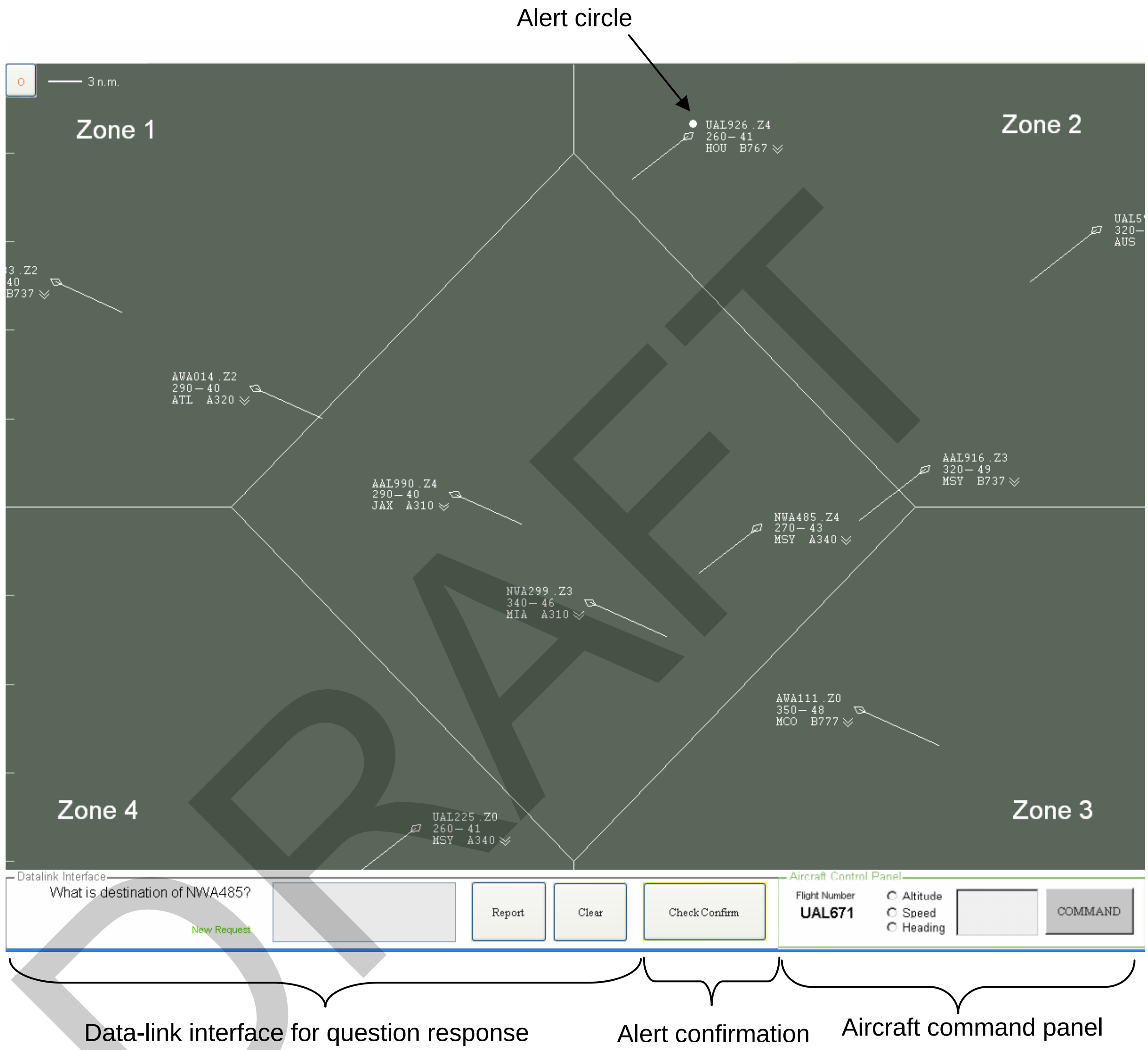

Figure 2. The RATE interface with collapsed data-blocks. Lower lines are accessed via the chevrons.

Procedure

For the primary task, controllers were required to observe incoming traffic, determine which of three possible egress zones each flight should be directed to, and then command the altitude and velocity required for each particular egress. Each egress zone had a specific altitude and airspeed that exiting aircraft must maintain prior to crossing the area threshold. The egress zone assignment for each flight was displayed on the first line of every data-block, (data block design will be discussed in more detail in the next section).

Controllers were also required to avoid conflicts, and had the option of bringing up three nautical mile diameter circles around each aircraft. Although controllers were formally in charge of the diamond sector, controllers could issue commands as soon as an aircraft appeared anywhere on screen. Thus, every aircraft appearing on screen had been effectively “handed over” to them, and controllers were informed of this in training. To command the aircraft, controllers first clicked on the aircraft symbol or heading vector, then issued commands in the control panel, and aircraft automatically maneuvered at rates typical of commercial aircraft. All aircraft were fully compliant with commands, although maneuver rates included small random variations.

As discussed previously, in every session a secondary workload task was included that consisted of answering a series of sixteen real-time evenly-spaced questions, a method similar to the SPAM technique for measuring situation awareness (Durso & Dattel, 2004; Durso et al., 1998). The order was predetermined and consistent between subjects, but randomized for each subject’s test session. The questions required the controller to respond with a single piece of information about a flight (for example, a flight’s altitude or final destination). These simple questions were prompted by an aural tone and green “New Request” text which appeared in the data-link window. Controllers scanned the spatial display to locate the relevant information, which required a search within either a single or multi-layered data-block to retrieve the requested information. They then typed their response in the data-link window. The response was confirmed by clicking the “Report” button.

