ote technical note technical

Moving Toward an Air Traffic Control Display Standard: Creating a Standardized Target Symbology for Terminal Situation Displays

Ferne Friedman-Berg, Ph.D., Human Factors Team – Atlantic City, ATO-P
Kenneth Allendoerfer, Ph.D., Human Factors Team – Atlantic City, ATO-P
Atul Deshmukh, Ph.D., Hi-Tec Systems, Inc.

August 2010

DOT/FAA/TC-TN-10/04

U.S. Department of Transportation
Federal Aviation Administration

William J. Hughes Technical Center
Atlantic City International Airport, NJ 08405

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/TC-TN-10/04		

4. Title and Subtitle

Moving Toward an Air Traffic Control Display Standard: Creating a Standardized Target Symbology for Terminal Situation Displays

5. Report Date

August 2010

6. Performing Organization Code

AJP-6110

7. Author(s)

Ferne Friedman-Berg, Ph.D., Human Factors Team – Atlantic City, ATO-P
Kenneth Allendoerfer, Ph.D., Human Factors Team – Atlantic City, ATO-P
Atul Deshmukh, Ph.D., Hi-Tec Systems, Inc.

8. Performing Organization Report No.

DOT/FAA/TC-TN-10/04

9. Performing Organization Name and Address

Federal Aviation Administration
Human Factors Team – Atlantic City, ATO-P
William J. Hughes Technical Center
Atlantic City International Airport, NJ 08405

10. Work Unit No. (TRAIS)

11. Contract or Grant No.

12. Sponsoring Agency Name and Address

Federal Aviation Administration
Human Factors Research and Engineering Group
800 Independence Avenue, S.W.
Washington, DC 20591

13. Type of Report and Period Covered

Technical Note

14. Sponsoring Agency Code

ATO-P

15. Supplementary Notes

16. Abstract

In this study, we applied human factors best practices to the design of an enhanced target symbol set for terminal displays. To design new symbols, we first identified types of information that would be most operationally useful. We selected four types of information that would provide the greatest operational benefits and tactical support for the controllers. These included aircraft category (Small, Large, Heavy, and Super Heavy), aircraft heading, aircraft altitude, and aircraft conformance to its assigned route or altitude. To evaluate the effectiveness of the symbols, we used search, sorting, and selection tasks that measured symbol preference, reaction time, and symbol identification. Our results indicated that the controllers used consistent heuristics for categorizing symbols into different size and conformance categories. For both the heading and category coding, we also found benefits in terms of both increased accuracy and decreased reaction times. We did not find any benefits for the altitude coding. By using meaningful symbols to convey relevant tactical information, such as aircraft category and heading, we can both increase visual search speed and increase target detection accuracy. On the basis of our findings, we propose a set of symbols and provide recommendations for creating standardized symbology for terminal situation displays and other safety critical systems.

17. Key Words

Aircraft Category
Aircraft Conformance
Altitude
Heading
Symbol Design
Visual Search

18. Distribution Statement

This document is available to the public through the National Technical Information Service, Springfield, Virginia, 22161. A copy is retained for reference at the William J. Hughes Technical Center Library.

19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	72	

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

Table of Contents

List of Illustrations

Tables Page

Acknowledgments

We thank the Air Traffic Organization Human Factors Research and Engineering Group Program Office for sponsoring this project. We also thank Jim McGhee (Northrop Grumman) and Rich Ridgeway (Northrop Grumman) for testing and retesting the software and for meticulously checking 256 slides for inaccuracies; remaining slide inaccuracies are entirely our responsibility. Thanks to Anton Koros, who assisted with the background research and experimental design. We thank Bina Pastakia, Liesl Powers, and Stephen Nogar (Hi-Tec Systems, Inc.), who assisted with the data collection. Thanks to Sonia Alvidrez (Hi-Tec Systems, Inc.), who assisted with the data entry. Thanks to the facility managers who graciously allowed us to visit their facilities and helped to coordinate our visits. We thank Robert Bortu (Engility Corporation) for his help with computer hardware and software. Thanks to all of the participants who kindly agreed to volunteer for this study. We also thank April Jackman (Northrop Grumman) for her careful review and editing of this document.

Executive Summary

The target symbology used on the Automated Radar Terminal System (ARTS) and Standard Terminal Automation Replacement System (STARS) displays reflects the display and surveillance technology available when they were designed. Given today's advances in technological capabilities, along with the introduction of new surveillance techniques, it is time to reexamine the target symbology used on terminal displays. The Air Traffic Organization Human Factors Research and Engineering Group sponsored the Human Factors Team−Atlantic City (ATO-P) to research potential aircraft symbology. The assessment described herein focuses on incorporating human factors standards and best practices into the development of symbols for the next generation of terminal displays.

As different systems compete for the highly prized real estate of the target symbol on Air Traffic Control (ATC) displays, researchers need to determine which of all potentially useful information types are the most tactically relevant for the controllers and the most optimal way to code this information. Poorly designed symbology can lead to errors and can make a system unsuitable for performing certain functions.

In this study, we applied human factors best practices in the design of an enhanced target symbol set for terminal displays. To design new target symbols for terminal displays, we first identified types of information that would be most operationally useful. We selected four types of information that would potentially provide the greatest operational benefits and tactical support for the controllers. These included aircraft category (*Small*, *Large*, *Heavy*, and *Super Heavy*), aircraft heading, aircraft altitude, and aircraft conformance to its assigned route or altitude.

We examined research on the design of target symbols for other displays that use symbolic representations of aircraft (e.g., Department of Defense radars and Cockpit Displays of Traffic Information) and selected some of the basic features, coding, and shapes suggested by that research. We coded heading using a triangle that pointed in the direction of flight. To code altitude, we used shades of green with lighter greens to indicate higher altitudes. To code category, we used the triangle as the base shape and then added distinctive features (such as lines, outlines, or holes) to indicate category membership. To code aircraft conformance, we surrounded the target symbol with chevron-like brackets and used color (red, yellow, and green) to indicate conformance states.

To evaluate the effectiveness of the proposed symbol sets, we used search, sorting, and selection tasks that measured symbol preference, reaction time, and symbol identification. Thirty-seven controllers performed search tasks using both existing coding schemes and novel coding schemes. By comparing performance on the tasks, we could determine how the proposed coding schemes affected search times.

We found that the controllers used consistent heuristics when assigning symbols to different size categories. For example, the controllers preferred simple, basic shapes (such as a triangle) as symbols for Small aircraft. For symbols representing intermediate size categories, such as Large and Heavy, they preferred basic shapes with a distinctive feature (such as a line or a hole). For a Super Heavy aircraft symbol, the controllers preferred basic shapes with an added outline, multiple lines, or other coding to symbolize the largest category.

For both heading and category coding, we found benefits in terms of both increased accuracy and decreased reaction times. We believe that by using meaningful symbols to convey relevant tactical information, such as aircraft category and heading, we can both increase visual search speed and increase target detection accuracy. If controllers can identify an aircraft target more quickly and more accurately with the addition of simple coding, then we believe such coding warrants further research, perhaps in a human-in-the-loop simulation, because it may have a significant positive impact on operational efficiency and safety. As implemented in our experiment, we did not find that altitude coding conferred any benefit on either accuracy or reaction time. However, we believe it may be beneficial to explore alternate ways to code altitude because of its tactical relevance for terminal controllers.

Feedback from the controllers indicated that we should not clutter the radar display, the target, or the datablock with too much coding. Although the aircraft heading and aircraft category coding were acceptable, controllers believed that too much unnecessary or difficult-to-process data had the potential to distract them from their primary task of separating aircraft.

The controllers also consistently assigned symbols to the four aircraft conformance categories. Feedback from the controllers indicated that these symbols were intuitive and easy to learn. Also, the colors used to indicate the status level for an aircraft matched standard color coding conventions. We believe that as conformance monitoring becomes more tactically relevant in the Next Generation Air Transportation System (NextGen) timeframe and as tailored arrivals become commonplace, simple and intuitive coding for information (such as route or altitude conformance) will become a requirement.

As we move toward NextGen, there are a number of hurdles that designers will need to clear. Many new systems will be fielded and much more information will be available. However, it will only be possible to present a fraction of this information on the controller's tactical display. With multiple NextGen tools, such as Automatic Dependent Surveillance-Broadcast (ADS-B) and Data Communications (Data Comm), competing for the limited datablock real estate, it makes sense to move the most tactically relevant ATC information from the datablock into the target symbol.

When creating target symbols that code tactically relevant information, system designers must ensure that they are easy to interpret and that they take advantage of both perceptually salient and culturally learned coding conventions. Our assessment is a starting point that designers can use when making decisions about how to create new candidate symbols for future ATC terminal displays or how to prioritize the information to be coded in these symbols. Designers and researchers can use the recommendations provided not only when developing and testing new symbols for terminal displays but also when designing displays for other safety-critical domains.

1. INTRODUCTION

In 2008, the Federal Aviation Administration (FAA) issued the 2009–2013 Flight Plan outlining a five-year strategic plan to meet the increasing airspace demand (FAA, 2008). A cornerstone of this plan is the implementation of Next Generation Air Transportation System (NextGen). This technology will dramatically transform the way the National Airspace System (NAS) provides air navigation services. The focus will shift from one of Air Traffic Control (ATC) to one of Air Traffic Management. The five transformational NextGen programs include Automatic Dependent Surveillance-Broadcast (ADS-B), System-Wide Information Management (SWIM), NextGen Data Communications (Data Comm), NextGen Network Enabled Weather (NNEW), and NAS Voice Switch (NVS).

Under NextGen, aircraft will use ADS-B to determine the position, speed, and direction of an aircraft and will automatically broadcast this information to air traffic managers and pilots. Reliance on ADS-B capabilities instead of ground-based radars will allow for the display of more precise positional information on ATC terminal displays, including Terminal Radar Approach Control (TRACON) displays and tower displays,[1] without employing display artifacts designed for traditional radar systems. Current target symbology reflects the display and surveillance technology that was available when the original Automated Radar Terminal System (ARTS) and Standard Terminal Automation Replacement System (STARS) displays were designed. Given new technological capabilities and new surveillance techniques, it is time to reexamine the target symbology on terminal displays.

For instance, it may not be necessary to display both a primary target and a beacon target to indicate different surveillance sources. Because ADS-B uses GPS to determine aircraft positions, positional uncertainty varies minimally as an aircraft moves through the airspace. Having a target that orients toward a sensor or that adjusts the target length relative to distance from the radar will no longer provide tactically useful information. As system capabilities evolve, we believe that it is appropriate to reconsider how we display aircraft information on terminal displays and investigate optimal target symbology.

1.1 Purpose

The Air Traffic Organization, Human Factors Research and Engineering Group, has sponsored the Human Factors Team–Atlantic City (ATO-P) to research potential aircraft symbology for this new generation of displays. The assessment described herein focuses on incorporating human factors standards and best practices into the development of symbols for the next generation of terminal displays. This study will allow us to provide appropriate guidance to the developers and vendors of new display systems.

1.2 Background

Designing appropriate symbology for safety critical systems is an extremely important component of good display design. Poorly designed symbology can lead to errors and can make a system

[1] Tower displays require additional considerations regarding viewing distance and lighting conditions that we do not address in this work.

unsuitable for performing certain functions. Even given the importance of symbol design, display designers often do not follow human factors best practices when creating new symbol sets. For example, system developers often design symbols by consensus without evaluating their proposed designs scientifically. In addition, although much guidance exists on avoiding bad designs, there is much less guidance on developing good designs (Carney, Campbell, & Mitchell, 1998).

In this study, we applied human factors best practices in the design of an enhanced target symbol set for terminal displays. To design new target symbols for terminal displays, we first identified types of information that would be operationally useful. Some critical information types in ATC include aircraft separation distances, speeds (both relative and absolute), headings, times until conflict, altitudes, communication types, intent, destination, data reliability, and separation standards (Nunes & Mogford, 2003). However, encoding all of these information types into the target would be impossible. Therefore, we selected four types of information that would potentially provide the most operational benefits and tactical support for the controllers. These included aircraft category (*Small*, *Large*, *Heavy*, and *Super Heavy*), aircraft heading, aircraft altitude, and aircraft conformance to its assigned route or altitude. We examined research on the design of target symbols for other displays that use symbolic representations of aircraft (e.g., Department of Defense radars and Cockpit Displays of Traffic Information) and selected some of the basic features, coding, and shapes suggested by that research (Palmer, Clausner, & Kellman, 2008; Smallman, St. John, Oonk, & Cowen, 2001a, 2001b; Tam, Corker, & Duong, 2003). To employ human factors best practices when designing these new symbol sets, we also took advantage of positive transfer by using features from old symbol sets in the design of the new symbol sets (Cardosi & Murphy, 1995).

To evaluate the effectiveness of the proposed symbol set, we needed to determine which test or measure to use in our evaluation. Determining which measure to use depends both on the purpose of the symbol and where it will be used (Mackett-Stout & Dewar, 1981). There are some measures, such as symbol comprehension and symbol preference, that are essential for evaluating all potential symbols (Green & Pew, 1978). For safety critical domains like ATC, reaction time (RT) measures are also essential because safety critical information must draw attention rapidly. Therefore, we used search, sorting, and selection tasks that measured symbol preference, RT, and symbol identification to evaluate the candidate symbols. For example, we asked the controllers to perform search tasks using both existing coding schemes and novel coding schemes to determine how the proposed coding schemes would affect search times. Using these findings, we evaluated controller stereotypes for the coding of target symbols. On the basis of our findings, we propose a number of changes to the target symbols used on terminal displays.

2. METHOD

2.1 Participants

Thirty-seven FAA Air Traffic Controllers participated in this study: Twenty-six were TRACON controllers and 11 were en route controllers on travel to the William J. Hughes Technical Center (WJHTC) in support of other projects. Twenty-nine participants were male and 8 were female. The average age of the participants was 50 yrs (s = 8 yrs).

2.2 Early Termination Criteria

This was a minimal risk study and participation in the study was voluntary. We instructed the controllers that they could end their participation at any time, for any reason, without penalty. In addition, if the researcher determined that terminating the session would be in the best interests of the participant, the researcher could end the session at any time. However, none of the sessions were terminated prior to the completion of the experimental session.

