

Partitioning Complexity in Air Traffic Management Tasks

M.L. Cummings

Chris G. Tsonis

Massachusetts Institute of Technology

Cambridge, MA

Abstract

Cognitive complexity is a term that appears frequently in air traffic control (ATC) research literature, yet there is little principled investigation of the potential sources of cognitive complexity. Three distinctly different sources of cognitive complexity are proposed which are environmental, organizational, and display. Two experiments were conducted to explore whether or not these proposed components of complexity could be effectively partitioned, measured, and compared. The findings demonstrate that sources of complexity can be decomposed and measured and furthermore, the use of color in displays, a display design intervention meant to reduce environmental complexity, can actually contribute to it.

Introduction

Addressing the difference between environmental and innate human complexity (often referred to as cognitive complexity), Herb Simon describes an ant's path as it navigates across a beach. The ant eventually reaches its destination, but because the ant must constantly adapt its course as a result of obstacles, the path seems irregular, laborious, and inefficient. Simon points out that while the ant's path seems complex, the ant's behavior is relatively simple as compared to the complexity of the environment. Simon proposes the following hypothesis as a result, "Human beings, viewed as behaving systems, are quite simple. The apparent complexity of our behavior over time is largely a reflection of the complexity of the environment in which we find ourselves. (Simon, 1981, p. 53)"

This distinction between innate, or cognitive complexity and environmental complexity is especially relevant in light of the considerable research conducted in air traffic controller cognitive complexity. Several studies have investigated air traffic control (ATC) information complexity issues (see Hilburn, 2004; Majumdar & Ochieng, 2002 for a review). In this literature, several common "cognitive complexity" factors have emerged to include traffic density, traffic mix, aircraft speeds, and transitioning aircraft. However, in light of Simon's ant parable, these factors really represent environmental complexity factors that influence cognitive complexity, as compared to other factors such as separation standards which are intentional organizational factors externally imposed on controllers. This is an important distinction because as can be seen in Figure 1, there are several sources of complexity that can affect an individual's cognitive complexity level.

Figure 1 illustrates the decomposition of "complexity" as it applies to complex sociotechnical systems in which operators interact with their environment, mediated by procedures and technology. Cognitive complexity is actually the lowest form of complexity in this notional model which can be affected by some or all of the preceding levels, although not necessarily in a linear fashion. A true, objective state of complexity exists in the world (environmental complexity), as perceived by one or more humans (cognitive complexity). In ATC, factors that affect environmental complexity include number of aircraft, weather, congestion, etc., which essentially are the same elements identified in ATC cognitive complexity literature. As illustrated in Figure 1, operators' perceived complexity can be mitigated through two strategies, organizational guidance and/or displays. Organizations like the FAA and DoD will institute policies and procedures to mitigate environmental complexity such that operations are both safe and meet specified goals. For example, separation standards exist to mitigate complexity for controllers

(and promote safety); however, when airspace becomes saturated, the need to maintain these organization-driven constraints causes situation complexity, and thus workload, to increase.

In human supervisory control domains, displays are nominally designed to further reduce complexity by representing the environment so that a correct mental model can be formed and suitable interactions can take place (Woods, 1991). Thus, displays should reduce complexity and hence workload through transforming high-workload cognitive tasks such as mental computations into lower workload tasks through direct perception, i.e. visually (Miller, 2000). However, in complex and dynamic sociotechnical domains such as ATC, it is not always clear whether displays and organizational interventions actually alleviate or contribute to the problem of complexity.

While previous research has been critical in addressing controller workload as a function of cognitive complexity due to environmental and to a lesser degree, organizational factors, significantly less attention has been paid to the role of displays in either adding to or mitigating complexity in the ATC environment. As illustrated in Figure 1, only examining environmental factors as sources of complexity is necessarily limited and does not capture the actual complexity perceived by operators. In an attempt to partition and quantify the influence of environmental, organizational, and display complexity factors on cognitive complexity, two experiments were conducted to determine if these components could be independently measured, and if so, determine possible implications for decision support design.

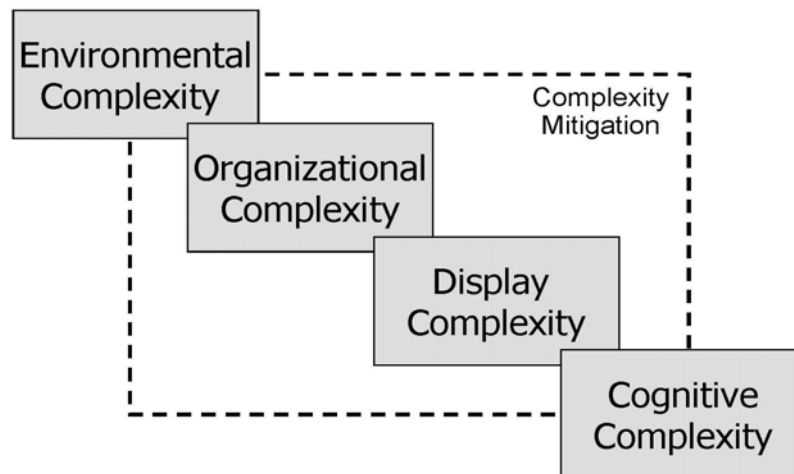


Figure 1: Human Supervisory Control Complexity Chain

Experiment 1

In addition to traffic density and related factors, it has also been hypothesized that the underlying airspace structure is a critical complexity factor (Histon et al., 2002). In theory, airspace structure provides the basis for mental abstractions which allows controllers to reduce complexity and maintain situation awareness. Histon et al., (2002) proposed that these mental abstractions, known as structured-based abstractions, can be generalized to standard flows (reminiscent of Pawlak's (1996) "streams"), groupings, and critical points. Providing air traffic controllers with these interventions, either explicitly through airspace design or implicitly through policy, in theory should help controllers improve performance and reduce overall complexity, as well as reduce perceived workload.