An additional secondary task consisted of confirming a displayed alert, which was the yellow dot previously described. A total of four alerts appeared throughout each scenario. Alerts remained active until they were acknowledged, the aircraft left the screen, or the next alert appeared which randomly ranged from 20-280 seconds. Controllers were required to maintain an appropriate level of vigilance in order to quickly observe the appearance of an alert symbol. As the goal was to test spare visual attention, and there was no associated aural tone. Furthermore, the symbol was not particularly prominent, with the same color as the data-block text in order to increase the sensitivity of the associated detection metric.

The subject controllers completed a detailed tutorial presentation that discussed the nature of the experiment, and explained the task and the software. Controllers were then quizzed to ensure that they had memorized the meaning of each data-block information entry. If they made errors, they were corrected and re-tested prior to continuing with the experiment. All subject controllers then completed a single five-minute practice scenario before beginning the four nine-minute test sessions. After completing all sessions, they answered a brief demographic survey.

Experimental Design

Three independent variables were examined: number of data-block lines, information priority across each data-block, and task load level, which will be discussed in detail below.

Four different data-block design were tested (Figure 3), and while each contained identical information, the four types (or factor levels) varied by how much information was presented on the single base layer (as opposed to toggled from embedded layers using the chevrons). The first data-block had two lines (2 Line DB) and could be expanded to present the third, then the fourth and finally the fifth line, with a sequence of mouse clicks (one click per chevron on each line revealed the single line below). Similarly, the next data-block (3 Line DB) had three base layer lines (the two base layer lines from the 2 Line DB, plus the addition of the 3rd line containing destination and aircraft type, as seen in Figure 3). It could be expanded to the fourth and the fifth lines with two distinct mouse clicks. The 4 Line DB had four base layer lines (the three from the 3 Line DB, an additional line with Computer ID and Origin, and the fifth line could be revealed with a click. Finally, the 5 Line DB presented all five lines of information at all times, and thus represents the single layered data-block form. Only one data-block could be expanded at any given time. Thus for scenarios with expandable data-block levels, as one data-block was expanded, any other already expanded data-blocks would contract to their base number of lines.

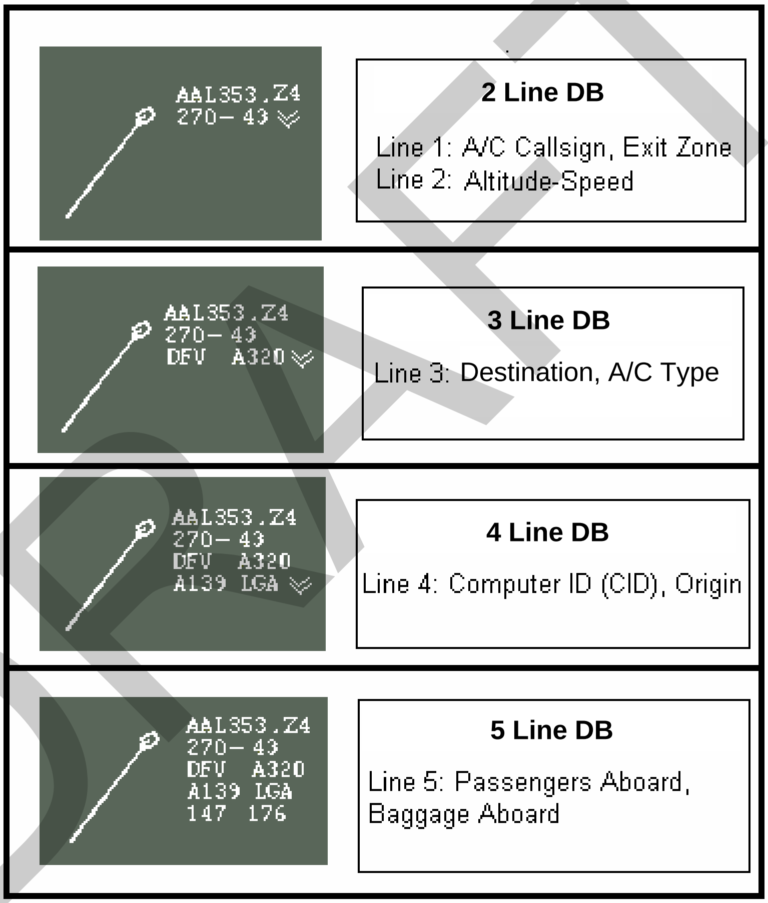

Figure 3. Four data-block types. Each data-block varied in the number of lines on the base layer. Controllers activated the next lower line by clicking on the chevron.

The information priority factor consisted of two levels, equal and prioritized, which examined the impact of information priority across the data-block lines via the secondary task. For the equal factor level, controllers were asked an equal number of questions (four) about the second through fifth lines in the data-blocks, for a total of sixteen questions. For the prioritized factor level, controllers were asked a higher number of questions about the second line (ten) and fewer questions about the other three data-block lines (six total). The intent for the prioritized level was to organize information such that more frequently needed information would always be on the top layer, and less important information on the lower levels.