2.3 Personnel

Two engineering research psychologists from the Human Factors Research and Engineering Group, Human Factors Team-Atlantic City (ATO-P, AJP-6110) and one contract support personnel from Hi-Tec Systems, Inc. conducted the assessment. Two Subject Matter Experts (SMEs) and a summer intern from Rutgers University, New Brunswick, NJ, provided support for the data collection activity.

2.4 Facilities

We conducted this study at the Research Development and Human Factors Laboratory (RDHFL) and at three field sites. The field sites visited included Potomac TRACON, Atlantic City TRACON and Tower, and Philadelphia TRACON and Tower.

2.5 Equipment

We conducted this study on laptop computers using E-Prime (Version 2.0) experimental software program software.

2.6 Procedure

Experimental sessions lasted approximately 45–60 minutes. This included time for instructions, informed consent, questionnaires, training, data collection, and debriefing. All participants completed seven target symbology tasks. The data collection program provided the instructions for the individual tasks and recorded all relevant data.

Prior to data collection, we briefed each participant regarding the goals of the project, the procedures and data collection techniques, and their rights as participants. Each participant signed an Informed Consent Form (see Appendix A) and completed a Background Questionnaire (see Appendix B). After completing the Informed Consent Form and Background Questionnaire, the participants read an introduction that explained the general purpose of the experiment (see Appendix C).

The participants then completed the following seven symbology-related tasks.

- Task 1: Target Symbology Coding Convention (A)
- Task 2: Target Symbology Coding Convention (B)
- Task 3: Baseline STARS Symbology
- Task 4: Enhanced STARS Symbology with Heading Information
- Task 5: Enhanced STARS Symbology with Heading and Category Information

- Task 6: Enhanced STARS Symbology with Heading, Category, and Altitude Information
- Task 7: Enhanced STARS Symbology with Conformance Information

We based our procedures for the coding convention tasks (Tasks 1 and 2) on procedures used in an evaluation of symbols for electronic displays of aeronautical charting and navigation information (Chandra & Yeh, 2007; Yeh & Chandra, 2005, 2006).

2.6.1 Task 1: Target Symbology Coding Convention (A)

For Task 1, the participants viewed 24 target symbols (see Figure 1). The symbols that we created and tested are similar to *symbicons*, created by Smallman et al. (2001a, 2001b) for tactical military displays. According to Smallman et al., symbols use nonmeaningful coding (i.e., current TRACON target symbols), icons use overly realistic coding (e.g., a picture or an outline of an aircraft), and symbicons code use more simplified visual coding to code meaningful information. For a more thorough discussion of how the symbols we created are more like symbicons see Friedman-Berg, Allendoerfer, and Deshmukh (2010).

Figure 1. Screenshot of Task 1 display.

Prior research by Smallman et al. (2001a, 2001b) found that there are many benefits to using symbicons to code information. Symbicons not only can increase situational awareness but also can increase visual search speed and decrease error rates when compared to both icons and symbols. These benefits are highly relevant in a domain like ATC, where situational awareness, speed, and accuracy are so important.

When creating the symbols for this task, we first calculated the appropriate target size using the methodology recommended by both the human factors design standard (Ahlstrom & Longo, 2003, HFDS 8.2.5.6.5) and the American National Standards Institute (ANSI, 1988) standard. The standards advise that for maximum legibility, the optimal character height is 20–22 minutes of visual arc. ATC displays are about an arm length from the user, and the functional arm reach of the 95th percentile male is 94.2 cm. Converting minutes of visual arc to millimeters using the 95th percentile distance, the recommended size for a target that will still be legible for the 95th percentile male is 5.75 mm. This converts to a pixel height of 23. The current size of the beacon target is 6 pixels high by 15 pixels wide. Therefore, all of our symbols were created to be 23 pixels high by 15 pixels wide.[2] We only deemed a symbol to be acceptable if the details of the symbols were still distinguishable at these dimensions.

The E-Prime (Version 2.0) experimental software program presented all 24 symbols simultaneously, but randomized the presentation order of the target symbols for each participant. The participants received instructions before the start of Task 1. The instructions directed the participants to select the four symbols that they believed best represented the four aircraft categories: Small, Large, Heavy, and Super Heavy (see Appendix C). The participants dragged the selected symbol and then dropped it in the selected category.

The purpose of this task was to identify preexisting symbol stereotypes (Yeh & Chandra, 2006) for the four categories and to ensure that future symbol sets contain intuitively coded symbols that align with controller expectations. Yeh and Chandra define symbol stereotypes as those symbols that have "key features that are necessary for recognition" or, in other words, those symbols and the features of those symbols that are most commonly associated with a particular meaning. After the participants selected their preferred coding scheme, we asked them to rate the goodness of the scheme on a 7-point scale ranging from 1 (*worst possible scheme*) to 7 (*best possible scheme*); see Figure 2.

[2] The pixel dimensions represent the maximum width and height of the symbol.

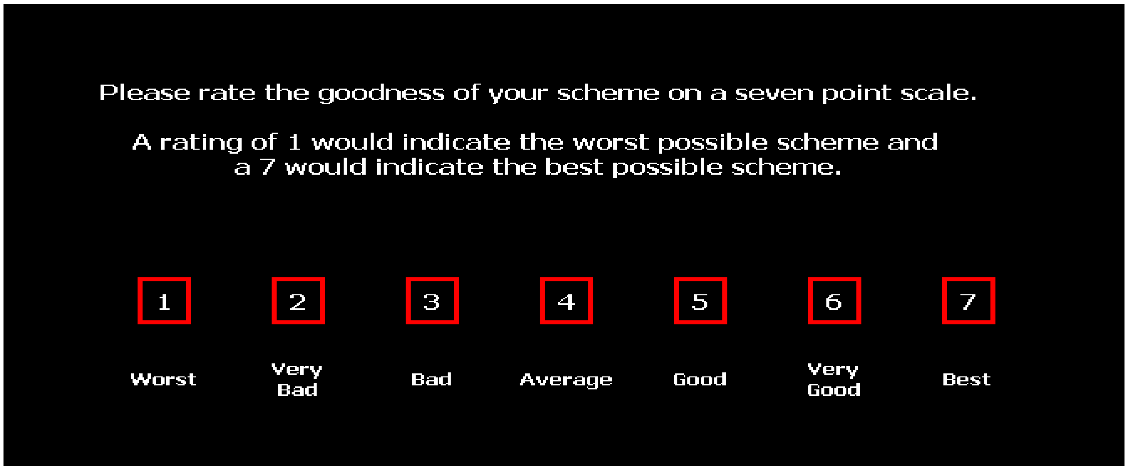


Figure 2. Screenshot of scale.

2.6.2 Task 2: Target Symbology Coding Convention (B)

For Task 2, the participants viewed 12 of the 24 target symbols seen in Task 1. The E-Prime (Version 2.0) experimental software program presented each symbol as a separate trial and randomized the presentation order of the trials (see Figure 3). The instructions told the participants to assign one of the four aircraft size categories (i.e., Small, Large, Heavy, or Super Heavy) to each symbol (see Appendix C).

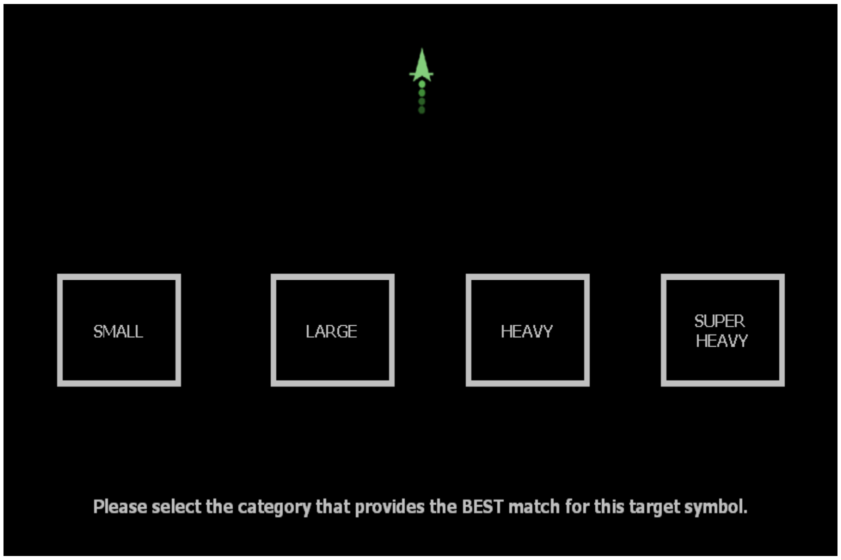


Figure 3. Screenshot of Task 2 trial.

Although Task 1 allowed us to investigate the most popular symbol stereotypes for the different aircraft categories, it did not provide information about stereotypes for less frequently selected icons. Task 2 allowed us to collect category assignment preferences for some of these less frequently selected symbols so that we could identify their stereotypes.

2.6.3 Tasks 3 through 6: Visual Search Tasks

For Tasks 3, 4, 5, and 6, we evaluated how different types of symbol modifications affected target detection. We coded three types of aircraft information: direction, category, and altitude. We tested how these three types of coding affected visual search. Task 3 used current STARS symbols, and Tasks 4, 5, and 6 used novel coding schemes. To measure the effect of each coding type on visual search performance, the participants performed a visual search task where they searched for an aircraft of a particular size (Small, Large, Heavy, or Super Heavy), heading (North, South, Northwest, or Southwest), altitude (Highest, Lowest, 0–3,000 ft, 3,000–6,000 ft, 6,000–9,000 ft, and 9,000–12,000 ft), or category + heading. For category + heading queries, there was always one distractor target from the same category and one distractor target with the same heading to prevent the participants from selecting the correct aircraft on the basis of a single feature.

The participants responded to all four question types in each of the four tasks. For each of the four question types, the location of the correct answer was counterbalanced across the display. This was to ensure that the average visual search RT was a function of search difficulty and not the distance of the correct answer from the initial fixation location. To counterbalance the location of the correct answer, we divided the display into a 4 x 4 grid. For each query type, the correct response appeared once in each of the 16 quadrants on the display.

2.6.3.1 Task 3: Baseline STARS Symbology

For Task 3, the controllers were told that they would be seeing a mock-up of a radar display and that their task would be to search for a target based on the cue given before each trial (see Appendix C for instructions). Before each trial, we instructed the participants to focus on a fixation cross at the center of the screen. After a short lag, the E-Prime software removed the fixation cross and then replaced it with the cue for that trial. The cue appeared in the same location on the screen as the fixation cross.

After reading the cue, the participants pressed the *Enter* key to begin the trial. This key press brought up a radar display image depicting 10 aircraft targets randomly distributed across the display (see Figure 4). The key press also moved the cursor to the center of the display and moved the instructions to the bottom of the screen. The participants' task was to select the aircraft that matched the query.

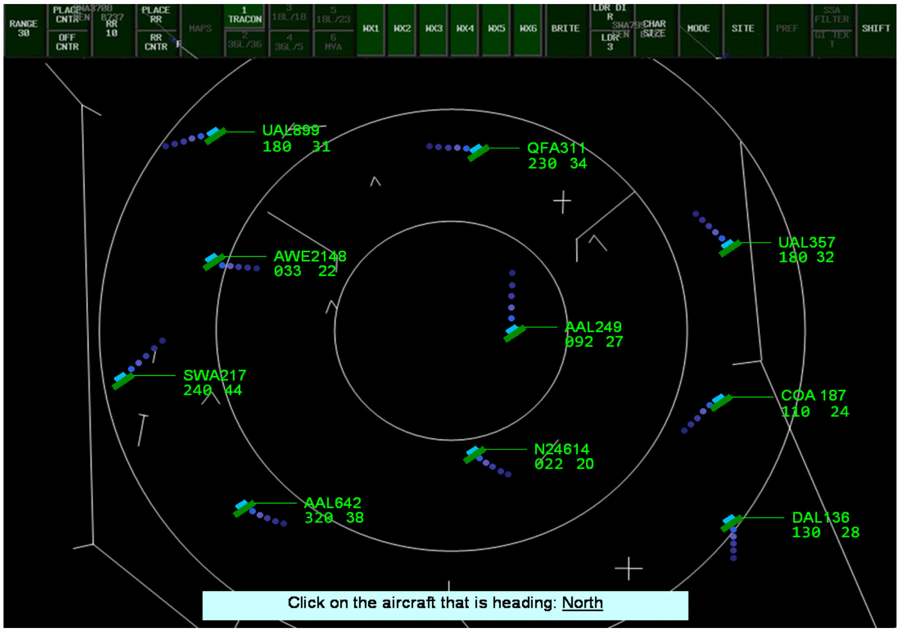


Figure 4. Screenshot of baseline STARS symbology.

The participants completed a total of 64 trials, 16 for each query type (see Appendix D). For this task, we used target symbols from a typical STARS terminal display operating in single sensor mode (i.e., the target and beacon symbols were oriented to the radar). The E-Prime experimental software program randomized the presentation order for all 64 trials.

2.6.3.2 Task 4: Enhanced STARS Symbology with Heading Information

Task 4 utilized the same procedures used for the baseline STARS symbology condition (Task 3). However, for Task 4, we replaced the standard STARS target and beacon symbols with an acute isosceles triangle because it used shape coding to indicate aircraft heading (see Figure 5). The point of the triangle indicated the direction of flight (see Appendix C for instructions).