In a study investigating judgment and complexity, Kirwan et al., (2001) determined that airspace sector design was second only to traffic volume in terms of contributing to cognitive complexity. In terms of the model in Figure 1, airspace sector design straddles both the organizational and display complexity categories. Designed by humans to mitigate environmental complexity, airspace structure is an organizational policy. However, airspace structure contains significant visual components represented through displays, thus it is an environmental complexity intervention both from an organizational and display perspective.

Including interventions in airspace sector design such as critical points (points through which aircraft must pass) and designated standard flows (such as jet ways) can increase order and improve predictability, and thus lower cognitive complexity. However, it is also possible that when uncertainty levels increase, usually as a function of dynamic environmental factors such as changes in weather and available airspace, these same airspace structures could actually add to complexity. Airspace structure and procedures mitigate complexity in what are termed "nominal" situations, but when an "off-nominal" condition occurs, such as an emergency or unexpected weather phenomena, the resultant increasing uncertainty causes complexity to grow (Athenes et al., 2002). In order to address whether or not airspace structure as an organizational/display factor is a significant contributor of cognitive complexity, an experiment was conducted which is described below.

Experiment 1 Method

Apparatus, participants, and procedure. To objectively investigate the impact of complexity sources described in Figure 1 on controller performance, a human-in-the-loop simulation test bed was programmed in MATLAB® (Figure 2). Since the subject pool consisted primarily of college students, it was necessary to devise a simplified and abstract task that represented fundamental elements of enroute ATC control. Subjects as controllers were only

concerned with providing heading commands to the aircraft, and aircraft velocities were held constant. Twenty egress points were located in the periphery of the airspace (Figure 2), and each incoming aircraft was assigned one of these egress points. The primary goal was to direct the incoming aircraft to the assigned egress, and when an aircraft exited correctly, subjects received points. Additional waypoints were located throughout the airspace sector by which subjects could collect more points by directing aircraft through these points.

A path connecting the waypoints with the highest values was the optimal path which encouraged controllers to fly aircraft on a specified route and not directly to their exits. To discourage controllers from flying aircraft through all waypoints (which is unrealistic in actual enroute ATC control), scores were penalized based on an aircraft's total time transiting the airspace sector. A final component of the score was a penalty for flying through a no-fly-zone, which represented restricted airspace. Maximization of total score was the subjects' goal, which could be achieved by flying the optimal path that was not necessarily the most direct path. Subjects' total scores were displayed in real-time. To provide a strong incentive for subjects to perform well and attempt to maximize their total score, a \$50 prize was offered to the subject who attained the highest score.

Two sources of complexity were investigated using this test bed, dynamic airspace sectors (environmental complexity factor) and airspace structure (display complexity factor). To investigate display complexity as a function of airspace structure, in certain scenarios, subjects were given structure through the display of the actual

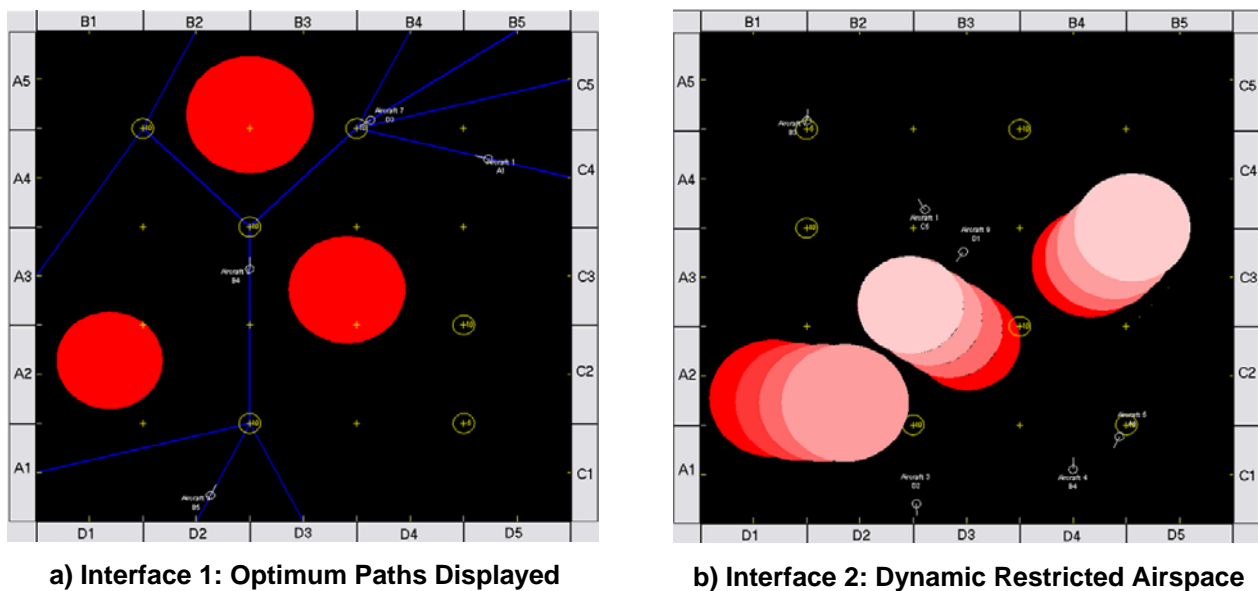


Figure 2: Experiment 1 Test Bed

optimum paths (those that maximized the score as a function of waypoints and time) (Figure 2a). As demonstrated in Figure 2b, some subjects were not given this same visual path structure, although the waypoints were the same, so the optimal paths were the same, but just not clearly linked with lines.