Increasing task load, known to potentially negatively impact controller performance, was defined in this experiment by aircraft count, which has been cited throughout the literature as a leading source of controller workload (Hilburn, Bakker, Pekela, & Parasuraman, 1997; Kopardekar, 2003; Majumdar & Ochieng, 2002; Metzger & Parasuraman, 2001). In this experiment, the task load factor had two levels, low and high, and was determined by the number of aircraft a controller was responsible for at a given time. Under low task load conditions, approximately 5-6 aircraft were controlled, as opposed to 11-12 in the high task load condition. These levels were determined both because of previous research with similar levels (Hilburn, Jorna, Byrne, & Parasuraman, 1997), and with pilot testing to ensure an adequate, but not overly excessive increase in workload.

The experiment therefore consisted of a 4x2x2 mixed factorial design (data-block lines x information priority x task load). The data-block independent variable was administered between subjects, and the task load and data-block information priority independent variables were administered within subjects. Scenarios were counterbalanced to control for learning effects. Multiple dependent variables were gathered to determine the performance of controllers due to the influence of the independent variables on the various tasks, as outlined in the following discussion.

Performance of the primary en-route vectoring task was measured by egress fraction (*EF*), a weighted metric that was developed for the RATE interface to capture the participants' adherence to the pre-specified rules that each aircraft exit a specific egress zones at a certain altitude and velocity. Thus, EF consists of the sum of the number of correct egress locations (*e*), correct egress altitudes (*a*) and correct egress velocities (*v*), divided by three times the maximum number of possible correct egresses (*T*) within a scenario (to account for the sum of three variables), as shown in Equation 1. All three variables appeared on the first two lines of each data-block. A perfect score indicating compliance with all states exit rules equaled 1.

$$EF = \frac{e + a + v}{3T} \tag{1}$$

For the secondary task questions, correctness and response time were measured, since secondary task reaction times can be used as an indirect measure of spare mental capacity and thus give insight into mental workload (Wickens, 1995). In addition, the number of alert confirmations was recorded for each session, as was the number of six-second intervals in which two or more aircraft potentially conflicted, called the near miss measure. In both cases the number and lines of data-blocks, and hence display clutter, could degrade performance. In the case of alert confirmations, the yellow dot is more difficult to discriminate from the surrounding screen elements, and thus is an indication of how much display clutter contributed to workload. For the near-miss measure, trajectory extrapolation also is more difficult as more and larger data-blocks fill the display.

To summarize the relationships and hypotheses between the independent and dependent variables, display clutter is a primary concern with increasing numbers of data-block lines, compounded by an increasing number of aircraft. Increasing data-block base layer lines that are continuously exposed will increase display clutter, and thus controller performance is likely to degrade on clutter-sensitive tasks. We hypothesize that in terms of this experiment, the near miss, alert confirmation, and secondary task dependent variables are most likely influenced by the degree of display clutter (as a function of data-block lines and number of aircraft.)

While the number of data-block base layer lines and the number of aircraft are hypothesized to degrade controller performance as a function of increased display clutter due to quantity and lines of data-blocks, another important consideration is the organization and prioritization information across a data-block. Thus we hypothesize that the information prioritization factor will significantly affect the situation awareness of controllers as measured by the secondary task question response accuracy and near miss metrics.

The egress fraction dependent variable is a meta-performance metric that accounts for controller performance in terms of adherence to the pre-specified RATE rules (that an aircraft must exit the correct egress location at the correct altitude and correct velocity). Thus while it does not explicitly capture any effects due to display clutter or information prioritization, this aggregate measure demonstrates how well controllers were able to perform in light of the levels of independent variables over the entire experiment, and thus can be seen as an efficiency metric.

RESULTS

Comprehensive statistical analyses were conducted to determine the effect of the independent variables on each of the dependent variables, which is detailed below. For all tests, $\alpha = .05$, and there were no outliers in the data set. Means are reported for parametric data and medians are reported for non-parametric data.

Egress Fraction

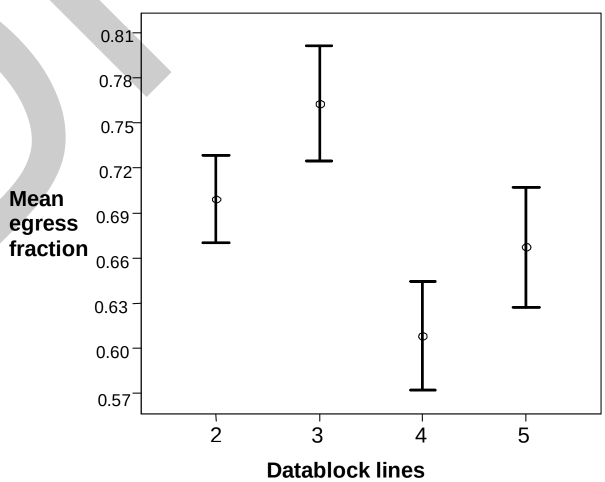

Figure 4. Mean operator egress fraction (with one standard error) by data-block type.

The egress fraction dependent measure was an objective aggregate assessment of the performance on the primary en-route ATC vectoring task, i.e., how well the controllers adhered to the pre-specified rules. A 4x2x2 (data-block lines x information priority x task load) mixed ANOVA was used to analyze the data, and all assumptions were met. The primary data-block line factor was found not to be a statistically significant factor ($F(3,18) = 1.78$, $p = .187$, η^2 (effect size) = .23). Although the data-block line factor was not statistically significant, a plot of egress fraction versus data-block line (Figure 4) shows a trend towards decreased performance for data-blocks with 4 or more base layer lines. While the omnibus test was not significant, the clustering of data deserved further investigation. The decrease was shown to be statistically significant using a contrast between the 2 and 3 line data-blocks as compared to the 4 and 5 line data-blocks ($p = .012$).