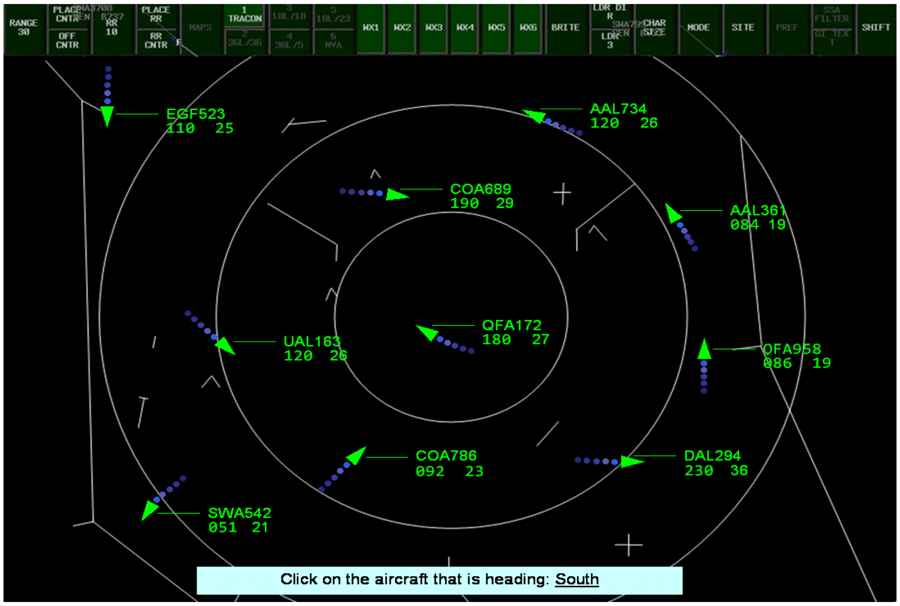
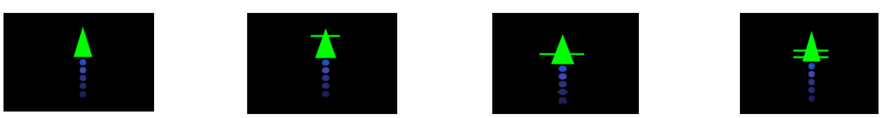


Figure 5. Screenshot of enhanced STARS symbology with heading information.

2.6.3.3 Task 5: Enhanced STARS Symbology with Heading and Category Information

Task 6 used the same procedures as used for Tasks 3 and 4. However, for this condition, we replaced the standard STARS target and beacon symbols with symbols that also coded aircraft category (Small, Large, Heavy, and Super Heavy; see Figures 6 and 7). We designed the symbols so that the placement or absence of lines indicated category membership. The triangle with no lines indicated the Small category. We believed that the smallest aircraft should have the most basic symbol to maintain cognitive consistency between the symbol and category. As the size category increased, the number of attributes on the symbol increased. Large aircraft were given a line in the front of the triangle, Heavy aircraft were given a line at the back of the triangle, and Super Heavy aircraft were given two lines at the back of the triangle (see Appendix C for instructions).



a b c d

Figure 6. Test symbols for the categories (a) small, (b) large, (c) heavy, and (d) super heavy.

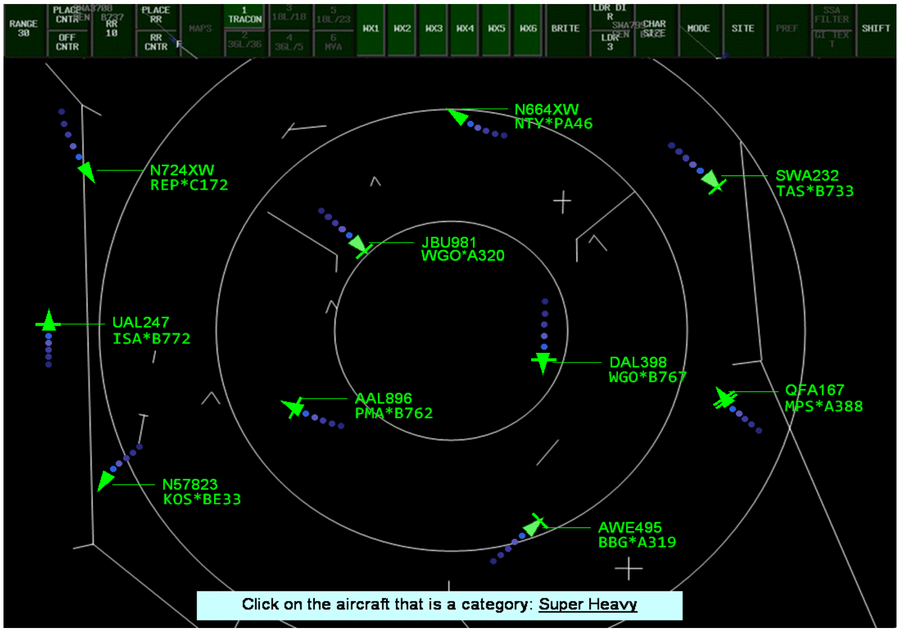


Figure 7. Screenshot of enhanced symbology with heading and category information.

2.6.3.4 Task 6: Enhanced STARS Symbology with Heading, Category, and Altitude Information

We used the same procedures for Task 6 that we used for Tasks 3, 4, and 5. However, for Task 6 we replaced the standard STARS target and beacon symbols with symbols that also coded altitude (see Figure 8). The aircraft shaded the darkest green were those at the lowest altitudes, while the lightest greens were those at the highest altitude (see Appendix C for instructions).

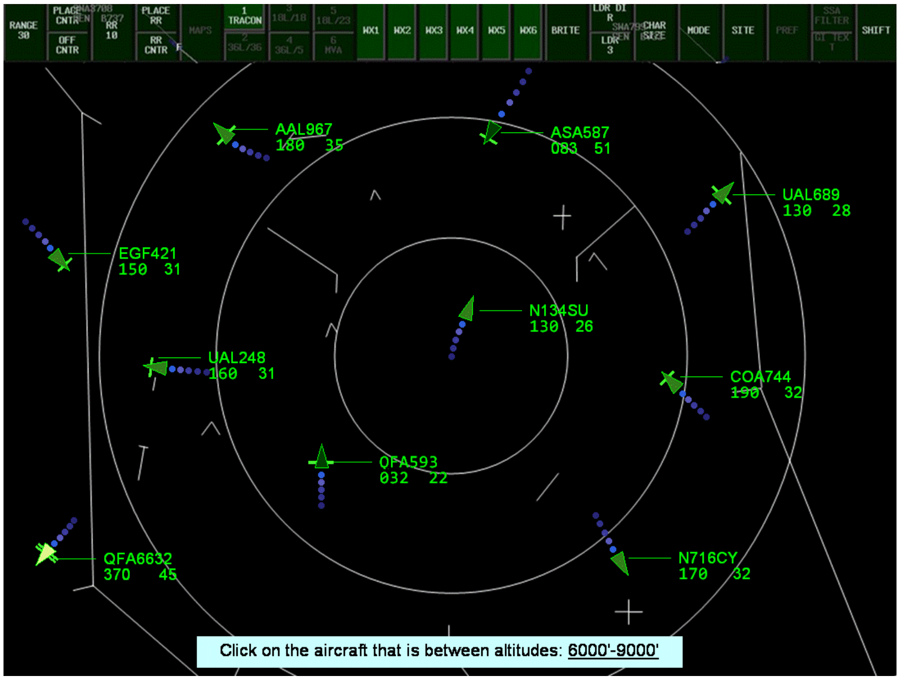


Figure 8. Screenshot of enhanced STARS symbology with heading, category, and altitude information.

2.6.3.5 Tasks 3 Through 6: Counterbalancing and Methodological Issues

Regarding our methodology, we did not control for a task order effect for Tasks 3, 4, 5, and 6; the task presentation order was the same for all participants. Tasks 3, 4, 5, and 6 were considered to be a set with the coding for each task in the set building upon the previous coding.

From a learning standpoint, we believed it was optimal to start with a baseline condition using today's target symbols. We incrementally added coding for direction, then category, and then altitude to the symbol set. Our rationale for adding the coding in this order is that we added the most conceptually simple and tactically relevant information first and the most conceptually complex last.

Because Task 3 used the current target symbols and required no learning on the part of the controller, it provided the baseline RT and accuracy measures for the four question types. Task 4 introduced coding for heading and required a minimal amount of learning. If the coding of heading into the target symbol had an effect on performance, we would expect the accuracy and the RTs for the heading questions to differ significantly between Task 3 and Task 4.

Task 5 maintains the coding of heading from Task 4 and adds a second form of coding for aircraft category. If the coding for aircraft category had an effect on performance, then the accuracy and RTs for the aircraft category questions should differ significantly between Task 5 and Tasks 3 and 4 (the corresponding statistical contrast is T5 - ½(T3 + T4). We also measured whether the addition of the category coding had any detrimental impact on the effect of coding for heading. That is, if there is an effect of coding for heading in Task 4, is the effect diluted or reversed by adding category coding in Task 5?

Task 6 maintains the coding learned in Tasks 4 and 5 and adds a third form of coding for altitude. Although it is possible to add altitude coding before category coding, category coding is conceptually simpler and potentially more operationally useful than altitude coding. Therefore, we chose to introduce category coding prior to the altitude coding.

For the heading questions for Tasks 3, 4, 5, and 6, we asked about both cardinal directions (i.e., north, south) and noncardinal directions (i.e., northeast, southeast). We believed that it might be easier to search for aircraft heading in cardinal directions than in noncardinal directions (Wolfe, 1998). Therefore, we analyzed performance overall, and then we compared performance on the cardinal and noncardinal queries.

We selected the category coding for Tasks 3, 4, 5, and 6 prior to obtaining feedback on potential symbols from the controllers. We then used the findings from Tasks 1 and 2 and the subjective feedback to determine whether our coding was optimal. On the basis of an analysis of data from Tasks 1 and 2 and feedback from participants, we determined that our symbols for Large and Heavy aircraft were not as optimal as our symbols for Small and Super Heavy aircraft. Therefore, for our data analysis we first analyzed overall performance on the category queries, and then we compared controller performance on queries about optimally coded symbols to controller performance on queries about nonoptimally coded symbols.

The altitude queries asked about both relative (relational) altitudes (e.g., Find the aircraft that has the highest altitude.) and specific (absolute) altitudes (e.g., Which aircraft is between 3,000 and

6,000 ft?). We wanted to evaluate how these different query types affected performance. Because the controllers often are concerned with the relative position of aircraft, we believed that relative altitude questions might be easier to answer than specific altitude questions. Therefore, we analyzed the data from altitude queries overall, then we compared performance on the relative and specific altitude queries to determine the effects, if any, of query type on performance.

There were four different category + heading query types: optimal and cardinal queries, optimal and noncardinal queries, nonoptimal and cardinal queries, and nonoptimal and noncardinal queries. For the category + heading data analysis, there were a number of relevant questions. When participants had to use two different types of aircraft information to find a target aircraft, did either type have an impact on performance? Did the ease or accuracy of their search change with the addition of heading coding in Task 4? Because the participants were performing a conjunction search using two features, did the ease or accuracy of the search change only when both features were coded into the target in Task 5?

2.6.4 Task 7: Enhanced STARS Symbology with Conformance Information

Task 7 investigated whether the controllers could intuitively interpret symbols that indicated some type of aircraft conformance (e.g., route conformance, altitude conformance). When an aircraft has the potential to be monitored for conformance with a restriction of any kind, it has four potential states. The aircraft can be (a) not monitored; (b) monitored and in conformance with its restrictions; (c) monitored and in conformance, currently, with a projection to go out of conformance in the near future; or (d) out of conformance currently (see Figure 9). We instructed the controllers to indicate which level of conformance they believed was associated with each symbol by dragging and dropping one symbol in each of the four conformance categories (see Appendix C for instructions).

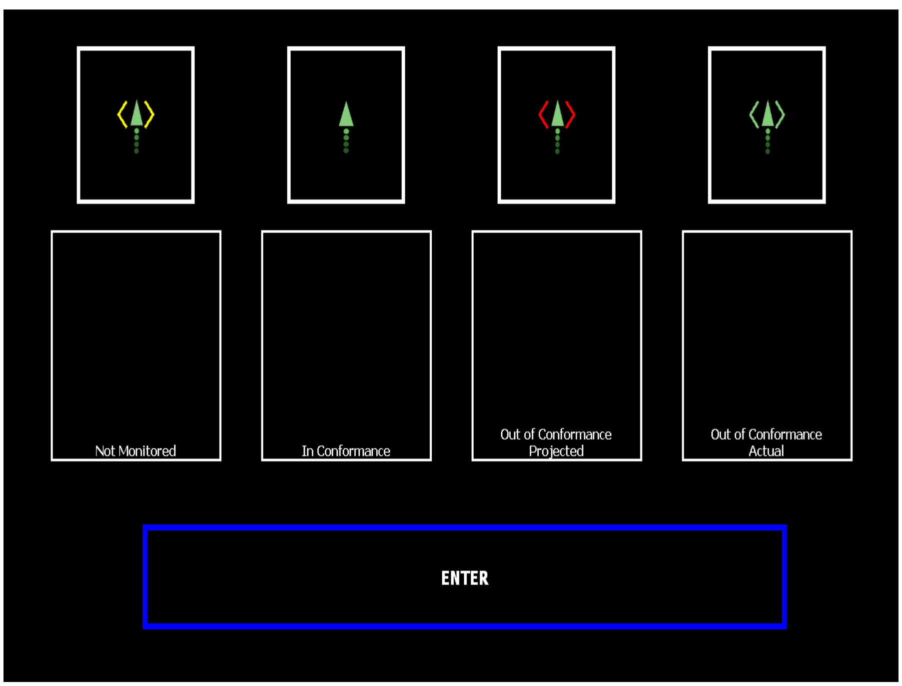


Figure 9. Screenshot of Task 7 with the four conformance monitoring symbols.

The E-Prime program randomized the order of the four novel symbols across the participants. We determined the most commonly identified coding conventions on the basis of their responses. The participants also rated their responses using a 7-point Likert Scale (as shown in Figure 2).

2.6.5 Exit Questionnaire

After completing all seven tasks, the participants completed an Exit Questionnaire (see Appendix E). The questionnaire consisted of yes or no and open-ended questions about the symbology used in this study. On the questionnaire, the controllers also had the opportunity to provide comments and recommendations about the suitability of the symbols and the cost-benefits of introducing enhanced target symbology into the operational ATC environment.

3. DATA ANALYSIS

For the data analysis, we computed descriptive statistics (e.g., frequencies, averages) for the Background Questionnaire data. We present descriptive statistics (e.g., category assignment proportions, common symbol groupings) for Tasks 1, 2, and 7 and inferential statistics (e.g., *t*-tests, analysis of variance, and contrasts) for Tasks 3 through 6. We also summarized the subjective responses from the Exit Questionnaire data regarding the use of enhanced target symbology. These data helped us to identify issues and benefits related to implementing enhanced target symbology in the TRACON and tower. On the basis of our analyses, we created a set of human factors recommendations regarding the use of enhanced symbology on TRACON displays.

3.1 Background Questionnaire

The 37 Air Traffic Controllers who participated in this study had a variety of job titles (see Figure 10). The majority were Certified Professional Controllers, followed by Front Line Managers. They also had a wide range of experience in different ATC settings (see Table 1) and with different ATC systems (see Table 2).