The second potential source of complexity was the use of dynamic as opposed to static airspace sectors, termed no-fly zones. In the dynamic case, the no-fly-zones moved at rates of approximately two-fifths the aircraft velocity and were meant to represent areas of airspace that can change rapidly such as military operating areas (MOAs) or areas of severe weather conditions. In these cases, air traffic controllers must develop new routes for affected aircraft under time pressure. Similarly, in this test bed, as airspace changed, subjects were required to reroute aircraft but still attempt to maximize their overall score. For those subjects with the displayed paths between optimal waypoints, the displayed lines were still the optimum, but only in cases where they were not obstructed, and thus subjects had to replan for a new best case.

The primary purpose of the experiment was to investigate whether or not visual structure in an airspace sector, in combination with changes in the external airspace environment, added to or mitigated perceived complexity as measured through performance. In theory, under nominal conditions (static airspace), subjects performance should be higher with the displayed lines than no lines, since the optimal path is clearly laid out. However, in the off-nominal case when airspace is dynamic, the visually compelling paths may no longer be optimal and could provide a distraction to subjects who must mentally generate new possible paths. In this case, the more unstructured airspace without the lines may be easier for subjects to use to generate contingency plans.

Training and testing were conducted using a Dell personal computer with a 19-inch color monitor with a screen area of 1024x768 pixels, 16-bit high color resolution, and a Pentium 4 processor. During testing, all user responses were recorded in separate files specific to each subject and scenario. A Visual Basic script scored and compiled the data into a single spreadsheet file for the subsequent statistical analysis. After signing required consent forms, subjects completed a tutorial that discussed the nature of the experiment, explained the context and use of the interface, and gave them the opportunity to understand the scoring mechanism. Subjects completed four practice scenarios that exposed them to every combination of independent variables. They then began the randomly ordered four test sessions, which lasted until all aircraft had exited the airspace (approximately 6-7 minutes).

Design of experiment. As discussed previously, two independent variables were investigated: 1) Structure of airspace (either with lines or no lines), and 2) The environment state (either static or dynamic no-fly-zones.) A

single dependent variable of total performance score was used. As described previously, the score was a linear weighted function of correct aircraft egress actions, bonus waypoints, incurred penalties for no-fly-zone violations, and total time transiting in sector. In the case of egress scores, aircraft did not receive points for exiting through the wrong egress and the maximal egress score was obtained by exiting near the center of the egress, with a minimal score awarded towards the edge. The four experimental scenarios were ninety degree rotations of each other, and the rotation was automated in a scenario generation tool (Tsonis et al., 2005). The statistical model used was a 2x2 fully crossed ANOVA and the four scenarios were randomly presented to a total of 20 subjects.

Experiment 1 Results and Discussion

The ANOVA linear model revealed that the environmental factor of dynamic airspace was highly significant ($F(1,74) = 54.55$, $p < .001$, all $\alpha < .05$); whereas the displayed airspace structure and the environment*displayed structure interaction were not significant. Figure 3 depicts the average performance scores across all four conditions. It can be seen on inspection that the performance scores were clearly higher in the static environment scenario as opposed to the dynamic environment phase. Whether or not subjects had less or more airspace structure did not significantly affect their scores. These results demonstrate that for this representative ATC task, the environment was a significant contributor to complexity and performance, causing lower scores. Thus as environmental complexity increased through changing airspace, workload, and thus cognitive complexity increased, with no apparent effect, either positive or negative, from increasing airspace structure. This is not to say aircraft structure was not an important contributor, just that its effects could not be measured due to overwhelming effect of changing airspace.

In terms of the model in Figure 1, this experiment demonstrated for this representative ATC task, the main component of complexity associated with controller workload was environment, and not organizational or display-related. Dynamically changing airspace structure was far more influential than the design of the airspace itself. Thus while sector design may be a contributing factor to air traffic controller performance, environmental complexity factors such as thunderstorms and special use airspace that intermittently become available can be significantly larger contributors to individual cognitive complexity.

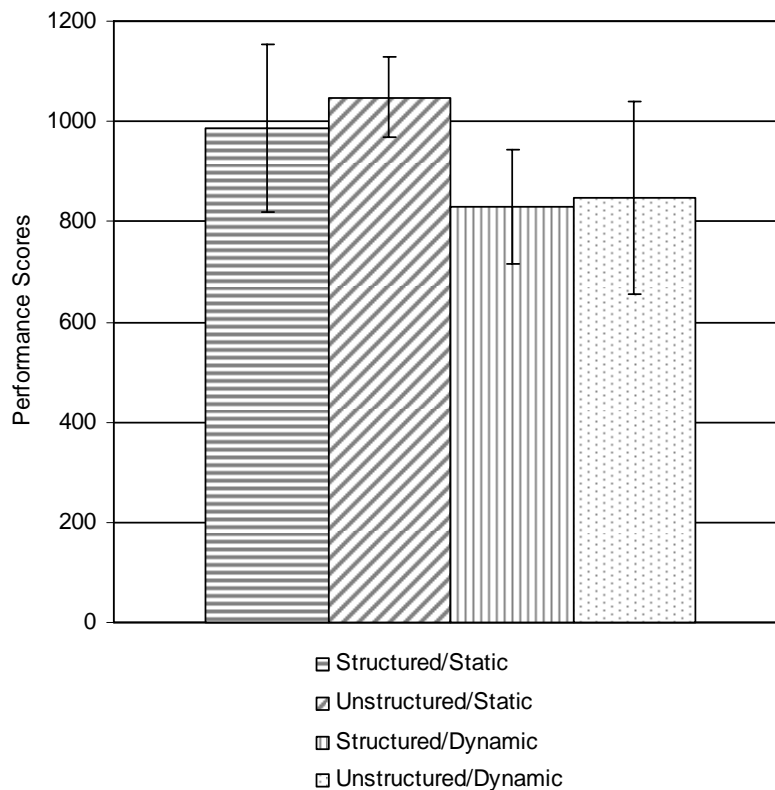


Figure 3: Experiment 1 Results

Furthermore, these results provide quantitative support for previous subjective assessments of controllers that active special use airspace increases complexity and would benefit from some display intervention (Ahlstrom et al., 2001). In light of the results reported here, it is likely that special use airspace (SUA), an organizationally driven constraint, can increase workload for controllers not because of the actual structure of the airspace, but because the status can change. SUA exists to mitigate environmental complexity in that, for example, it separates military and commercial traffic. However, when SUAs cycle between active and inactive, especially in short time periods, situation complexity can increase.