While the information priority factor was not significant ($F(1,18) = 2.141$, $p = .161$, $\eta^2 = .11$), the interaction between information priority and task load was statistically significant ($F(1,18) = 13.24$, $p = .002$, $\eta^2 = .42$). The equal priority factor level essentially added to overall workload by imposing a greater level of cognitive effort than the prioritized level under high task load conditions, which reduced overall performance by 12.5% from the low task load condition (as measured by the egress fraction). This can be seen by point *a* in Figure 5 (equal information priority, high task load) and is confirmed by a t-test comparison ($t(22) = -5.73$, $p < .001$) with point *b* (prioritized information at high task load.)

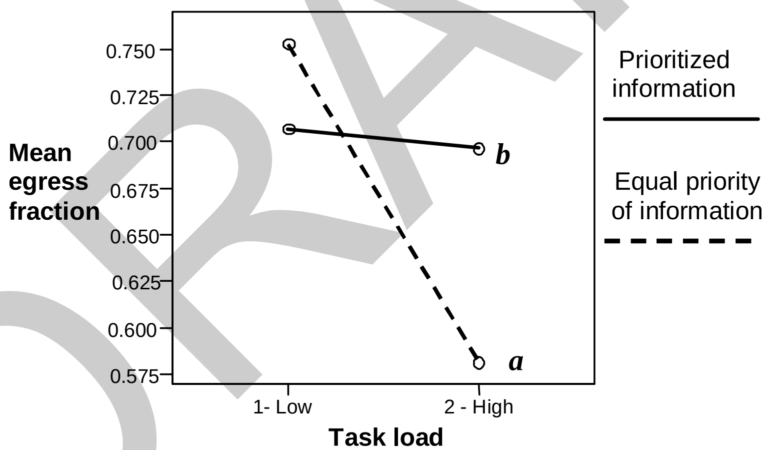

Figure 5. Interaction between task load and data-block information priority factor.

As expected, task load was significant ($F(1,18) = 7.96$, $p = .011$, $\eta^2 = .31$), such that as the number of aircraft increased, the vectoring task became more difficult and performance dropped significantly (mean low = .73, high = .64). This corroborates previous studies that have reported aircraft count as a primary source of controller workload and cognitive complexity (Hilburn, Bakker, Pekela, & Parasuraman, 1997; Kopardekar, 2003; Majumdar & Ochieng, 2002).

Secondary Task Reaction Time

In this experiment, spare mental capacity was measured through questions introduced through the data-link interface, which required controllers to seek out a particular piece of information. As with the egress fraction dependent variable, a mixed 4x2x2 ANOVA was carried out in order to uncover the significant main effects and interactions. The three independent variables were once again the between-subjects factor of data-block lines and the two within-subjects factors of information priority and task load. As before, the assumptions required for ANOVA were met.

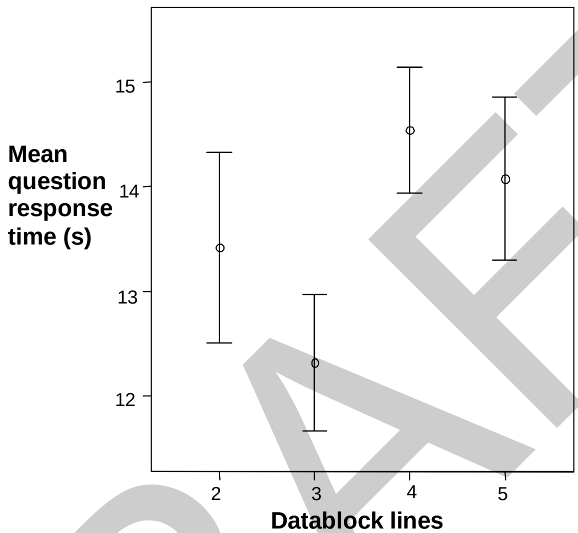

Figure 6. Mean question response time (with one standard deviation) by data-block.

The data-block line factor was not statistically significant in affecting the mean question response times ($F(3,18) = .988$, $p = .421$, $\eta^2 = .14$). The data-block means for question response time are plotted in Figure 6. While the omnibus test was not significant, Figure 6 demonstrates an overall trend towards increased reaction time based on increasing the number of base layer lines in the data-block, particularly between the data-blocks with 2 embedded lines (3 Line DB) and 1 embedded line (4 Line DB), which was a similar trend seen in the egress fraction performance score. When the reaction times were grouped for the data-blocks with 2 and 3 lines and compared with the 4 and 5 line data-blocks (also combined), the difference is marginally significant ($p = .061$). While not as significant as the trend seen for egress fraction, there is a marginally significant difference between data-block with three or less base layer lines as compared to those with four or more.

The information priority factor was significant ($F(1,18) = 5.40$, $p = .032$ $\eta^2 = .23$), with equal priority resulting in overall longer response times than the prioritized condition (equal mean = 14.2s, prioritized mean = 12.5s). Thus, those data-blocks with less frequently needed information placed on the lower priority levels contributed to faster response times as compared to those data-blocks with an equal priority across all data-block lines. Figure 7 shows how information priority affected response times for each of the four data-block lines. Interestingly, the data-blocks with the smallest base layer of 2 lines and the largest base layer of 5 lines gained the most benefit from prioritized information in terms of secondary task response time. In the case of the 2 line data-block, this result is likely due to the fact that for the prioritized information condition, the correct answer could be found without uncovering the embedded layers 63% of the time. Thus controllers generally did not have to click through the embedded layers. In the 5 line data-block, all information was present all the time, but since the required information could more often be found in the first 2 lines, this priority scheme helped them to more quickly find the required information.