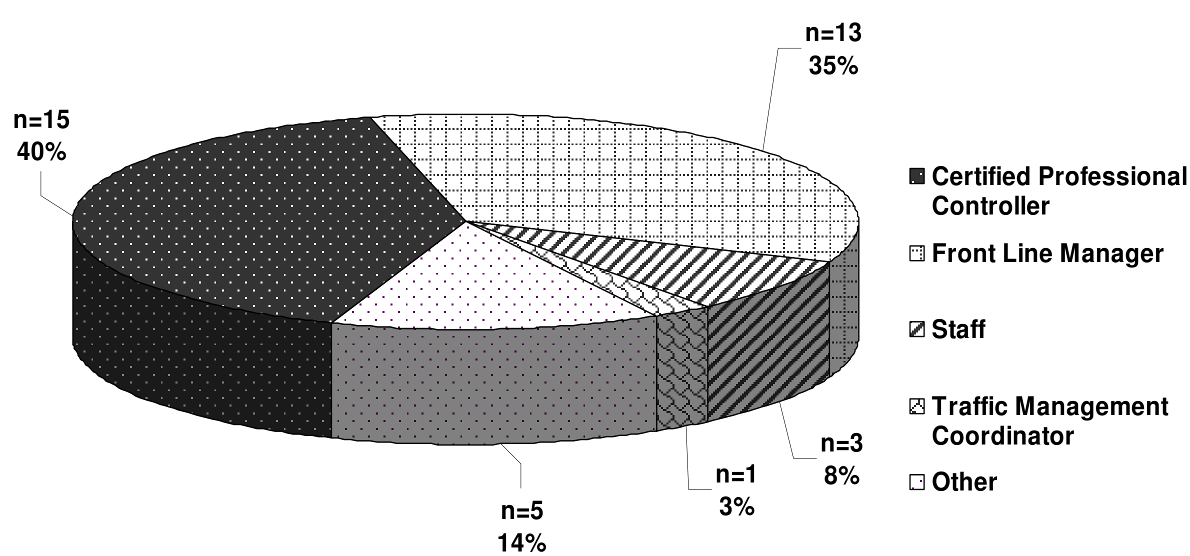


Figure 10. Percentage of participants with each job title.

Table 1. Mean Years Experience in Different ATC Settings

CPC/FLM Experience	$\bar{X}$ years	*s*	*n*
TRACON	20	7	31
En Route	17	9	17
Tower	19	7	25
TMU	12	10	7
Command Center	21	NA	1

Note. CPC = Certified Professional Controller; FLM = Front Line Manager; TMU = Traffic Management Unit.

Table 2. Mean Years Experience with Different ATC Systems

System Experience	$\bar{X}$ years	*s*	*n*
STARS	7	2	13
CARTS	13	7	8
ARTS IIIA	14	7	17
ARTS IIIE	13	7	12
Other	11	7	25

Note. STARS = Standard Terminal Automation Replacement System; CARTS = Common Automated Radar Terminal System; ARTS = Automated Radar Terminal System.

3.2 Stereotype Symbol Preferences for Aircraft Category Coding: Results for Tasks 1 and 2

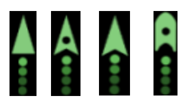
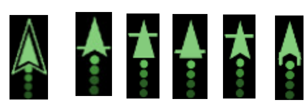
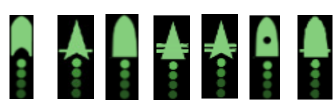
Prior research has investigated the coding of altitude into the aircraft target (Johnston, Horlitz, & Edmiston, 1993; Palmer et al., 2008; Tam et al., 2003), but we could find no work that has evaluated aircraft category coding. Yet, in the terminal environment, aircraft category may be at least as operationally useful as aircraft altitude. Aircraft category is important to TRACON controllers because different size aircraft have different performance characteristics. These characteristics may affect how different aircraft respond to ATC control instructions in the terminal environment. They may have different climb, descent, and turn rates; and they may also have different wake turbulence-related separation requirements.

3.2.1 Task 1 Results

For Task 1, the controllers selected four symbols from a set of 24 to best represent the four aircraft size categories: Small, Large, Heavy, and Super Heavy. Of the 24 choices, four did not receive more than one vote for any category, so we dropped them from the analysis. Of the remaining 20, one received an equal number of votes for two categories, so we list it as one of the preferred choices for each of those categories. We assigned the remaining 19 symbols to the category for which it was selected most frequently.

For Task 1, the controllers appeared to use certain heuristics or symbol stereotypes when selecting the symbols for the four size categories (see Table 3). The symbols in Table 3 are presented in descending frequency of selection, from left to right. The symbols the controllers most commonly assigned to the Small aircraft category were those that were more basic (Rosch, 1973), for instance, a simple triangle or minor variants of a triangle (i.e., a chevron). They had no lines or outlines and were, for the most part, not bullet shaped. Two of the preferred symbols had holes. This may be because holes represent "less" of some thing; in this case, less aircraft.

Table 3. Most Frequently Selected Aircraft Size Category for Symbols and the Similarities of Symbols within a Category

Category		Similarities
Small		• Basic Symbol • Hole • Triangle/Chevron
Large		• Nonbasic Symbol • Lines • Triangle/chevron
Heavy		• Nonbasic Symbol • Bullet-shaped • Double lines • Lines toward the "tail"
Super Heavy		• Nonbasic Symbol • Lines • Outline • Bullet-shaped

Note. The symbols are ordered from most frequently assigned to least frequently assigned.

The symbols the controllers most frequently assigned to the Large category also had many features in common. They tended to be nonbasic symbols, in that they often had lines or outlines. Similar to the symbols selected for the Small category, the primary shapes of the symbols selected for the Large category were triangles and chevrons.

As with the Large and Small categories, the symbols selected by the controllers for the Heavy category shared a number of common features. The controllers assigned many more bullet-shaped symbols to the Heavy category. In addition, the symbols selected for the Heavy category were nonbasic, and they typically had at least one line. Double lines were common, and the lines (whether single or double) were consistently placed toward the tail of the triangle or bullet shape.

For the Super Heavy category, the controllers typically assigned symbols that were nonbasic and bullet shaped. The symbols all contained lines or outlines.

The controllers also provided goodness ratings for their coding schemes on a scale of 1 to 7, with 1 indicating the *worst possible scheme* and 7 indicating the *best possible scheme*. The average rating was 4.72, *SE* = .21. Most of the ratings were 4 or higher (see Figure 11).

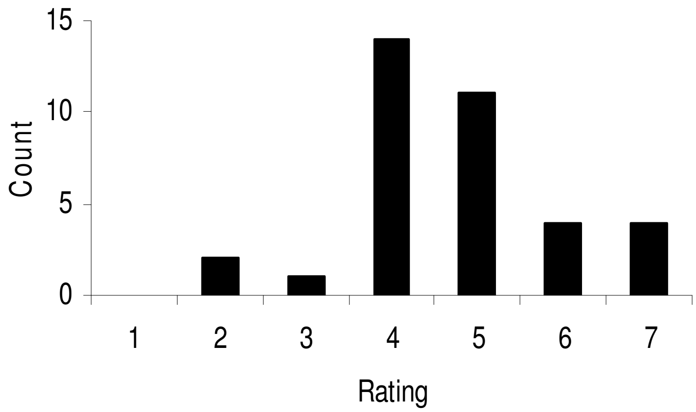


Figure 11. Frequency distribution of goodness ratings for Task 1.

3.2.2 Task 2 Results

For Task 2, the controllers assigned a preferred aircraft category to 12 symbols (see Table 4). We tested only 12 of the set of 24 symbols tested in Task 1, and all of the symbols we tested were triangles and chevrons. We reduced the number of symbols because of time constraints and because of the number of subtasks in the experiment.

Table 4. Proportion of Aircraft Category Assignment by Symbol Type

Category												
Small	.65	.70	.68	.43	.19	.5	.32	.16	.14	.8	.5	.24
Large	.30	.22	.24	.41	.51	.57	.49	.51	.46	.19	.19	.22
Heavy	.5	.8	0	.11	.27	.32	.16	.30	.32	.43	.43	.16
Super Heavy	0	0	.8	.5	.3	.5	.3	.3	.8	.30	.32	.38

For Task 2, the controllers categorized the symbols using similar heuristics to the ones they used for Task 1. They preferentially assigned the simpler symbols to the Small category. The symbols were basic triangles or chevrons and one had a hole. For the Large category, the controllers selected more nonbasic shapes. They preferentially assigned triangles and chevrons with a single line in the front or back to this category. Participants were somewhat indecisive in assigning the chevron with two lines projecting from the rear and the triangle with two lines projecting from the rear. Although there was a slight preference for the Small category for the chevron and the Large category for the triangle, they were both assigned almost equally to the

Small and Large categories. Both symbols with double horizontal lines toward the rear of the symbol were assigned to the Heavy category, and the one symbol with an outline was assigned to the Super Heavy category.

3.3 Visual Search: Tasks 3 through 6 Results

For Tasks 3 through 6, we compared visual search performance in baseline conditions to performance when using enhanced symbology. For each type of coding, we discuss observed differences in performance between the enhanced and baseline presentations. For our dependent measures, we evaluated how the addition of each type of coding affected performance. We measured whether each additional type of coding had a positive or negative impact on either accuracy or RT. If coding had an impact on speed or accuracy, either positive or negative, we also evaluated the point at which it began to have an impact and whether its impact was diluted or reversed by the addition of any other type of coding.

For the RT data, we first prepared the data for analysis. For our RT analysis, we followed the standard convention of including only correct responses (Ratcliff, 1993). To avoid positively skewing the average RT, we also dropped long RTs, which we defined as responses longer than 20 seconds. Ratcliff points out that using a lower RT cutoff (e.g., 1,000 ms) leads to greater power, and a higher cutoff can result in means that are more variable. A higher cutoff can reduce power and can make it more difficult to detect differences. However, because we conducted this study in the field and not in a controlled laboratory setting, our participants frequently had very long RTs. To avoid eliminating too many valid responses from our RT analysis, we used the higher cutoff along with a nonrecursive elimination procedure to eliminate extreme outliers. By using the higher cutoff, we ensured that we did not exclude valid but long responses. By using the nonrecursive elimination procedure to uniquely determine a cutoff for each participant, we ensured that the responses used to calculate a participant's average RTs were not too dissimilar.

For the nonrecursive outlier deletion procedure, we used a shifting z score criterion (Thompson, 2006; Van Selst & Jolicoeur, 1994) to eliminate responses that were at least 2.016 standard deviations above the mean RT. The z score criterion for each cell for each participant ranged from 2.016 to 2.8068. It varied as a function of the number of correct responses in that cell and it excluded any RT greater than 20 seconds.

3.3.1 Aircraft Heading Queries

We analyzed the accuracy data for the heading queries by calculating the difference in accuracy between Tasks 3 and 4. In the Task 3 display, aircraft heading could be determined by looking at the history trails but not by looking at the target symbols. In the Task 4 display, aircraft heading could be determined by looking at either the target symbols or the history trails. Although accuracy was high for all tasks, there was a significant increase in accuracy between Tasks 3 and 4. The mean difference between Tasks 3 and 4 was .086, $t(36) = 4.92$, $p < .001$, 95% CI = [.051, .122], $SE = .018$. Task 4 accuracy was not significantly different from Task 5 and Task 6 accuracy (see Figure 12).[3]

---

[3] All error bars in all figures are within subject confidence interval error bars (Cousineau, 2005).

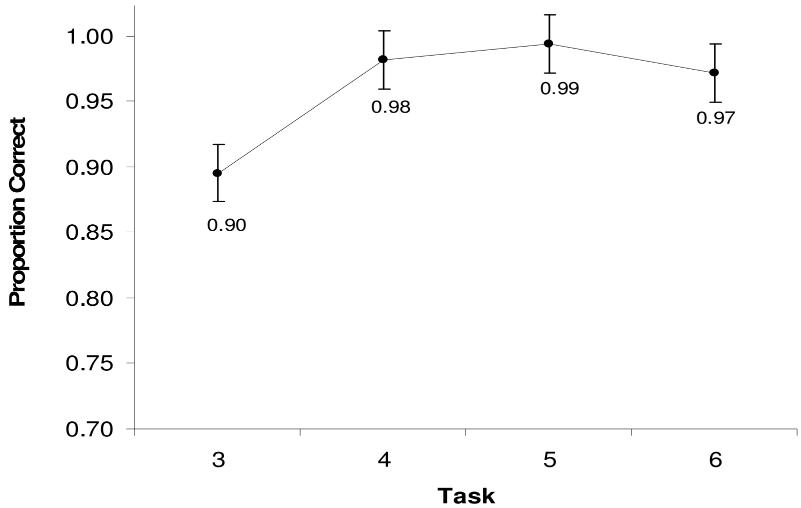


Figure 12. Proportion correct for heading queries by task.

When performing the same analysis on the RT data, we found that it mirrored the accuracy data. With the addition of coding for heading, there was a significant decrease in RT between Task 3 and Task 4. The mean difference between Tasks 3 and 4 was 700.25 ms, $t(36) = 5.51$, $p < .001$, 95% CI = [442.60, 957.91], $SE = 127.04$. The mean RT for Task 4 was not significantly different from the mean RT for Task 5 and Task 6. The coding used for heading both increased accuracy and decreased RT. The addition of category and altitude coding (Tasks 5 and 6) did not appear to dilute any benefit gained by the coding of heading into the target symbol (see Figure 13).

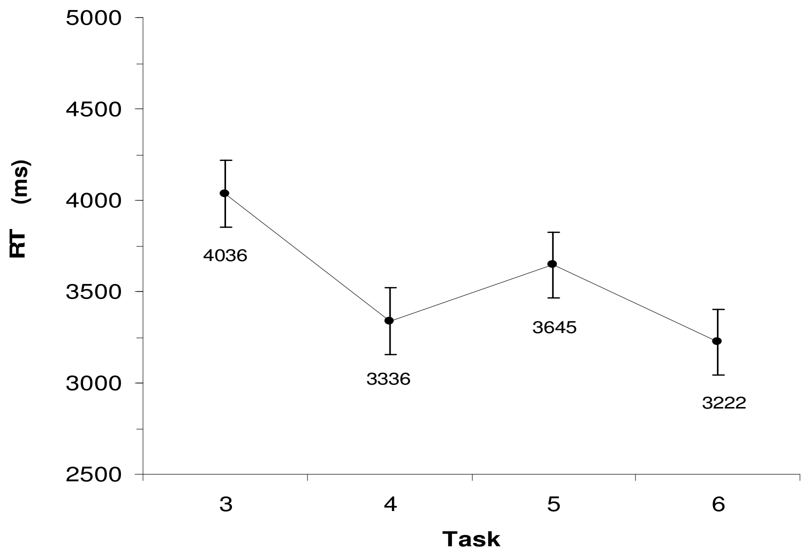


Figure 13. Reaction times (ms) for heading queries by task.