Experiment 2

While experiment 1 addressed complexity concerns due to airspace design, a second experiment was conducted to examine more specifically complexity due to display design in terms of a generic timeline. Timelines are currently in use for the Center-TRACON Automation System (CTAS) Traffic Management Advisor (TMA) in order to provide for coordination of arrival traffic between Center and TRACON Airspace (Hansman & Davison, 2000). In addition, while not yet operational, the Surface Management System integrates both tactical and strategic timelines; however, results of exploratory and operational trials have been mixed (Atkins et al., 2004; Brinton et al., 2002; Spencer et al., 2002). For example, in one recent operational trial, local and ground controllers rejected timelines while traffic managers felt they were very useful (Atkins et al., 2004). However, in all trials, evaluation of the timelines has been subjective based on controllers' opinions, so no performance data has yet been published that objectively demonstrates whether or not timeline displays, or features of these timelines, actually help controllers in their tasks.

One specific concern in the design of timelines is the use of color. With advanced display technologies that allow designers to use hundreds of color conventions with no added system cost, there has been a recent increase in the amount of color used in ATC displays. However, there has been little consideration of how much color might be too much from an information processing perspective. Xing and Schroeder (2004) have documented the extensive and inconsistent color use in ATC displays. The FAA has issued no formal requirements for the use of color in ATC displays, and consequently, manufacturers of ATC technologies are free to decide their own color schemes. Indeed even ATC facilities and individual users are allowed to determine their own color preferences in some tools, such as the Enhanced Traffic Management System (ETMS) and the Information Display System (IDS). While guidelines exist for the general use of color in ATC display technologies (Cardosi & Hannon, 1999; Reynolds, 1994), they generally address optimal perception conditions, and not how the use of color could improve or degrade task performance

Color in ATC displays is typically used for three primary reasons: 1) To draw attention, 2) To identify information so that search time is minimized, and 3) To organize information through segmentation such as airspace boundaries (Xing & Schroeder, 2004). Research has shown that color is superior to achromatic visual attributes (e.g., luminance, shapes, and text) in search and organization tasks primarily because color-coded information can be

processed more quickly (Christ, 1975). For example, the use of red in displays to convey warning information allows operators such as pilots and ATC personnel to quickly assess an urgent problem state.

While color can provide improvements in search and organization tasks, previous research has shown that while subjects believed that color improved their performance, color did not improve target detection or identification (Jeffrey & Beck, 1972). In addition, the use of color can cause cognitive tunneling or “inattentive blindness (Simons, 2000)”, in which operators may miss other important information on a display because they fixate on the more salient color change. Another possible problem with the current use of color in ATC displays is that many displays use the same color to identify different types of information, such as using blue to indicate both an airline type as well as a particular phase of flight. As the number of display colors increases, with possible multiple meanings, both the perceptual and cognitive load of controllers is increased, subsequently adding to cognitive complexity.

To address whether or not display complexity could be captured due to the use of increasing color categories in a generic timeline, an experiment was conducted using the framework in Figure 1 for partitioning sources of complexity.

Experiment 2 Method

Apparatus, participants, and procedure. To objectively investigate the impact of environmental and display complexity factors associated with a timeline, a human-in-the-loop simulation test bed was developed in MATLAB® (Figure 4). Since the subject pool consisted of college students, a simplified task was needed that contained realistic decision support tools, yet did not require years of expertise to operate. Thus the subjects’ task was that of a low-level surface manager of incoming and outgoing traffic, responsible for assisting a supervisor in managing personnel for baggage handling, ground personnel, and galley service.

The radar screen in Figure 4 represents incoming and outgoing traffic in the terminal control area. The circle in the middle represents the airport area. The timeline contains two essential elements, much like what is used in actual ATC timelines, incoming traffic (left) and outbound traffic (right). The incoming side of the timeline represents the time until the expected aircraft gate arrival. The outgoing side of the timeline represents the time that an aircraft begins loading passengers and baggage at the gate until it becomes airborne. Each aircraft label contains the flight number, number of passengers, number of baggage items, assigned gate number, velocity (when airborne) and altitude (when airborne).