As with the egress fraction performance dependent variable, task load was statistically significant in determining question response time ($F(1,18) = 6.30$, $p = .022$, $\eta^2 = .26$). Those people with more aircraft expectedly took longer to answer the questions, and thus had less spare mental capacity (mean for 6 aircraft = 12.8s, mean for 12 aircraft = 14.4s). The primary source of this response time increase is most likely the increased search time penalty across the aircraft, as well as longer delays in attending to the secondary tasks. There were no significant interactions for any of the independent variables.

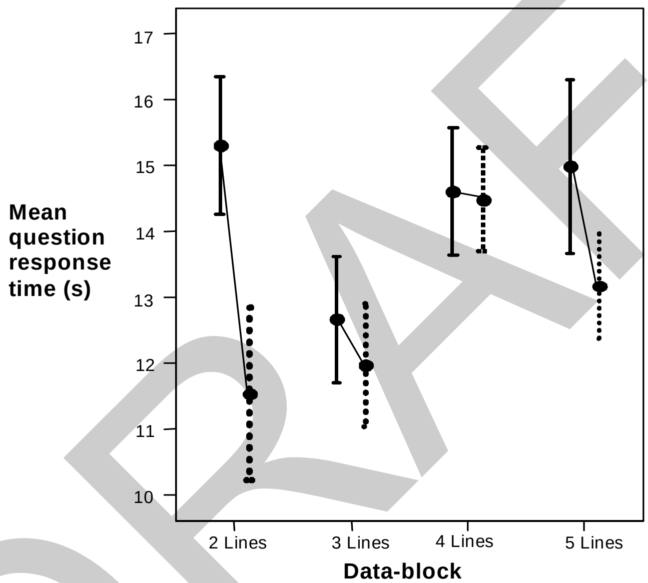

Figure 7. Mean question response time for equal (solid) and prioritized (dashed) information (with one standard error) by data-block type.

Question Response Accuracy

Question response accuracy, which measured through an on-line probe method, was measured as the number of correct responses within each scenario. As this was not interval data, a non-parametric Kruskal-Wallis test was conducted. It was expected that the data-block line factor would not affect the accuracy of the question responses because as long as operators could get to the information, it was hypothesized that they would be able to respond correctly, even though they may take longer doing so. There was generally no speed-accuracy trade-off as those people who answered the most questions correctly generally had the lowest response times.

The chi-square statistic of the data-block ranks was not significant ($\chi^2(3, 22) = .41$, $p = .94$). Thus as expected, the question response accuracy did not depend on the type of data-block. To investigate whether the information priority factor was significant in determining accuracy, a Wilcoxon Signed-Rank test was performed, which also showed non-significance (Z = -.51, p = .61). However, the task load factor, also examined with a Wilcoxon Signed-Rank test, showed significance (Z = -4.41, p = .000). In low task load scenarios, controllers had on average 13.8 correct responses per scenario, as compared with 11.4 correct responses per scenario under high task load (out of a maximum of 16 possible correct responses).

Thus the question accuracy results demonstrated that accuracy was only dependent on task load, and not on the number of data-block lines from which the information is extracted, or on the information priority. As long as the information in a data-block could be found, there appears to be no significant difference in the accuracy of responses.

Alert Confirmation

As discussed previously, controllers were instructed to acknowledge alerts by pressing a button on the interface (Figure 2). The actual metric used was whether or not the alert was confirmed. As expected, the higher task load (more planes to manage) reduced the ability of controllers to detect alerts on the display (Wilcoxon Z = -2.95, p = .009). However, a Kruskal-Wallis test ($\chi^2(3, 88) = 3.34$, p = .34) showed that the number of data-block lines did not significantly affect the ability of controllers to detect a change on the display.

The effect of data-block information priority on alert confirmation accuracy was analyzed using a Wilcoxon Signed-Rank test and was found to be significant (Z = -2.22, p = .027). Controllers using data-blocks with evenly distributed information across the data-block lines missed more alert confirmations than those controllers with prioritized information (mean equal percentage missed = 16%, mean prioritization percentage missed = 10%). This difference is likely due to the fact that, on average, with the evenly distributed information, controllers had to search longer in order to access a required piece of information, regardless of whether any information was embedded. This extra attention and resultant time meant that they were less likely to detect a visual alert on the display.

Thus the number of aircraft was the only significant contributor to missed alert confirmations in terms of added display clutter. In addition, while not an explicit display clutter issue but also an important design consideration, the organization of the information and the frequency of access also added to diminished controller performance.

Near Misses

The final dependent metric was a count of the number of near misses, which was measured when any two aircraft were separated by less than three nautical miles in 6 second intervals. This was chosen as a metric as it is representative of the severity of a traffic violation. In general, the more time aircraft are in conflict, the greater the chance of a collision, which is why this metric has a temporal component. In addition, this metric was important to consider since all the previous metrics focused on controller performance per aircraft or per task, but did not take into account the interactions between aircraft which is arguably the most important metric since an air traffic controller is responsible for the safety of all aircraft interacting in space and time. Thus the near miss metric provides an overall human-system performance measurement.