We also divided the heading questions into two subtypes: questions about cardinal directions (i.e., north, south) and questions about noncardinal directions (i.e., northeast, southeast). Contrary to our intuitions, the accuracy for the noncardinal directions was higher than the accuracy for the cardinal directions, $\overline{X}_{diff}$ = .035, $t(36) = 3.602$, $p = .001$, $SE = .010$, 95% CI = [.015, .055]; see Figure 14. However, both types of heading queries benefited similarly from the coding of heading in Task 4 when compared to Task 3, $\overline{X}_{diff\_cardinal}$ = .12, $t_{cardinal}(36) = 4.22$, $p < .001$, $SE = .029$, 95% CI = [.063, .180] and $\overline{X}_{diff\_noncardinal}$ = .051, $t_{noncardinal}(36) = 3.40$, $p = .002$, $SE = .015$, 95% CI = [.020, .081].

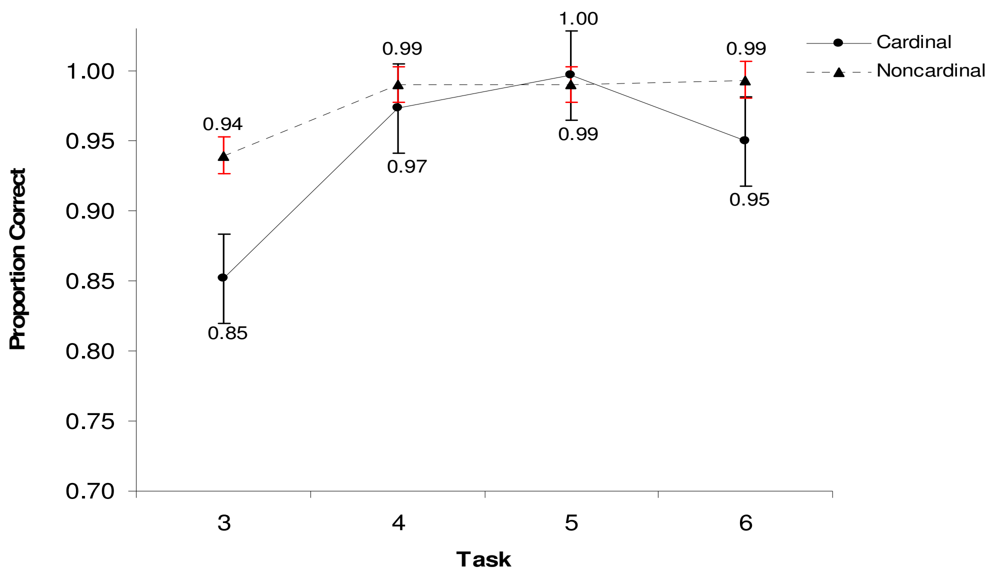


Figure 14. Proportion correct for heading queries by task and direction type.

The mean RT for the cardinal directions was faster than the mean RT for the noncardinal directions, $\overline{X}_{diff}$ = 445.851 ms, $t(36) = 6.36$, $p < .001$, SE = 70.074, 95% CI = [303.735, 587.968]; see Figure 15. However, noncardinal heading queries benefited from the coding of heading when comparing Task 3 to Task 4, $\overline{X}_{diff}$ = 1,096.418 ms, $t_{noncardinal}(36) = 8.19$, $p < .001$, SE = 133.953, 95% CI = [824.748, 1368.088], whereas cardinal heading queries did not.

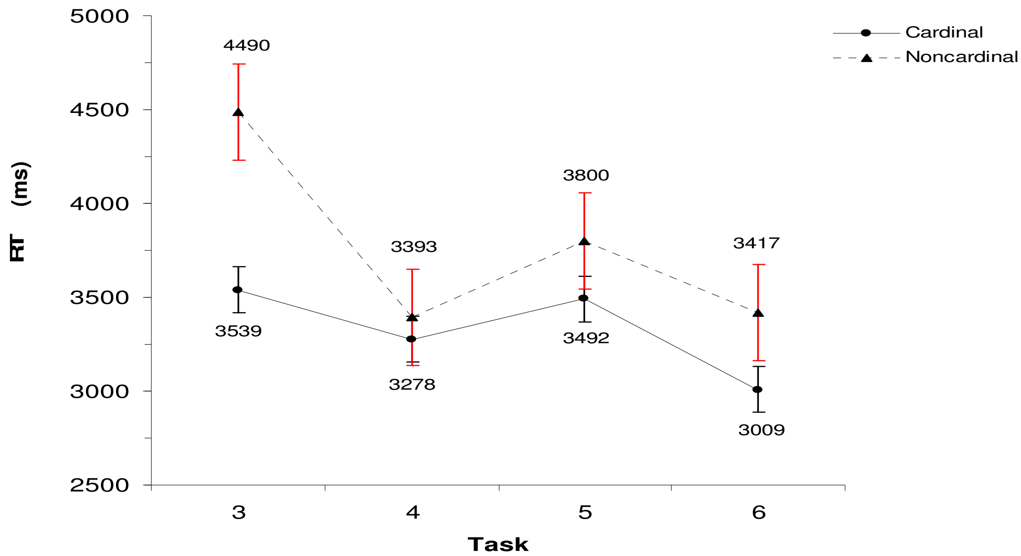


Figure 15. Reaction times (ms) for heading queries by task and direction type.

3.3.2 Aircraft Category Queries

For the queries related to aircraft category, we focused on the difference in accuracy between Task 5 and Tasks 3 and 4. The relevant contrast for this comparison is Task 5 - ½(Task 3 + Task 4). The average increase in proportion correct for Task 5 over that of Tasks 3 and 4 was .098, $t(36) = 4.71$, $p < .001$, 95% CI = [.056, .140], $SE = .021$. The proportion correct for Task 5 was not significantly different from the proportion correct for Task 6 (see Figure 16).

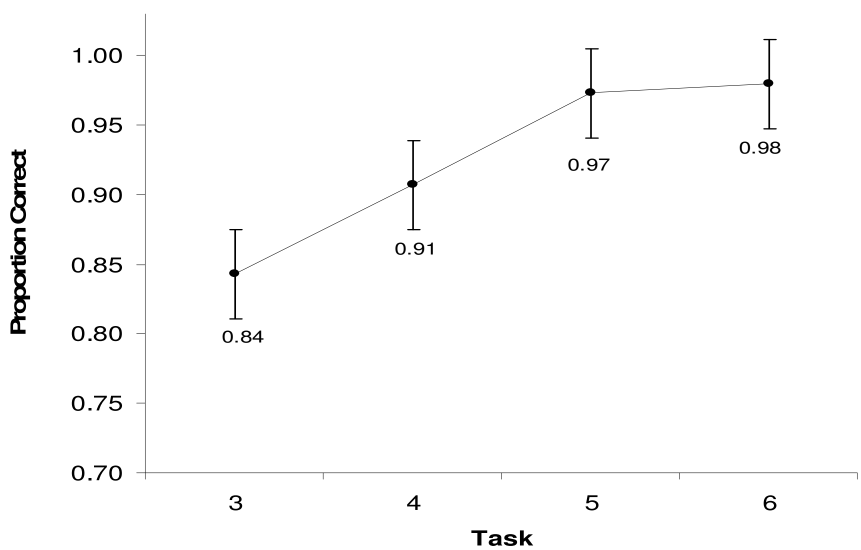


Figure 16. Proportion correct for category queries by task.

The average decrease in RT for Task 5 over the average RT for Tasks 3 and 4 was 820.81 ms, $t(36) = 3.23$, $p = .003$, 95% CI = [306.52, 1335.09], $SE = 253.58$. The average RT for Task 5 was not significantly different from the average RT for Task 6 (see Figure 17). The participants did see benefits from the coding of aircraft category, as shown by both the increase in accuracy and decrease in RT when comparing performance with and without the coding.

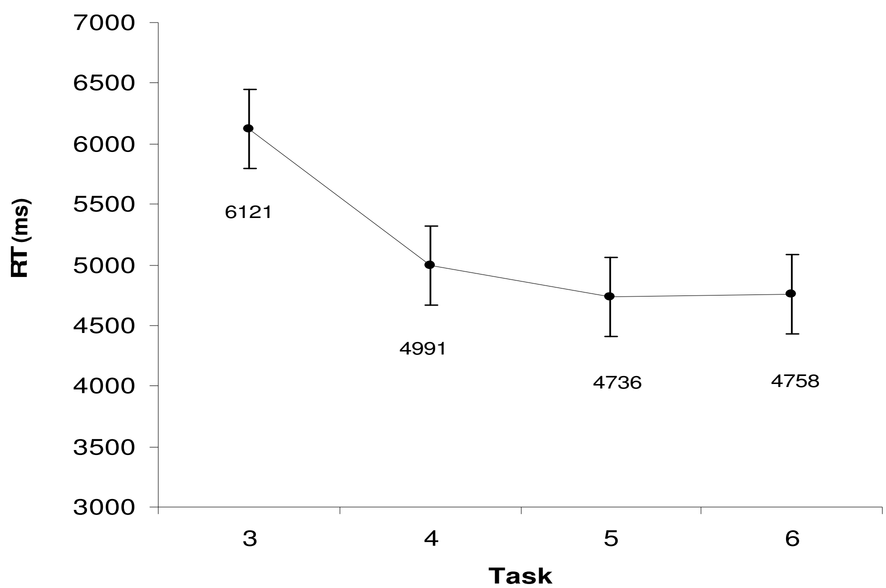


Figure 17. Reaction times (ms) for category queries by task.

For both accuracy and RT, because there appeared to be a difference between Task 3 and Task 4 performance, we also performed post hoc pairwise comparisons using a Bonferroni adjusted *p* value cutoff of .008. We found no significant increase in accuracy between Task 3 and Task 4 performance, but there was a significant decrease in RT even though category coding was not introduced until Task 5. The mean decrease in RT was 1,223.01 ms, $t(35) = 6.92$, $p < .001$, 95% CI = [728.68, 1717.34], $SE = 176.76$. This may be due to a learning effect, or it may be because the display used for Task 4 is easier to search than the Task 3 display. In the Task 3 display, the beacon and target symbols orient towards the radar while the history trails point toward the direction of flight, and this may create a more cluttered looking display that contains visually inconsistent information. However, on the basis of only the findings from this experiment, we cannot tease apart these two possibilities.

To explore the data further, we divided the category questions into two subtypes; those that asked about Large and Heavy aircraft and those that asked about Small and Super Heavy aircraft. We used this distinction because the participants indicated that the coding for Small and Super Heavy aircraft was easier to learn than the coding for Large and Heavy aircraft. Overall, accuracy for questions about Small and Super Heavy aircraft, which was the preferred or more optimal coding, was higher than accuracy for the questions about Large and Heavy aircraft, which used the less preferred or nonoptimal coding (see Figure 18). For the optimal coding, the average increase in proportion correct for Task 5 over that of Tasks 3 and 4 was .073, $t(36) = 4.71$, $p < .001$, 95% CI = [.034, .111], $SE = .019$. For the nonoptimal coding, the average increase in proportion correct for Task 5 over that of Tasks 3 and 4 was .118, $t(36) = 4.79$, $p < .001$, 95% CI = [.068, .168], $SE = .025$. The increase in accuracy for Task 5 over Tasks 3 and 4 for the nonoptimal coding was slightly larger than the increase in accuracy for the optimal coding, $\overline{X}_{difference}$ = .046, $t(36) = 2.290$, p =.028, 95% CI = [.005, .086], SE =.020. For both types of category coding, the proportion correct for Task 5 was not significantly different from the proportion correct for Task 6.

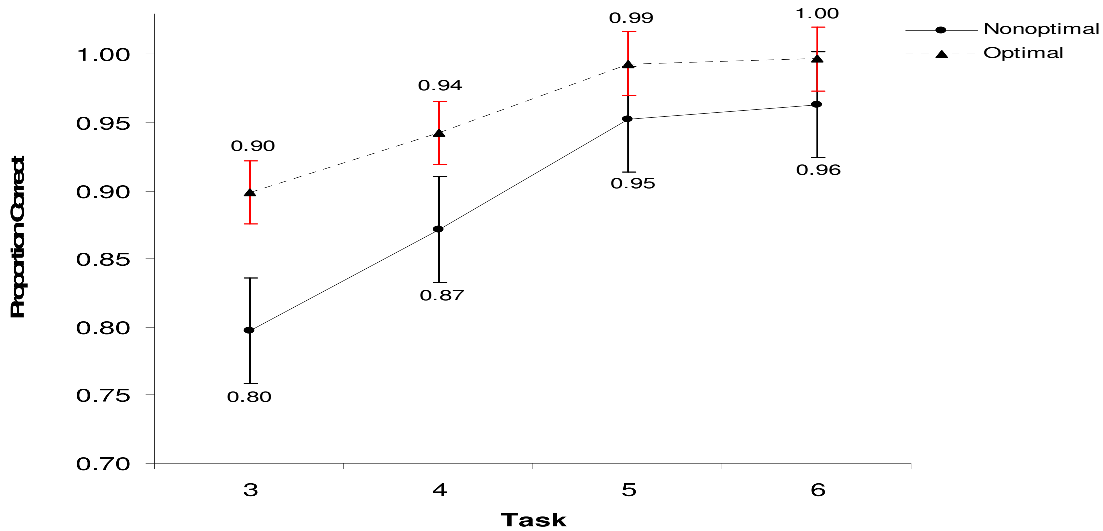


Figure 18. Proportion correct for category queries by task and query type.