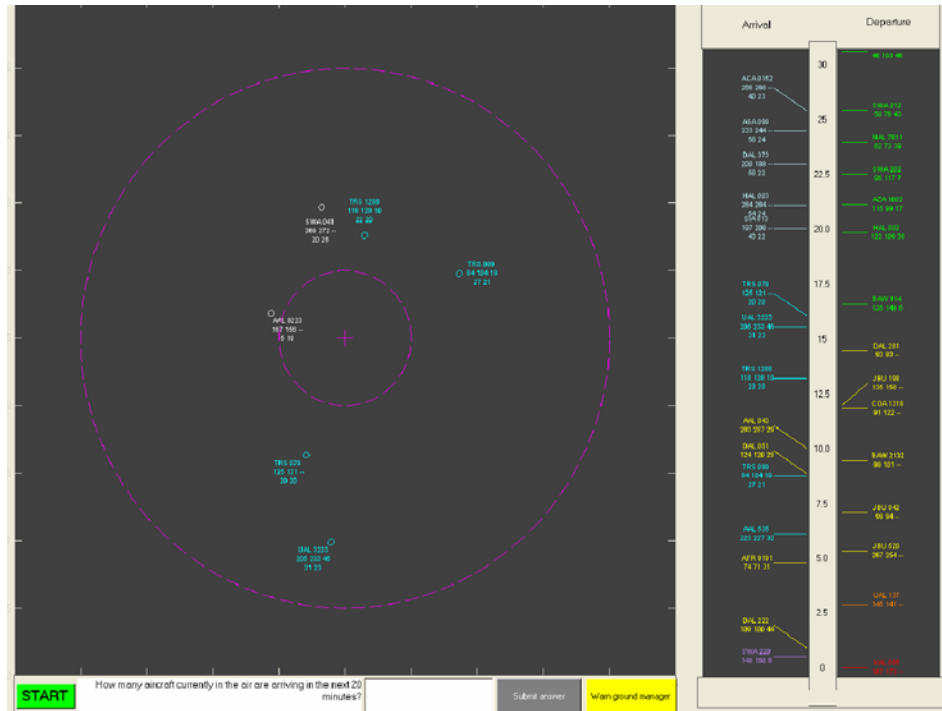


Figure 4: Experiment 2 Interface

After signing consent forms, subjects completed a tutorial and three practice scenarios. They then began the randomly assigned 18 test sessions, which lasted approximately 3.5 minutes each. Subjects were required to monitor both the radar display and timeline, and answer questions from their superior through datalink (text message) communication. Subjects were also required to notify their superior when aircraft of a particular airline entered the outermost circle.

Design of experiment. Two independent variables were used to represent environmental complexity, number of aircraft (10, 20, 30) and arrival sequence (sequential vs. non-sequential.) Aircraft number has been cited throughout the literature as a leading source of complexity (Hilburn, 2004), and it has been previously demonstrated in an ATC study that number of aircraft is the primary driver of operator performance (Kopardekar, 2003). Arrival sequence was included as an additional environmental complexity factor because in the context of a visual linear timeline display, the sequence of entries could affect a controller's ability to effectively search for information. Those aircraft that maintain their relative positions in the timeline are easier to track than those aircraft that appear to "jump" on the timeline, e.g., those aircraft that are put into holding patterns, disrupting the expected flow of traffic. This could be problematic for situation awareness as a timeline does not convey the uncertainty and dynamic

changes of non-sequential aircraft arrivals. Thus, we examined whether or not sequential arrivals versus non-sequential arrivals as displayed on the timeline influenced operator performance.

The third and final factor investigated was the number of color categories used to represent information about incoming and outgoing aircraft (3, 6, 9). As the levels of color increased, the additional colors added more detailed information for the general category so the meaning of the colors would not be a confound. In the 3 color condition, aircraft were generally grouped into arrival, taxiing, and gate status categories. Thus, as more colors were added, more refinements were made to the specific categories. For example, in the 3 color condition, a single color (white) represented arrivals whereas in the 9 color condition, arrivals were further subdivided into enroute, both in and outside of terminal area, as well as on final approach. As in actual timeline displays, colors were sometimes duplicated across arrival and departing aircraft (e.g., yellow on the left side of the timeline meant taxiing in while yellow on the right timeline meant taxiing out; thus, yellow always meant taxiing.)

The primary dependent variable was answer accuracy, measured through questions in an embedded datalink tool, a strategy which has been useful in developing workload metrics in military command and control (Cummings & Guerlain, 2004). The questions introduced through the datalink window fell into 2 categories: 1) Search questions which relied on primarily perceptual information processing, e.g., subjects had to locate a single piece of information such as the gate call sign of a certain aircraft, and 2) Problem solving questions, in which subjects were required to calculate or derive information from multiple sources, e.g., the number of aircraft at their gates. In each of the 18 scenarios, subjects were asked six questions which were generally evenly split between problem solving and search categories. In addition to answering questions, subjects were required to notify the supervisor when they first noted that a flight of a certain carrier entered the outermost radial circle on the radar display. Failures to recognize this situation resulted in an error of omission, which is the other dependent variable. The statistical model used was a 3x3x2 fully crossed ANOVA, and the 18 scenarios were randomly presented to a total of 29 subjects.

Table 1: Color Categories

Flight Status	Number of Colors		
	3	6	9
Scheduled to arrive	White		
En route		Blue	
En route outside airspace			Blue
En route inside airspace			Cyan
On final approach		Orange	Orange
Taxiing in	Yellow	Yellow	Yellow
On runway			Red
Ready to dock with gate			Purple
At gate	Green	Green	Green
Ready for pushback		Pink	Pink
Taxiing out	Yellow	Yellow	Yellow
In final queue (holding short)		Orange	Orange
	White	White	White

Experiment 2 Results

Accuracy results. For this experiment, subjects were classified as generally ‘accurate’ if they answered at least four out of six test questions correctly in a particular scenario. An overall comparison of subject performance for the search and problem solving question accuracy in each of the 18 test sessions reveals intriguing results. For the search questions, additional color categories increased inaccuracy, most predominantly from three to six color categories, however, the general subject population still answered accurately overall (Figure 5, Pearson Chi-Square showed marginal significance, $p = .069$).

For the problem solving questions, the increase in color categories actually improved the answer accuracy, although there was no difference between the use of six or nine color categories (Figure 6, Pearson Chi-Square test

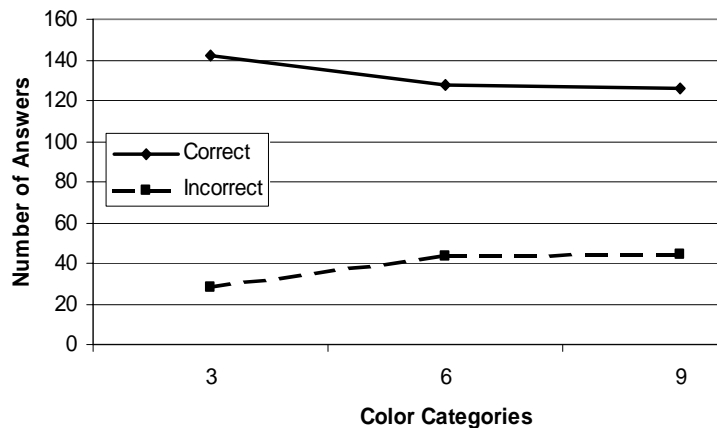


Figure 5: Accuracy for Search Questions

showed $p < .001$ between 3 colors versus 6 and 9). In general for both question types, wrong answers increased with increasing aircraft levels. Measures of association using Cramer’s V are reflected in Table 2. For the search questions, arrival patterns and number of aircraft affected correct answers while increasing aircraft and color categories were both moderately associated with wrong answers.