A Kruskal-Wallis test showed that the data-block line factor was not significant in affecting the near miss count ($\chi^2(3, 92) = .11$, p = .99). Regardless of how many lines were present on the base layer, controllers performed no differently. However, as expected, the effect of task load on near misses was significant (*Z* = -5.77, p = .000). Higher task load resulted in more near misses (low task load median = 6, high task load median = 15). While this number of near misses seems high, the RATE interface is an abstraction of the ATC task, and thus near misses in RATE are not indicative of actual operational errors. Moreover, due to Navy operational constraints, training, testing, and debriefing time was limited, thus no controller was an expert in the RATE interface, but all received equal training time.

Analysis of the information priority factor yielded intriguing results in that the Wilcoxon Signed-Rank test was also significant (*Z* = -5.43, p = .000). Prioritized information resulted in a significantly larger number of near miss counts as compared to the equal information priority condition (prioritized median = 11.5, equal median = 6.5). These results are unexpected in that for all other performance measures, prioritized information appeared to reduce controller workload (in comparison to equal information priority), such that controllers were able to attend to both primary and secondary tasks more quickly and accurately. However, in the case of the near miss metric, which took into account not only the current state of each aircraft, but its relationship to all other nearby aircraft, those controllers with equal information priority clearly had fewer near misses.

We hypothesize that the source of the seeming anomaly is the concept of situation awareness. Situation awareness (SA) is generally defined as 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. SA has three distinct levels, which are: 1) the perception of the elements in the environment, 2) the comprehension of the current situation, and 3) the projection of future status (Endsley, 1995). We propose that while equal information priority across the data-block lines required longer search times (as evidenced by the significance of the results for the secondary workload reaction times), this longer interaction forced controllers to better understand both the current and likely future states of each aircraft.

In the equal information priority condition, controllers spent approximately 75% of their time interacting with the bottom 3 lines of the data-block as compared to only 38% for those controllers with the prioritized information. This lower level interaction caused controllers to have more comprehensive and accurate mental models of the overall air traffic picture, which allowed them to better understand the relationships between aircraft. While the prioritized information allowed controllers to more quickly execute their tasks, the lack of in-depth interaction prevented them from building more detailed mental models. This lack of a more complete mental model was not manifested in the other performance metrics since they only examined individual performance of egressing aircraft or secondary communications tasks. It was only when the near miss metric was examined that this disadvantage of prioritized information across each data-block was revealed.

In order to further investigate this hypothesis, a Friedman test was conducted which compared the interactions between low and high task load, and equal and prioritized information, resulting in four distinct conditions: equal priority/low task load, equal priority/high task load, prioritized/low task load, and prioritized/high task load. The overall Friedman test was significant ($\chi^2(3, 22) = 56.45$, $p < .001$), and the medians of the four conditions are depicted in Figure 8. Conditions 1 and 3 represent low task load conditions (with equal and prioritized information, respectively), and not surprisingly, controllers in these conditions experienced fewer near misses than those in the high task load conditions. However, those controllers experiencing high task load with the prioritized information across the data-block (Figure 8) experienced, on average, 66% *more* near misses per test scenario than all other conditions. This lends credence to our hypothesis that controllers in this condition had lower situation awareness than in all other conditions, as we know they were spending less time interacting with the data-block (from the secondary task results), but were also under the maximum task load level which also allowed them less time to actually perceive, comprehend, and project aircraft states correctly.

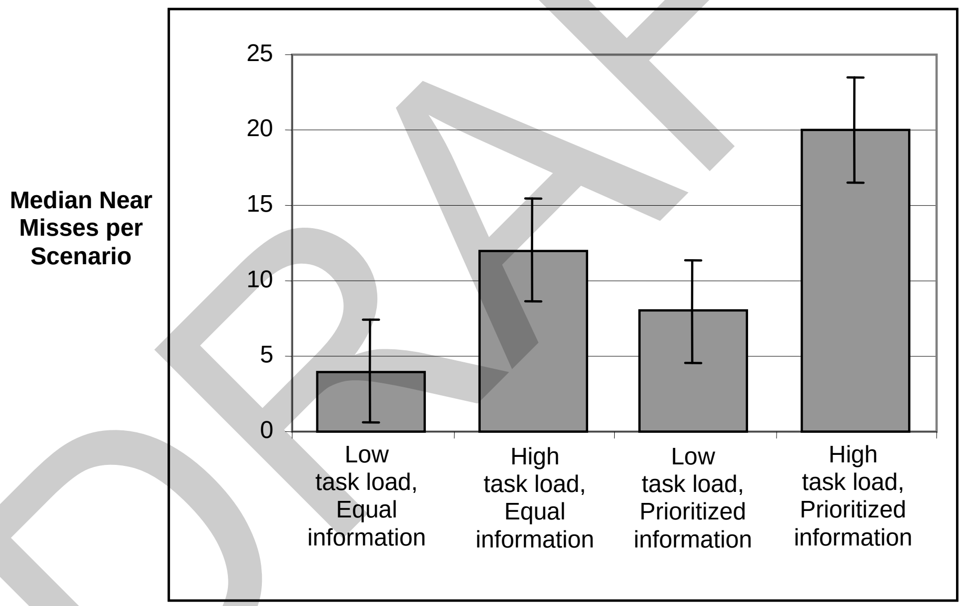

Figure 8. Medians of the near miss measure for information priority and task load

DISCUSSION

We originally hypothesized that it was possible that interactive data-block designs with embedded lines of lesser important information could improve operator performance due to reduction of display clutter. However, we also recognized that the actual act of clicking through an embedded data-block in order to reveal hidden layers would have some display overhead associated with it. According to Fitts' law, for the RATE interface, on average, it takes approximately 1.2 seconds to move the mouse cursor from the aircraft interface panel to the data-block and initiate the first data-block expansion (Card, English,

& Burr, 1978; Fitts, 1954). Additional clicks take around 200 ms each, if the user knows how many expansions are necessary before looking at any results. However, if the user must search each layer that was revealed to find the desired piece of information, the access time could increase significantly, making an interaction data-block counter-productive to efficient performance.