The average RT for questions about Small and Super Heavy aircraft, which used the more optimal coding, was faster than the average RT for the questions about Large and Heavy aircraft, which used the nonoptimal coding (see Figure 19). For the optimal coding, the average decrease in RT for Task 5 compared to the RT for Tasks 3 and 4 was 1,077.31 ms, $t(36) = 4.08$, $p < .001$, 95% CI = [541.49, 1613.14], SE = 264.20, but the average RT for Task 5 was not significantly different from the average RT for Task 6. For the nonoptimal coding, the decrease in RT for Task 5 over the RTs of Tasks 3 and 4 was not significant and the average RT for Task 5 was not significantly different from Task 6.

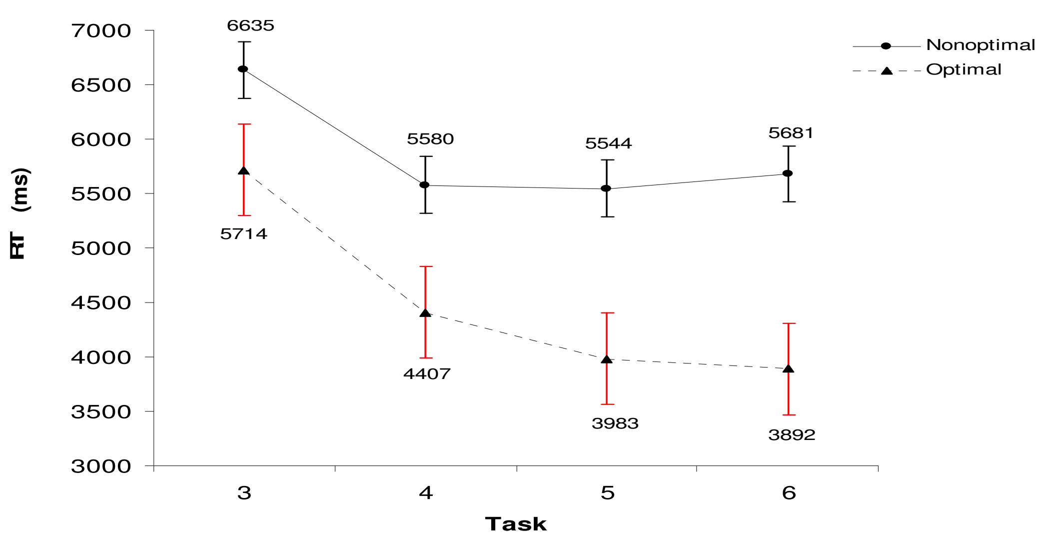


Figure 19. Reaction times (ms) for category queries by task and query type.

Because there appeared to be a difference between Task 3 and Task 4 accuracy and RTs for both the optimal and nonoptimal coding, we performed a post hoc test of these differences using a Bonferroni adjusted critical *p* value of .008. For accuracy, there was no significant difference between Task 3 and Task 4 performance for either the optimal coding or the nonoptimal coding. However, for RTs for both the optimal and the nonoptimal coding, there were significant decreases in RTs when comparing Task 4 to Task 3. The average decrease for optimal coding was 1,306.95 ms, $t(36) = 6.70$, $p < .001$, 95% CI = [911.05, 1702.84], $SE = 195.21$, and the average decrease for nonoptimal coding was 1,075.84 ms, $t(35) = 4.01$, $p < .001$, 95% CI = [531.05, 1620.63], $SE = 268.36$. Again, it is not clear why the addition of heading coding would benefit reaction times for category queries.

3.3.3 Aircraft Altitude Queries

For the altitude queries, we analyzed the difference in accuracy between Task 6, which coded altitude, and Tasks 3, 4, and 5, which did not (see Figure 20). The relevant contrast for this comparison is Task 6 - $^{1}/_{3}$(Task 3 + Task 4 + Task 5). We found an increase in accuracy for altitude queries of .032 when we added altitude coding in Task 6, $t(36) = 2.714$, $p = .010$, 95% CI = [.008, .056], $SE = .012$.

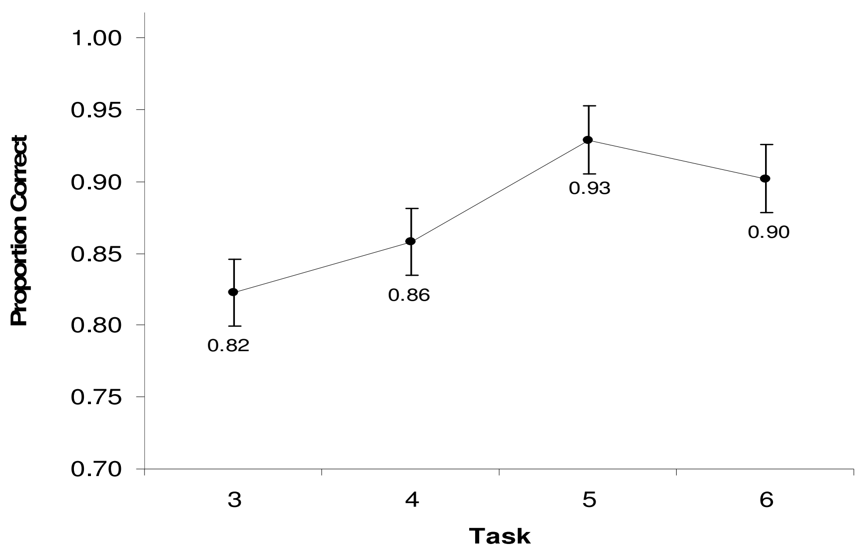


Figure 20. Proportion correct altitude queries by task.

The mean proportion correct for Task 5 was higher than the mean proportion correct for Task 6, so we also performed a post hoc test of the contrast, Task 5 - ½(Task 3 + Task 4), using a Bonferroni adjusted critical *p* value of .008. There was an increase in accuracy for altitude queries of .089 when we added category coding in Task 5, $t(36) = 4.218$ $p < .001$, 95% CI = [.046, .131], SE = .021. There was not a significant difference between the proportion correct in Task 5 and the proportion correct in Task 6. We are not sure why altitude queries would benefit from the addition of category coding given that the aircraft in our images were placed randomly.

For the altitude queries, we also looked at the difference in RTs between Task 6 and Tasks 3, 4, and 5. We saw a decrease in RT for altitude queries when we added altitude coding, with a mean decrease in RT between Task 6 and Tasks 3, 4, and 5 of 642.94 ms, $t(36) = 4.333$, $p < .001$, SE =

148.39, 95% CI = [232.34, 776.34]; see Figure 21. We performed post hoc pairwise comparisons for the RT data using a Bonferroni adjusted critical *p* value of .008, but we found no significant pairwise differences.

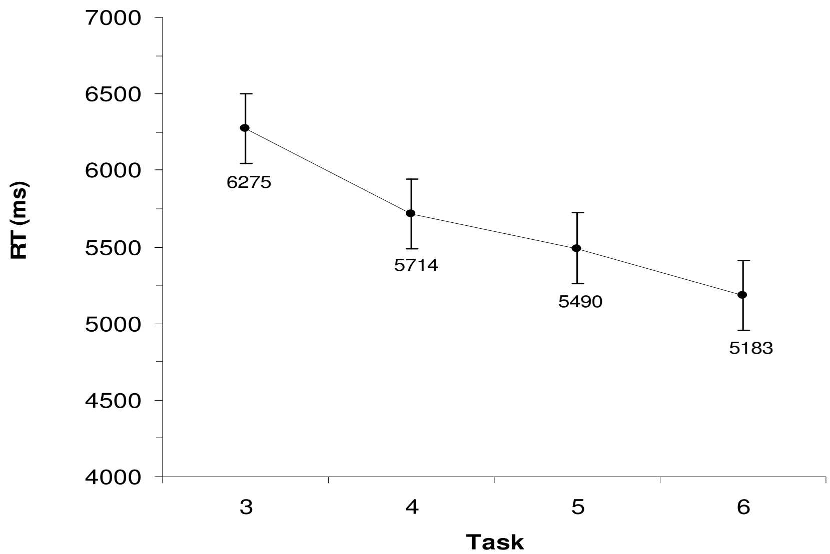


Figure 21. Reaction times (ms) for altitude queries by task.

The mean accuracy for the relative altitude questions was not significantly different than the mean accuracy for the absolute altitude questions (see Figure 22). For the relative altitude queries, there was no effect of altitude coding on accuracy. However, for the absolute altitude queries, there was an effect. The average increase in accuracy for absolute altitude queries when we added altitude coding in Task 6 (compared to the average of Tasks 3 through 5) was .071, $t(36) = 5.499$, $p < .001$, 95% CI = [.045, .097], $SE = .013$.

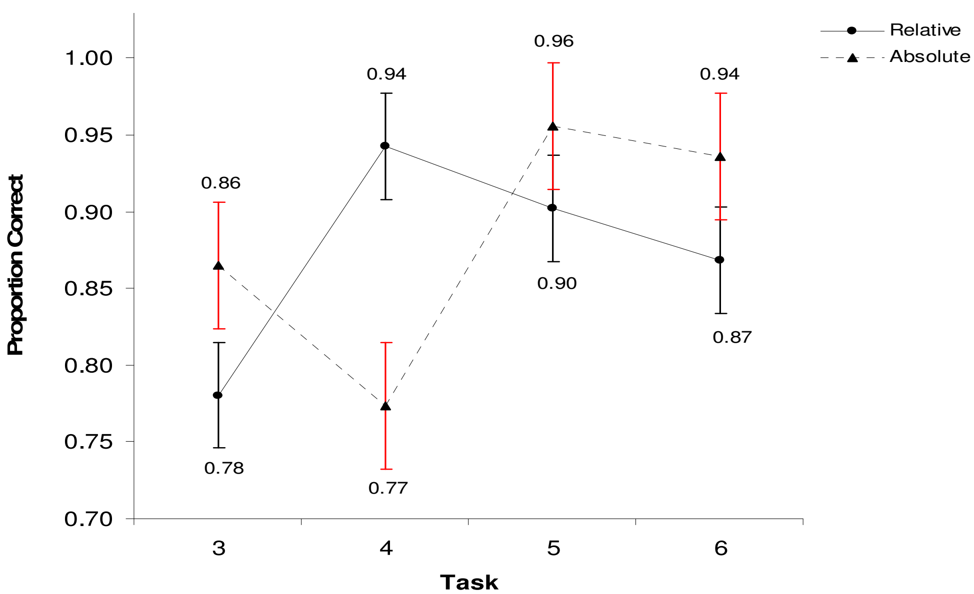


Figure 22. Proportion correct for altitude queries by task and query type.

We also found an effect of the type of query on RTs. The controllers were faster when responding to the absolute altitude queries than when they were responding to the relative altitude queries (see Figure 23). The mean decrease in RT from relative altitude queries to the absolute altitude queries was 1,057.395 ms, *SE* = 173.734, 95% CI = [705.045, 1409.744], $t(36) = 6.086$, $p < .001$. For both types of altitude queries, there was also an effect of altitude coding on RTs. For absolute altitude queries, the average decrease in RT that accompanied the addition of altitude coding was 673.003 ms, $t(36) = 3.761$, $p = .001$, *SE* = 178.925, 95% CI = [310.127, 1035.879]. For relative altitude queries, the average decrease was 558.602 ms, $t(36) = 3.504$, $p = .001$, *SE* = 159.435, 95% CI = [235.254, 881.951].

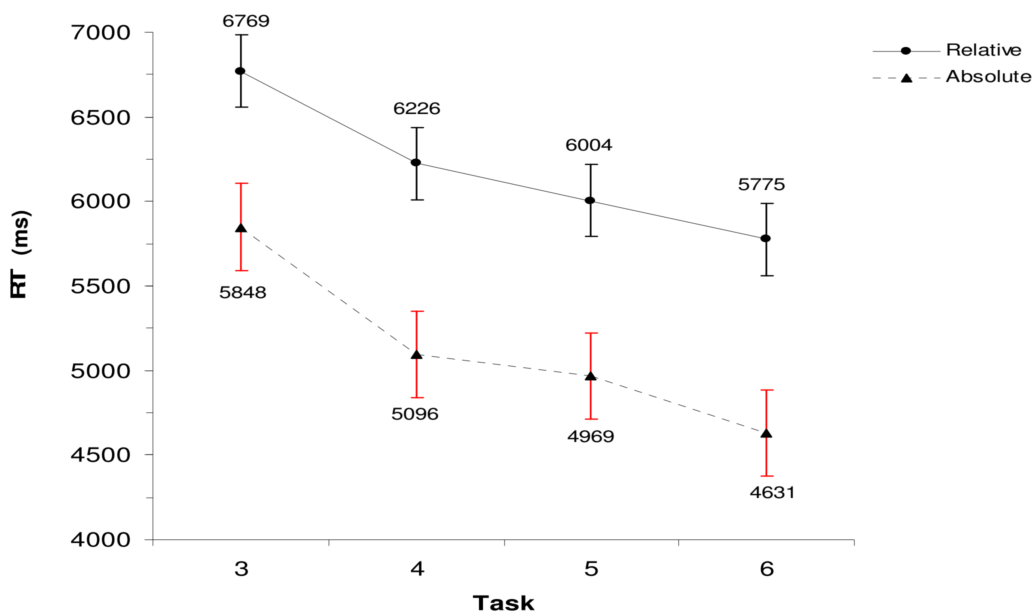


Figure 23. Reaction times (ms) for altitude queries by task and query type.

For the accuracy data, the results of the post hoc pairwise comparisons (using a Bonferroni adjusted critical *p* value of .008) were mixed. For absolute altitude queries, there was a significant decrease in the proportion correct between Task 3 and Task 4 of .091, $t(36) = 4.483$, $p < .001$, 95% CI = [.050, .132], *SE* = .020, and a significant increase between Task 4 and Task 5 of .182, $t(36) = 9.246$, $p < .001$, 95% CI = [.142, .222], *SE* = .020. For relative altitude queries, there was a significant increase in the proportion correct between Task 3 and Task 4 of .162, $t(36) = 4.839$, $p < .001$, 95% CI = [.094, .230], *SE* = .034. For the RT data, only the difference between Task 3 and Task 4 for the absolute altitude queries was significant. The mean decrease between the Task 3 and Task 4 RTs was 752.45 ms, $t(36) = 2.995$, $p = .005$, 95% CI = [242.92, 1261.97], *SE* = 251.23. Future work should try to examine in more detail why altitude queries might benefit from the addition of either heading or category coding.