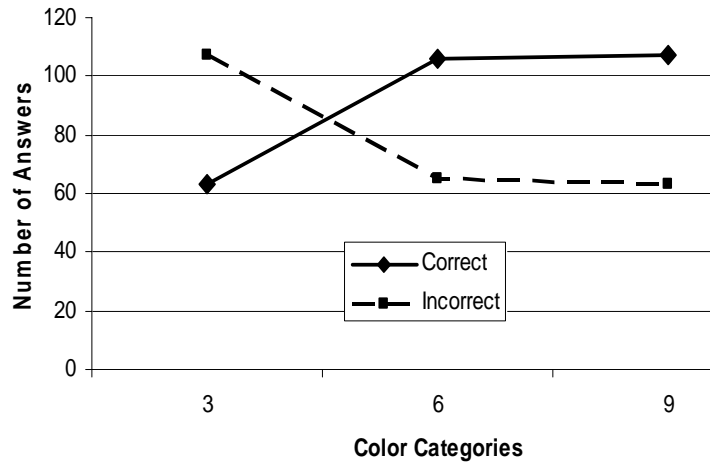


Figure 6: Accuracy for Problem Solving Questions

Table 2: Accuracy Measure of Association

	Search Accuracy	Problem Solving Accuracy
Color Categories	.102 ($p=.069$)	.240 ($p < .001$)
Aircraft	.269 ($p < .001$)	.275 ($p < .001$)
Arrival Pattern	.327 ($p < .001$)	Not significant

Errors of omission. A known negative consequence of human supervisory control high workload is an increase in human error, as illustrated in the previous section. The incorrect answers given by the subjects to datalink queries represent errors of commission. However, errors of omission are also prevalent and of particular concern in the air traffic control community. To this end, the influence of the three display metrics on omission error occurrences was investigated. Subjects were told that whenever an aircraft marked as BAW (British Airways) entered the area between the two dashed lines on the radar display in Figure 4, they were to notify their supervisor by clicking a single button on the lower right corner, marker “Warn ground manager”. For all independent variables, correct

notifications exceeded errors of omission (unlike accuracy for problem solving questions); however, there were important trends in error rates due to the display components.

Figure 7 represents the overall correct notifications as compared to the errors of omission for the color categories, which was significant (Wilcoxon, $p = .009$). A Mann-Whitney test between the errors for the six and nine color categories also was significant ($p = .001$). Figure 8 demonstrates a similar omission error increasing trend

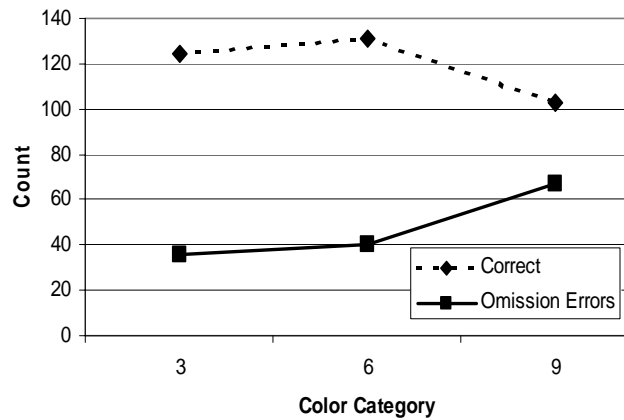


Figure 7: Omission Errors: Colors

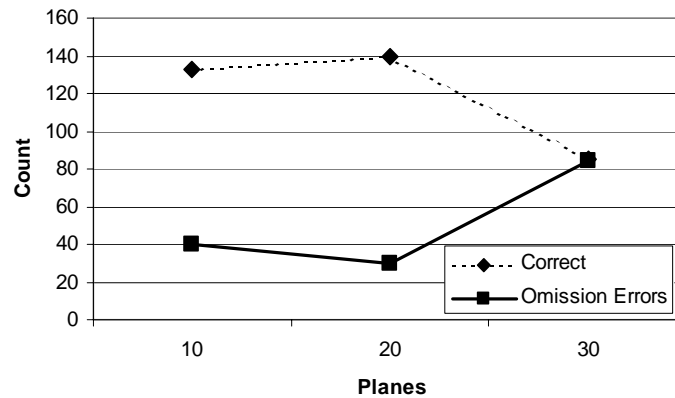


Figure 8: Omission Errors: Aircraft

between 20 and 30 aircraft (Mann-Whitney, $p < .001$). The arrival pattern effect is seen in Figure 9. Subjects' errors of omission increased when the arrival patterns were non-sequential, and this difference between sequential and non-sequential was significant (Mann Whitney, $p = .036$).