Counter to the concern that interactive data-blocks could cause increased reaction times and possible resultant degraded performance, the results in this experiment showed unequivocally across a number of dependent variables (performance, secondary task reaction time and accuracy, alert confirmations, and near miss counts), controllers' performance was relatively robust to the data-block design. In this ATC en-route simulation, regardless of whether a controller interacted with data-blocks with 2-4 base layer lines, or a 5 line data-block, performance was statistically no different.

The only caveat to this finding is that for both the egress fraction score and secondary task reaction time, there appeared to be a performance decrease trend between the data-blocks with 2/3 base layer lines and those with 4/5. Repeating this experiment with a larger population may shed more light on this possible implication, but based on these results, controllers generally appear to be able to effectively adjust to whatever data-block design they are given.

While controllers' performance was robust in the face of multiple data-block designs for egress fraction and response times, performance across all five dependent variables was not with increasing task load, as represented by an increasing number of aircraft to control. For the five different dependent variables, controllers performed statistically significantly worse when presented the higher (11-12) number of aircraft as opposed to the lower (5-6). These results provide additional evidence for the theory that aircraft number is a significant contributor to air traffic complexity and controller workload (Kopardekar, 2003; Majumdar & Ochieng, 2002).

The information priority independent variable (with equal and prioritized levels) demonstrated mixed results across the five dependent variables. For the egress fraction performance score, controllers with equal information priority across the data-blocks in the high task load scenario performed significantly worse than those controllers with prioritized information. In addition, for the secondary task response time and alert confirmation dependent variables, controllers with information distributed equally across data-block lines did not perform as well as those with the prioritized information. However, there was no difference between the two information priority levels for question accuracy.

For the near miss metric, which is arguably the most critical metric because of its safety implications, the results showed that the prioritized information led to significantly more near-misses than the equal priority condition across the data-blocks. Given the relatively consistent poorer performance of those controllers with equal information priority across the other dependent variables, this result was unexpected. This result was thought to occur because the longer interaction times caused by equal information priority (confirmed by the secondary task reaction time metric) contributed to higher situation awareness and fewer near-misses than for those controllers who did not spend as much time interacting with the data-block under the prioritized information condition. This effect was not seen in the meta-performance metric, egress fraction, most likely because even though egress fraction was considered an aggregate performance measure, it did not explicitly account for relational situation awareness (i.e., the relationships between aircraft), and only captured how well controllers handled individual aircraft.

This mix of results for the information priority factor highlights the cost-benefit analysis nature of interface design. If data-blocks are designed such that the information that is needed most is placed on the upper levels, controllers do not have to spend as much time interacting with each individual data-block. However, this reduction in interaction time is a double-edged sword in that some period of time is needed to build situation awareness, and if controllers do not spend enough time perceiving and comprehending current information, they have difficulty making future projections that are critical for such tasks as detecting possible future collisions. Thus a critical lesson learned in this research study is that faster is not always better when interacting with a data-block, and that some period of interaction time is needed to build situation awareness.

Lastly, these results should be interpreted in light of efficiency and safety, critical issues for ATC operations. The egress fraction dependent variable is representative of system efficiency (i.e., the percentage of aircraft that were correctly processed through the system), and the near miss metric is indicative of system safety. While the number of data-block lines and interactivity appeared not to induce any performance detriment, the number of aircraft and hence number of data-blocks did appear to negatively affect controllers for both efficiency and safety measures. However, these results also demonstrate that there is a cost-benefit tradeoff between efficiency and safety, which is a common problem across numerous supervisory control systems. The equal versus prioritized data-block design for information location and access did not significantly impact system efficiency (as measured by egress fraction), but there was a tradeoff for system safety in the two different designs. These results are not an endorsement for a specific data-block design strategy, but highlight that a design decision can affect efficiency and safety with sometimes conflicting results.

CONCLUSION

In a cognitive task analysis conducted at the FAA Technical Center in Atlantic City, NJ, one important aircraft data-block design issue raised was whether the information in data-blocks should be embedded and layered such that it is accessed on demand, as opposed to a single layer. One possible benefit of such an interactive, multiple-layer design would be reduced display clutter. However, one possible drawback to such a design would be increased interaction times such that any positive benefit could be negated in terms of controller performance.

To address this design conundrum, an experiment was conducted to compare four data-block designs, in addition to varying task load and data-block information priority factors. The results demonstrated that data-blocks with interactive, expanding layers did not have an overall effect on any of the dependent measures. Despite the lack of statistical significance, there was a trend towards reduced performance between data-blocks with 2 and 3 embedded lines, and data-blocks with 1 and no embedded lines, with the latter conditions producing the lowest performance.

Results for the objective task load factor, represented through increasing the number of aircraft under control, demonstrated that increasing task load reduced performance on the primary ATC en-route vectoring task, reduced secondary task question response accuracy, increased response times, impaired the ability to confirm alerts, and resulted in a higher number of near misses. These results provide additional evidence that aircraft numbers are a primary source of operator workload.