3.3.4 Aircraft Category + Heading Queries

For the category plus heading queries, we analyzed whether the addition of heading coding (Task 4) and category coding (Task 5) affected accuracy. The two relevant comparisons, which test the impact of coding (a) heading coding and (b) heading and category coding, are Task 4 - Task 3 and Task 5 - ½(Task 3 + Task 4), respectively. The addition of heading coding in Task 4 did not significantly increase accuracy. However, when comparing Tasks 3 and 4 to Task 5, there was an increase in accuracy of .056 when adding both heading and category coding, $t(36) = 2.765$, $p = .009$, 95% CI = [.015, .097], $SE = .020$ (see Figure 24).

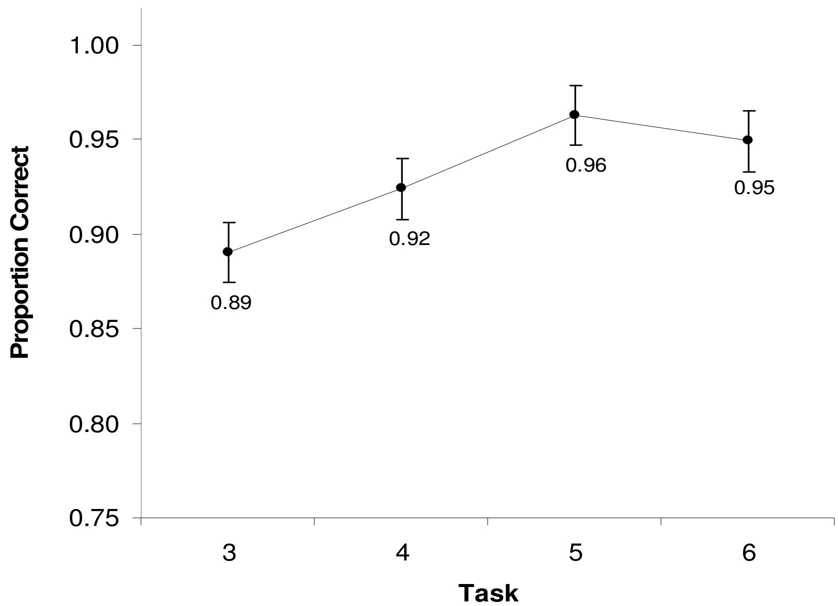


Figure 24. Proportion correct for category + heading queries.

For the category plus heading queries, we also evaluated how the addition of heading coding (Task 4) and category coding (Task 5) affected RTs. The relevant comparisons for the RTs are the same as those for accuracy. Unlike accuracy, the addition of heading coding in Task 4 did result in a decrease in RT when compared to Task 3, $\overline{X}_{difference} = 973.352$ ms, $t(36) = 5.041$, $p < .001$, 95 % CI = [581.791, 1364.912], $SE = 193.068$ (see Figure 25). When comparing Tasks 3 and 4 to Task 5, there was also a significant decrease in RT when adding both heading and category coding, $\overline{X}_{difference} = 640.183$ ms, $t(36) = 3.467$, $p = .001$, 95% CI = [265.648, 1014.718], $SE = 184.673$.

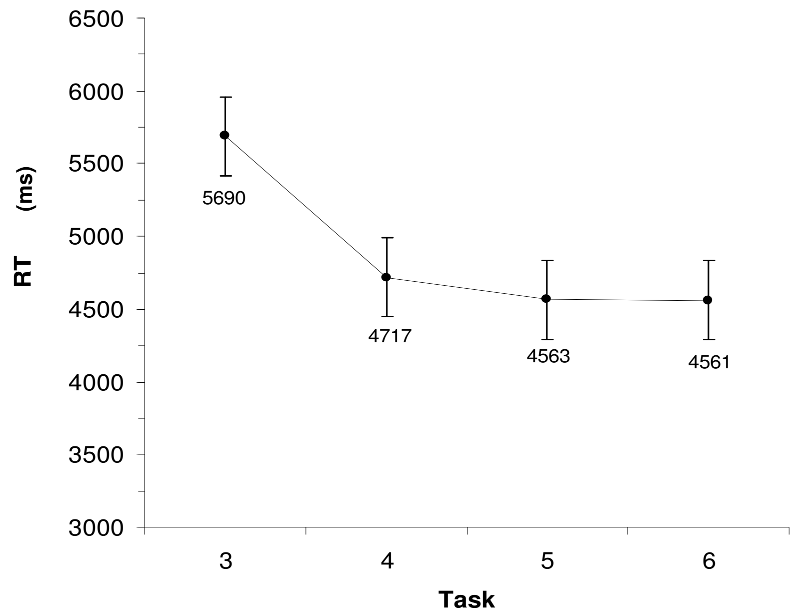


Figure 25. Reaction times (ms) for category + heading queries.

For these analyses, Task 4 - Task 3 and Task 5 - ½(Task 3 + Task 4) are the contrasts that we used to test the impact of coding heading alone and both heading and category. For ease of understanding, we have created two graphs for the data. One graph contains only the data for the optimal coding queries (see Figure 26) and one graph contains the data for the nonoptimal coding queries (see Figure 27).

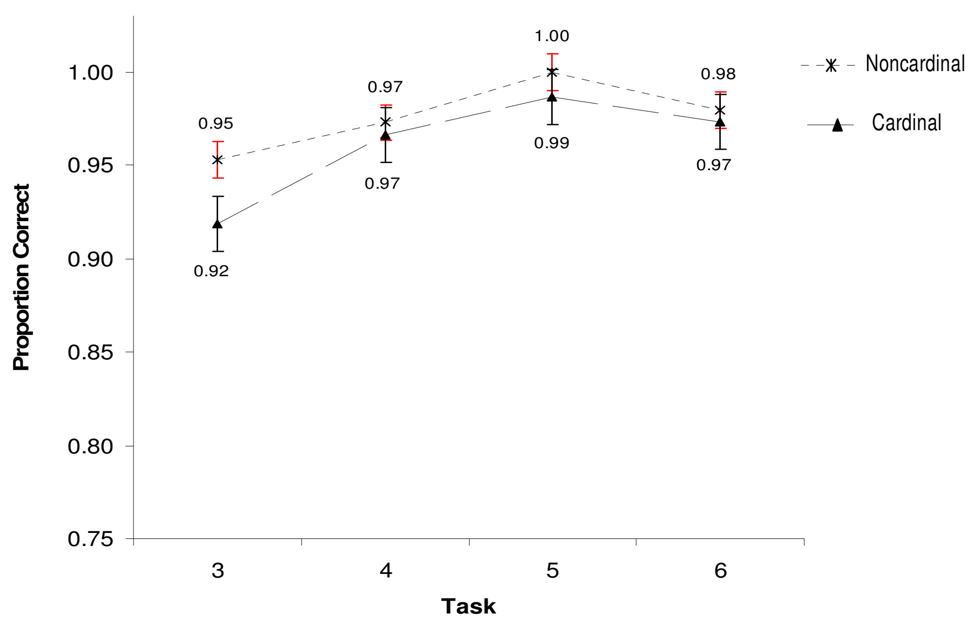


Figure 26. Proportion correct for category + heading queries by task and query type – optimal coding.

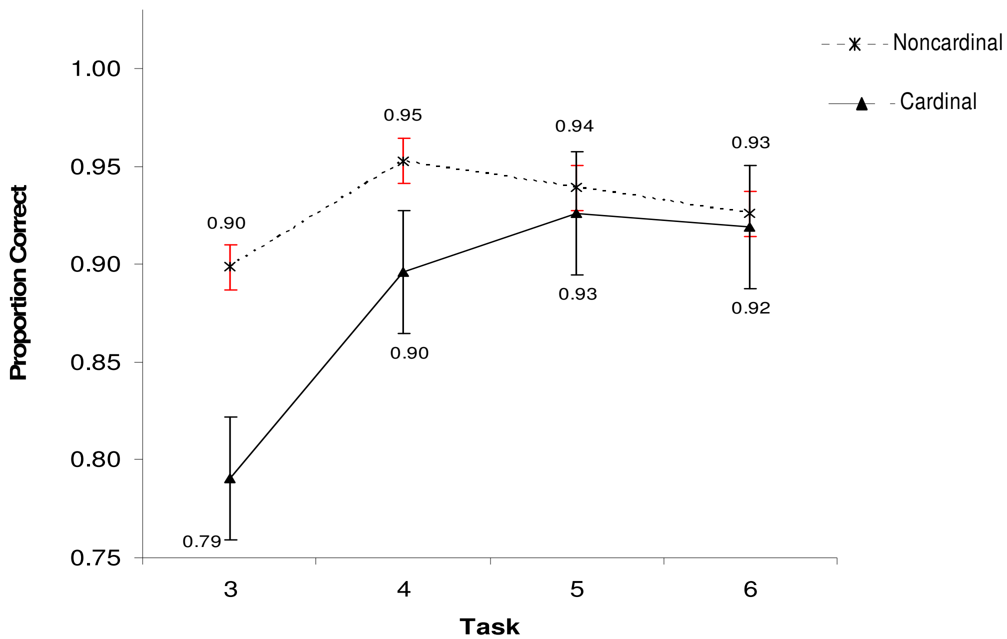


Figure 27. Proportion correct for category + heading queries by task and query type – nonoptimal coding.

For all four query types, there were no significant increases in accuracy when comparing Task 3 to Task 4. However, for every query type, except queries about noncardinal headings and nonoptimal coding, accuracy for Task 5 was greater than the average accuracy for Tasks 3 and 4:

$\overline{X}_{diff\_optimal\_cardinal}$ = .044, $t(36) = 2.252$, $p = .031$, 95 % CI = [.004, .083], $SE = .020$;
$\overline{X}_{diff\_optimal\_noncardinal}$ = .037, $t(36) = 3.479$, $p = .001$, 95 % CI = [.015, .059], $SE = .011$;
$\overline{X}_{diff\_nonoptimal\_cardinal}$ = .095, $t(36) = 2.751$, $p = .009$, 95 % CI = [.025, .164], $SE = .034$.

The relevant comparisons for the reaction time data are the same as those for the accuracy data. We compared RTs for Task 4 to Task 3 to test the impact of heading coding on RTs. We also compared the RTs for Task 5 to the average of the RTs for Tasks 3 and 4. This comparison tested the impact of coding both heading and category on RTs. Figure 28 contains the RT data for the optimal coding queries, and Figure 29 contains the RT data for the nonoptimal coding queries.

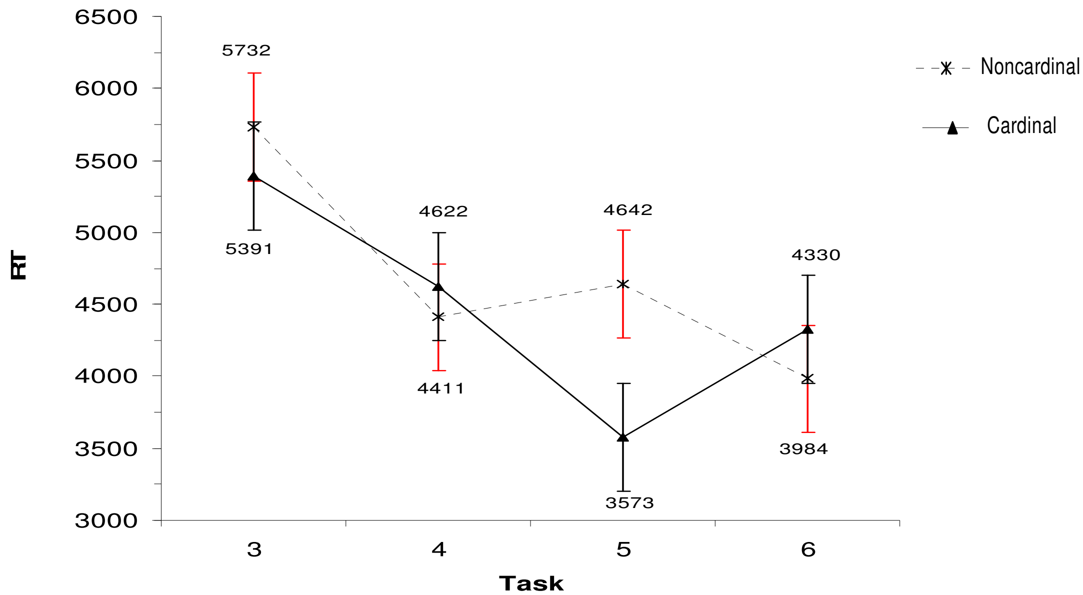


Figure 28. Reaction times (ms) for category + heading queries by task and query type – optimal coding.

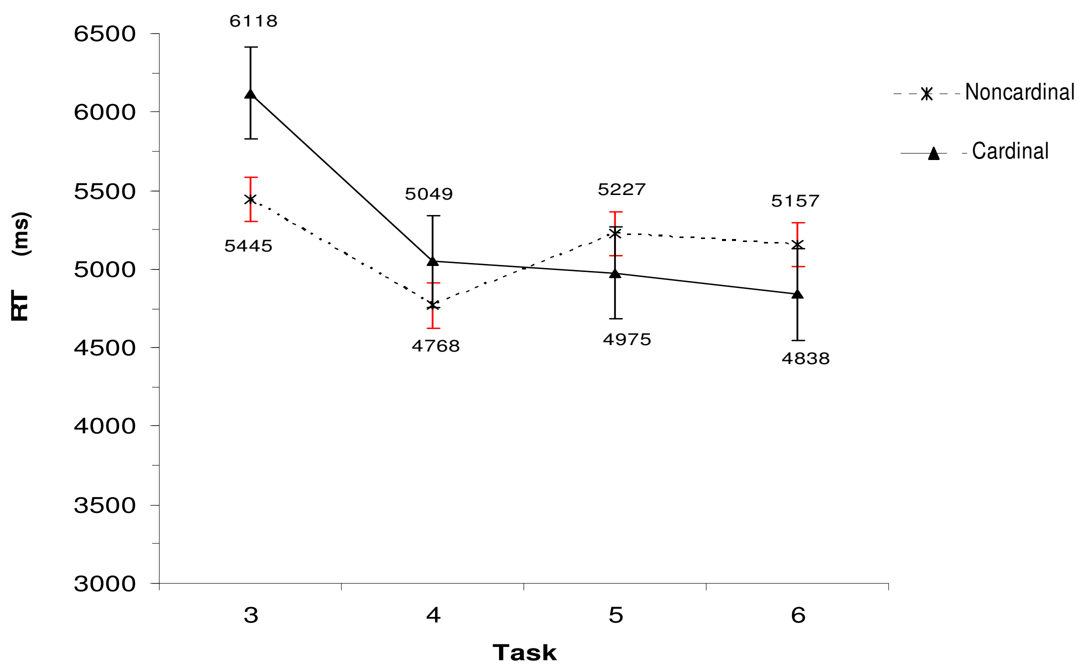


Figure 29. Reaction times (ms) for category + heading queries by task and query type – nonoptimal coding.