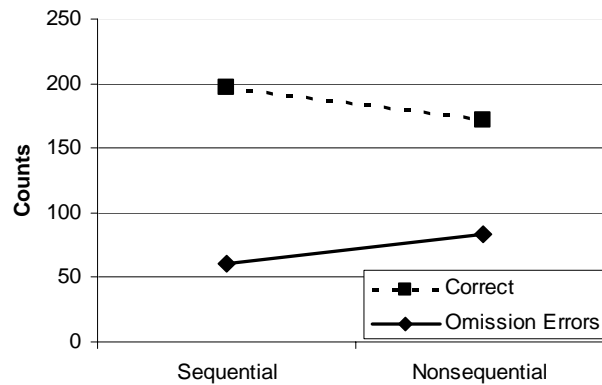


Figure 9: Omission Errors: Arrival Pattern

Given that all three independent variables showed significance through non-parametric testing, further investigation was warranted to determine the magnitude of any significant relationship. Association testing using the Kendall tau-b statistic between errors and the independent variables revealed significant associations for all three variables, as shown in Table 3. As color categories increased from 3 to 9 and as aircraft increased from 10 to 30, errors of omission increased, as indicated by the positive association. Subjects who experienced sequential arrival patterns experienced lower errors of omission as indicated by the negative sign. All associations were moderate, but the aircraft factor, a complexity factor driven by the external environment, contributed only slightly more to error rates than the color category, a display design intervention typically included to mitigate complexity.

Table 3: Measures of Omission Error Association

Factor	Association	Significance
Color Category	.330	p < .001
Aircraft	.355	p < .001
Arrival Sequence	-.293	p = .001

Experiment 2 Discussion

In this second experiment that investigated whether or not complexity from different sources could be effectively measured and compared, we identified two sources of environmental complexity (arrival sequence and number of aircraft), and complexity from displays (color coding). Using a generic timeline and ground traffic management task, we demonstrated that these sources of complexity can be partitioned, measured for increasing

levels, and then compared to one another to determine which factor(s) contribute the most to cognitive complexity as measured through performance. While the following sections will contain more details about the analysis of environmental and display complexity sources, in general we determined that increasing aircraft number is a primary driver of complexity. However, increasing use of color categories contributed only slightly less to poor performance.

Environmental complexity. The aircraft number factor was included because it represents a primary source of environmental complexity for controllers. In general, answer accuracy decreased and errors of omission increased as the number of aircraft increased. Increasing numbers of entities is a known source of cognitive complexity (Edmonds, 1999), and aircraft count has been identified as the primary environmental driver of ATC cognitive complexity (Kopardekar, 2003), so the result is expected.

Traffic flow is another commonly cited source of ATC complexity (Majumdar & Ochieng, 2002), and while arrival sequence was not a significant factor for answer accuracy, it did significantly affect subjects' errors of omission, but to a lesser degree than the aircraft and color categories. Arrival sequence did not directly cause errors of omission, but non-sequential patterns increased search time, and thus increased overall workload which diverted attention from the monitoring task. The negative influence of non-sequential arrivals on operator performance illustrates the consequences of change blindness, which occurs when operators fail to detect change in a visual display. This blindness typically occurs if a change occurs in concert with another visual event, such as animation or an eye blink, and has been shown to occur even for large changes in a visual display (Rensink, O'Regan, & Clark, 1997).

The arrival sequence result suggests that while arrival sequencing may be an environmental complexity factor, there is a display component associated with it, in the sense that there was no visual indication to users when aircraft unexpectedly changed sequence, which is typical of most timelines. Had some kind of situation awareness recovery tool been implemented for aircraft jumps in the timeline, this effect may have been mitigated.

Display complexity. For search questions that relied primarily on perceptual cueing, color coding was generally helpful, however, there was a slight drop in accuracy as color categories increased. For the problem solving answer accuracies, color coding significantly improved correct answers from three to six colors, but reached a plateau at six so nine color categories provided no additional benefit. However, one area where increasing color categories appeared to be detrimental was in the analysis of errors of omission, which significantly increased from

six to nine color categories. This is an important finding because color coding is a design intervention meant to mitigate complexity, not add to it.

In this generic timeline task, increasing color usage beyond six color categories appeared to promote inattentive blindness or cognitive tunneling. While increasing color categories do not “cause” people to forget actions, they do require that controllers spend more time in search and mapping tasks, and thus take time away from other tasks and increase the likelihood that a subsequent task will be forgotten. In addition, while number of aircraft exhibited a slightly stronger association, increasing color categories was a significant contributor to errors of omission more so than arrival rate. This result that suggests color usage should not exceed six color categories is in keeping with both previous ATC research (Cardosi et al., 1999), as well as general display design principles (Wickens & Hollands, 2000).

Conclusion

Complexity, as it applies to operators and the air traffic control and management environments or any other human supervisory control domain, cannot be simply categorized as “cognitive complexity.” We proposed there are different components of complexity (Figure 1) and demonstrated through two experiments that these sources can be partitioned and the impact on human performance can be measured. Specifically, we partitioned two sources of complexity, environmental and display, in a basic enroute ATC control task and a simplified ground traffic management task. In both experiments, the environmental complexity factors were predominant; however, in the case of the timeline task, the use of color in the displays added to, instead of reducing environmental complexity.

In the enroute experiment, the environmental complexity source of changing airspace was far more significant in influencing overall controller performance as compared to the complexity display factor caused by visual routes. These results support air traffic controllers’ subjective opinions that special use airspace (SUA) is a source of complexity, and that more work is needed for better display representation. Because of the temporal and cyclic nature of SUA, possible design interventions could include some kind of timeline display for SUA scheduling, as well as intelligent decision support agents that can predict in advance when airspace could become available or deactivated and possible courses of action.

In the ground traffic management task, investigation of environmental and display sources of complexity revealed that increasing numbers of aircraft affected subject performance only slightly more than color categories;

however, traffic arrival sequence patterns were not as strongly associated with degraded performance. While aircraft count is a known primary contributor to cognitive complexity, this research demonstrates that the increasing use of color is a likely source as well. Color has been shown to aid operators in search and organization tasks; however, this experiment generated evidence that beyond six color categories, performance was not aided, and overall task performance degraded through an increase in errors of omission. These research results highlight the importance of understanding sources of complexity in human supervisory control domains. Displays are the primary visual mechanisms through which controllers build mental models; however, as demonstrated in this study, even a simple addition of color categorization can significantly influence controller cognitive complexity.