Lastly, data-block information priority was significant across multiple dependent variables, and while equally prioritized information across the data-block increased workload, it also led to fewer near misses. Thus, as with all complex system design cost-benefit analyses, the design of new data-blocks should consider the balance between a reduction in individual data-block interaction time to allow for more time to interact with more data-blocks, against the need to allow *enough* interaction time to build situation awareness for supervisory control of multiple aircraft. This cost-benefit approach to providing controllers with just the right amount of information will be critical during the transition from the current system to the Next Generation Air Traffic Control System (NextGen).

ACKNOWLEDGEMENTS

This research was sponsored by the Civil Aerospace Medical Institute. The authors would like to thank the U.S. Navy, and in particular the ATC Training Center in Pensacola, Florida for providing subject controllers. Additional support was provided by Chris McQuin and Angela Ho who assisted in test bed development and conducting the experiments. In addition, the anonymous reviewers were very helpful in framing and streamlining this research effort.

REFERENCES

Abbott, T. S., Moen, G. C., Person, L. H., Jr., Keyser, G. L., Jr., Yenni, K. R., & Garren, J. F., Jr. (1980). *Flight investigation of cockpit-displayed traffic information utilizing coded symbology in an advanced operational environment* (No. NASA-TP-1684). Hampton, VA: NASA Langley Research Center.

Bennett, K. B. (1992). Graphical Displays: Implications for Divided Attention, Focused Attention, and Problem Solving. *Human Factors, 34*(5), 513-533.

Card, S. K., English, W. K., & Burr, B. J. (1978). Evaluation of the mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. *Ergonomics, 21*, 601-613.

Durso, F. T., & Dattel, A. R. (2004). SPAM: the real-time assessment of SA. In S. Banbury & S. Tremblay (Eds.), *A Cognitive Approach to Situation Awareness: Theory and Application* (pp. 137–154): Aldershot: Ashgate.

Durso, F. T., Hackworth, C. A., Truitt, T. R., Crutchfield, J., Nikolic, D., & Manning, C. A. (1998). Situation Awareness as a Predictor of Performance for En Route Air Traffic Controllers. *Air Traffic Control Quarterly, 6*(1), 1-20.

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors, 37*(1), 32 - 64.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*, 211-245.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology, 47*, 381-391.

Hilburn, B., Jorna, P. G., Byrne, E. A., & Parasuraman, R. (1997). The Effect of Adaptive Air Traffic Control (ATC) Decision Aiding on Controller Mental Workload. In *Human-automation Interaction: Research and Practice* (pp. 84-91). Mahwah, NJ: Lawrence Erlbaum.

Hilburn, B. G., Bakker, M. W. P., Pekela, W. D., & Parasuraman, R. (1997). *The effect of free flight on air traffic controller mental workload, monitoring and system performance.* Paper presented at the 10th International CEAS Conference on Free Flight, Amsterdam.

Kopardekar, P. (2003). *Measurement and Prediction of Dynamic Density.* Paper presented at the ATM 2003 R&D Seminar, Budapest.

Kroft, P., & Wickens, C. D. (2002). Displaying multi-domain graphical database information: An evaluation of scanning, clutter, display size, and user activity *Information Design Journal, 11*(1), 44-52.

Majumdar, A., & Ochieng, W. Y. (2002). *The factors affecting air traffic controller workload: a multivariate analysis based upon simulation modeling of controller workload.* Paper presented at the 81st Annual Meeting of the Transportation Research Board, Washington DC.

Metzger, U., & Parasuraman, R. (2001). The role of the air traffic controller in future air traffic management: An empirical study of active control versus passive monitoring. *Human Factors, 43*(4), 519-528.

Rasmussen, J., & Vicente, K. J. (1989). Coping With Human Errors Through System Design: Implications for Ecological Interface Design. *International Journal of Man-Machine Studies, 31*, 517-534.

Tsonis, C. G. (2006). *An Analysis of Information Complexity in Air Traffic Control Human Machine Interaction* (No. HAL2006-05). Cambridge, MA: MIT Humans and Automation Laboratory.

Wickens, C. D. (1995). *Situation awareness: Impact of automation and display technology.* Paper presented at the Situation Awareness: Limitations and Enhancements in the Aviation Environment, Neuilly Sur Seine, France.

Wickens, C. D., Nunes, A., Alexander, A. L., & Steelman, K. (2005). *3D Navigation and Integrated Hazard Display in Advanced Avionics: Workload, Performance, and Situation Awareness Display Clutter Issues in SVS Design*. Savoy, IL: Aviation Human Factors Division, Institute of Aviation.

Xing, J. (2004). *Measures of information complexity and the implications for automation design*. Oklahoma, OK: Civil Aerospace Medical Institute.

Xing, J., & Heeger, D. (2001). Quantification of contrast-dependent center-surround interaction. *Vision Research, 41*, 571-583.

Yeh, M., & Wickens, C. D. (2001). Attentional filtering in the design of electronic map displays: A comparison of color-coding, intensity coding, and decluttering techniques. *Human Factors, 43*(4), 543-562.

BIOGRAPHIES

Mary (Missy) L. Cummings received her Ph.D. in Systems Engineering from the University of Virginia in 2003. She is currently an assistant professor in the MIT Aeronautics & Astronautics Department and is the director of the Humans and Automation Laboratory.

Chris G. Tsonis received his S.M. in Aeronautics and Astronautics from MIT in 2006. He is currently a trader assistant at DRW Trading Group and was previously a graduate research assistant in the Humans and Automation Lab.

Andrew A. Rader is a doctoral candidate in the Aeronautics & Astronautics Engineering department at MIT. He earned his Masters of Aerospace Engineering at Carleton University in Ottawa, Canada in 2005.