The proposed components of environmental, organizational, and display complexity may not contribute in a linear, additive, or consistent manner to either cognitive complexity or performance and will likely be task and technology dependent. As new technologies and procedures are introduced in air traffic control and management tasks, if the different sources of complexity are not well understood, it is likely that the interventions themselves could add to overall complexity and not reduce it, as evidenced by our color category results. Thus it is important to be able to partition the sources of complexity and measure their impact so that areas for possible design interventions can be identified. Once the potential design interventions are identified and implemented, further testing can compare the new results to the earlier complexity tests in order to determine if the perceived complexity on the part of the operator has actually been mitigated.

Acknowledgements

This research was sponsored by the Civil Aerospace Medical Institute. We would like to thank Daniel Crestian Cunha, a visiting student from Instituto Tecnológico de Aeronáutica (ITA), as well as Kathleen Voelbel for their assistance.

References

- Ahlstrom, U., Rubinstein, J., Siegel, S., Mogford, R. H., & Manning, C. (2001). *Display Concepts For En Route Air Traffic Control* (DOT/FAA/CT-TN01/06). Washington DC: U.S. Department of Transportation.
- Athenes, S., Averty, P., Puechmorel, S., Delahaye, D., & Collet, C. (2002). *ATC Complexity and Controller Workload: Trying to Bridge the Gap*. Paper presented at the HCI-Aero.
- Atkins, S., Jung, Y., Brinton, C. R., Stell, L., Carniol, T., & Rogowski, S. (2004). *Surface Management System Field Trial Results*. Paper presented at the AIAA 4th Aviation Technology, Integration, and Operations (ATIO) Forum, Chicago, IL.

- Brinton, C. R., Carniol, T., Butler, T., Duley, J., & Hitt, J. M. (2002). *Air Traffic Management System Development and Integration (ATMSDI) Acquisition CTO-05 – Surface Management System (SMS) CTOD 5.11-1 – Phase 1 Report*. Moffett Field, CA: Raytheon.
- Cardosi, K., & Hannon, D. (1999). *Guidelines for the use of color in ATC displays* (DOT/FAA/AR-99/52). Washington DC: Office of Aviation Research.
- Christ, R. E. (1975). Review and analysis of color coding research for visual displays. *Human Factors*, 7, 542-570.
- Cummings, M. L., & Guerlain, S. (2004). *Using a Chat Interface as an Embedded Secondary Tasking Tool*. Paper presented at the 2nd Annual Human Performance, Situation Awareness, and Automation Conference, Daytona Beach.
- Edmonds, B. (1999). What is Complexity? - The philosophy of complexity per se with application to some examples in evolution. In F. Heylighen & D. Aerts (Eds.), *The Evolution of Complexity*. Dordrecht: Kluwer.
- Hansman, R. J., & Davison, H. J. (2000). *The Effect of Shared Information on Pilot/Controller and Controller/Controller Interactions*. Paper presented at the 3rd USA/Europe Air Traffic Management R&D Seminar, Napoli.
- Hilburn, B. (2004). *Cognitive Complexity in Air Traffic Control - A Literature Review* (EEC Note No. 04/04): Eurocontrol.
- Histon, J. M., Hansman, R. J., Gottlieb, B., Kleinwaks, H., Yenson, S., Delahaye, D., & Puechmorel, S. (2002). *Structural Considerations and Cognitive Complexity in Air Traffic Control*. Paper presented at the 19th IEEE/AIAA Digital Avionics Systems Conference, Irvine, CA.
- Jeffrey, T. E., & Beck, F. J. (1972). *Intelligence information from total optical color imagery*. Washington DC: US Army Behavior and System Research Laboratory.
- Kirwan, B., Scaife, R., & Kennedy, R. (2001). Investigating complexity factors in UK Air Traffic Management. *Human Factors and Aerospace Safety*, 1(2), 125-144.
- Kopardekar, P. (2003). *Measurement and Prediction of Dynamic Density*. Paper presented at the ATM 2003 R&D Seminar, Budapest.
- 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.
- Miller, C. (2000). The Human Factor in Complexity. In T. Samad & J. Weyrauch (Eds.), *Automation, Control, and Complexity: New Developments and Directions*. New York: John Wiley.
- Pawlak, W. S., Brinton, C. R., Crouch, K., & Lancaster, K. M. (1996). *A framework for the evaluation of air traffic control complexity*. Paper presented at the AIAA Guidance Navigation and Control Conference., San Diego.
- Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8, 368-373.
- Reynolds, L. (1994). Colour for air traffic control displays. *Displays*, 15(4), 215-225.
- Simon, H. (1996). *The Sciences of the Artificial* (3rd ed.). Cambridge: MIT Press.
- Simons, D. J. (2000). Attentional capture and inattention blindness. *Trends in Cognitive Sciences*, 4, 147-155.
- Spencer, A., Smith, P., Billings, C., Brinton, C. R., & Atkins, S. (2002). *Support of Traffic Management Coordinators in an Airport Air Traffic Control Tower Using the Surface Management System*. Paper presented at the International Conference on Human-Computer Interaction in Aeronautics, Cambridge, MA.
- Tsonis, C., Cunha, D. C., Histon, J., & Cummings, M. L. (2005). *Developing Visualization Tools for the Design of Air Traffic Control Experiments*. Paper presented at the 49th Annual Meeting of the Human Factors and Ergonomics Society, Orlando, FL.
- Wickens, C. D., & Hollands, J. G. (2000). *Engineering Psychology and Human Performance* (Third ed.). Upper Saddle River, NJ: Prentice-Hall Inc.
- Woods, D. D. (1991). The cognitive engineering of problem representations. In G. R. S. Weir & J. L. Alty (Eds.), *Human-computer interaction and complex systems*. London: Taylor and Francis.
- Xing, J., & Schroeder, D. J. (2004). *Reexamination of color vision standards. I. Status of color use in ATC displays and demography of color-deficit controllers*: FAA Tech Report.