For three of the query types, there was a significant decrease in RT from Task 3 to Task 4: $\overline{X}_{diff\_optimal\_cardinal}$ = 769.149 ms, $t(36) = 2.499$, $p = .017$, 95 % CI = [144.954, 1393.343], SE = 307.774; $\overline{X}_{diff\_optimal\_noncardinal}$ = 1,321.381 ms, $t(36) = 4.743$, $p < .001$, 95 % CI = [756.312, 1886.450], SE = 278.621; $\overline{X}_{diff\_nonoptimal\_cardinal}$ = 947.023 ms, $t(36) = 2.460$, $p = .019$, 95 % CI = [165.646, 1728.400], SE = 384.894.

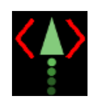
However, the RT for Task 5 was shorter than the average RT for Tasks 3 and 4 only for the query about cardinal headings and optimal coding, $\overline{X}_{diff\_optimal\_cardinal}$ = 1,433.122 ms, $t(36)$ = 6.548, $p < .001$, 95 % CI = [989.266, 1876.977], SE = 218.853.

3.4 Stereotype Symbol Preferences for Aircraft Conformance: Task 7 Results

We could not find any work that has evaluated aircraft conformance symbology. Aircraft conformance is relevant in the terminal and TRACON because as we try to create a more efficient NAS, aircraft may need to fly more tightly spaced, precision approaches. As part of NextGen, it will be critical for the system and for the controllers to monitor aircraft closely for conformance with their prescribed routes because any deviation from that route may create a safety hazard.

For the four aircraft conformance categories, there was substantial agreement among the controllers as to which symbol belonged with which definition of conformance (see Table 5). The green triangle without the brackets was assigned to the *Not Monitored* category by 76% of the controllers. The green triangle with the green brackets was assigned to the *In Conformance* category by 70% of the controllers. The green triangle with the yellow brackets was assigned to the *Out of Conformance-Projected* category by 97% of the controllers, and the green triangle with the red brackets was assigned to the *Out of Conformance-Actual* by 92% of the controllers.

Table 5. Proportion of Conformance Assignment by Symbol Type

Conformance				
Not Monitored	.76	.24	0	0
In Conformance	.24	.70	0	.5
Out of Conformance-Projected	0	0	.97	.3
Out of Conformance-Actual	0	.5	.3	.92

The controllers also provided goodness ratings for their coding schemes on a Likert scale of 1 to 7, with 1 indicating the *worst possible scheme* and 7 indicating the *best possible scheme*. The average rating was 5.41, SE = .25. Most of the ratings were 4 or higher (see Figure 30).

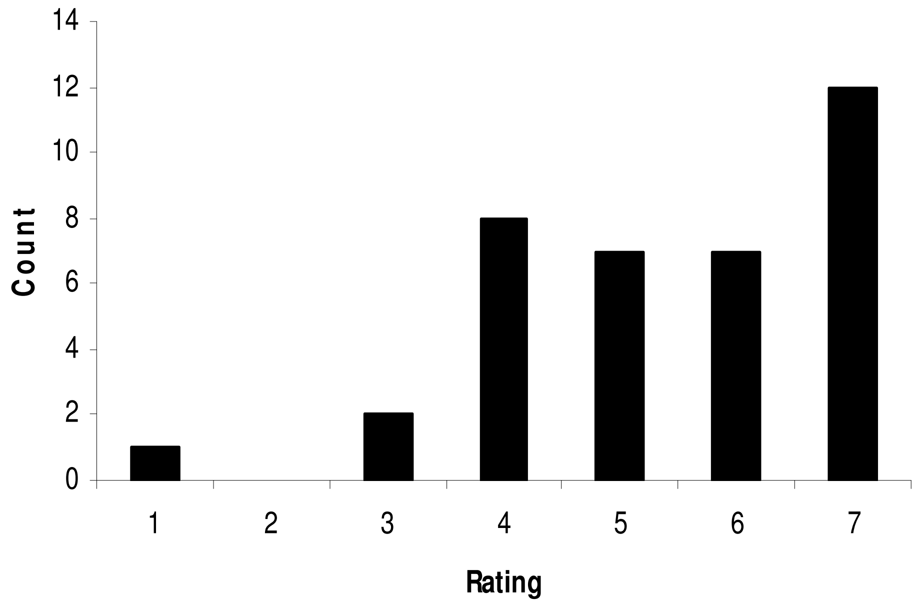


Figure 30. Frequency distribution of goodness ratings for Task 7.

There was a small but measurable degree of disagreement among the controllers about the assignments for the green triangle with the green brackets and the green triangle without the green brackets. Even though 76% of the participants assigned the green triangle without brackets to the Not Monitored category, 24% of the participants assigned the green triangle with the brackets to the In Conformance category. We speculate that the controllers who assigned the triangle with brackets to the In Conformance category may have been thinking about the default state of the system in the future NAS rather than the default state currently. In the NAS of the future, all aircraft may be monitored for conformance, therefore, In Conformance would be its default state. In this case, it might make sense for the most basic symbol (i.e., the triangle without brackets) to be the default symbol, that is, the symbol for In Conformance. However, as we transition to NextGen, most aircraft will not be monitored. During this transition phase, the default aircraft state would be Not Monitored. Therefore, we believe it would make more sense for the basic symbol to be assigned to the Not Monitored category.

From a human factors standpoint, we believe it makes more sense to design the symbology for the current state of the system rather than for the future state of the system. If we design a symbol to account for how things may be in the future, we may make life simpler for the future controller. However, we run the risk of creating symbols that are unintuitive or, worse yet, counterintuitive and difficult to learn given the controllers current mental model.

3.5 Exit Questionnaire Results

The controllers rated the impact of coding heading, aircraft category, and altitude information on their ability to control traffic on a scale ranging from 1 (*very negative*) to 7 (*very positive*). The average ratings were fairly neutral, in the range of 3.68 to 4.84 (see Table 6), but there were more positive ratings for Heading Coding, Aircraft Category Coding, and Conformance Coding than there were for Altitude Coding (see Figure 31).

Table 6. Mean Rating of Coding Types

	Coding Type			
	Heading	Category	Altitude	Conformance
Mean	4.84	4.62	3.68	4.46
SD	1.25	1.52	1.62	1.52

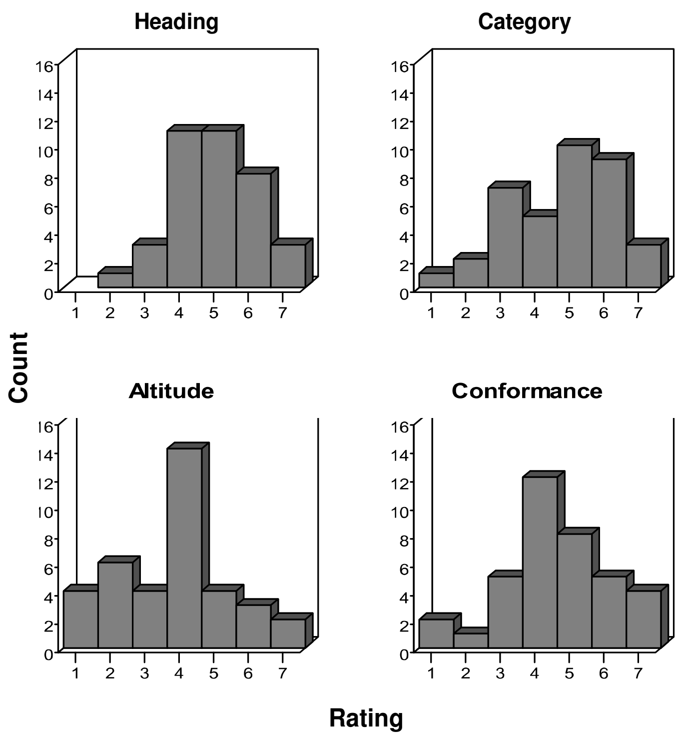


Figure 31. Frequency graphs of ratings by coding type.

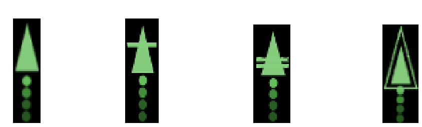
Although some controllers reported that they preferred using history trails, more than half of the participants reported that using triangles to code direction was helpful. They reported that the triangle coding was easy to recognize and understand and also helped them to determine direction of flight. This is consistent with work by Olson, Kaliardos, Zuschlag, and Kendra (2009) who found no negative impact of coding heading on pilot responses to Traffic Alert and Collision Avoidance System II (TCAS II) traffic advisory and resolution advisory events.

In some cases, we demonstrated improvement from Task 3 to Task 4 for no discernible reason. On the basis of these findings, it seems reasonable to conclude that the STARS target symbology may be presenting heading information to controllers in a suboptimal manner. On a static STARS display, the sole source for heading information is history trails. Therefore, it makes sense that providing additional and consistent heading information would lead to an improvement in the accuracy and RT of controllers. Interestingly, it also seems that when we provide consistent heading information embedded into the target symbol, we not only see performance improvements in searches for heading information but we also see performance improvements in searches for category and altitude information.

The controllers provided mixed feedback for the aircraft category coding. Many controllers reported that it was helpful and that it enabled them to quickly identify Small and Super Heavy aircraft symbols. However, a number of controllers reported problems with learning the Large and Heavy aircraft symbols. Both the Large and Heavy symbols consisted of a triangle with a single line; the only distinguishing feature was the location of the line at either the tip or the rear of the triangle. Differentiating Large aircraft from Heavy aircraft, therefore, appeared to require too much effortful processing by the controller. We did not anticipate that this similarity would cause such difficulties for the controllers. The data we collected in Task 2 and the feedback we collected in the visual search task also indicated that the controllers might have difficulty learning to differentiate the Large and Heavy symbols. We believe that for the symbols to be optimally beneficial, we would need to make them more perceptually distinct.

Although we selected the symbols for the visual search task prior to obtaining controller feedback, it is interesting that we selected symbols for the Small and Super Heavy that were consistent with the coding suggested by the heuristics used by the controllers in Tasks 1 and 2. The Small symbol was a simple triangle and the Super Heavy symbol was a nonbasic shape (i.e., a triangle with two lines at the back).

On the basis of our findings, we have created a revised set of symbols that is consistent with the findings from Tasks 1 and 2 (see Figure 32). We believe this scheme or some similar scheme designed using the heuristics identified in Tasks 1 and 2 would provide the benefits of coding aircraft category without creating the confusion that our original symbol set created.



a b c d

Figure 32. Redesigned symbols for size categories (a) small, (b) large, (c) heavy, and (d) super heavy.

Many controllers reported that they did not find the altitude coding to be very helpful, and more than half reported that it was confusing or distracting. They reported that the differences between the colors used for coding were not obvious, and they did not feel that they received any concrete performance benefits from the coding. In addition, they did not like the coding because it did not convey absolute altitude.

4. DISCUSSION AND CONCLUSIONS

We believe the information in this report will provide valuable information to the Human Factors Research and Engineering Group and Terminal Program Office, which they can use when developing specifications and guidelines regarding symbol selection for future ATC terminal displays.

The overall feedback from the controllers indicated that we should not clutter the radar display, the target, or the datablock with too much coding as we design new ATC displays. Although they indicated that the coding of direction of flight and aircraft category was acceptable, they believe that beyond that, too much unnecessary or difficult-to-process data can distract them from their primary task of separating aircraft.

We found benefits in terms of both increased accuracy and decreased reaction times for both heading and category coding. However, for the category coding the overall accuracy and reaction times were better for the optimal than the nonoptimal coding. Our reaction time and accuracy data are consistent with the findings of Smallman et al. (2001a, 2001b). By using meaningful symbols to convey relevant tactical information, such as aircraft category and heading to the controllers, we can both increase visual search speed and increase target detection accuracy.

We also found that altitude coding, as implemented in our experiment, was not beneficial for controllers. It did not have a consistent impact on either accuracy or reaction time. It may be beneficial to explore other ways to code altitude because of its tactical relevance for terminal controllers.

Our results indicate that the controllers use consistent heuristics for categorizing symbols into different size categories, and designers need to take advantage of this. We believe for a more thorough understanding of the heuristics used for the bullet shaped symbols, it would be useful to test the entire set of 24 symbols from Task 1 in Task 2. We recommend that designers adhere to the following design principles. We believe that to code the size category Small into an aircraft target symbol, designers should use a simple, basic shape. To code intermediate categories, (e.g., Large and Heavy) designers should use the basic shape as the base form and then add something distinctive to the basic shape (e.g., lines, holes). To code the Super Heavy category, designers should use the basic shape as the base form and then add an outline or other coding to indicate that it is the most extreme or largest category.

By using a consistent basic shape, system users will have the ability to quickly identify that a group of symbols all belong to the same set or category. If there is more than one intermediate level, the designers must ensure that they are perceptually distinct to avoid confusion on the part of the controller.

Our findings indicate that there was strong agreement among the controllers when assigning the symbols to the four aircraft conformance categories. Feedback from the controllers indicated (a) that these symbols were easy to learn and (b) that the colors used to indicate the status level of an aircraft matched standard color coding conventions. We believe that as conformance monitoring becomes more tactically relevant in the NextGen timeframe and as tailored arrivals become commonplace, simple and intuitive coding for information like route or altitude conformance will become a requirement.

There are clear informational hurdles that designers will need to deal with as we move towards NextGen. Although much more information will be available to present to the controllers, only a small portion of this information should be displayed on the controller's tactical display if we want to avoid information overload. With multiple NextGen tools (such as ADS-B and DataComm) competing for the limited datablock real estate, it makes sense to move the most tactically relevant ATC information from the datablock into the target symbol.

We believe that one of the most critical tasks ATC researchers will have in the next 10 years will be determining (a) which of all potentially useful information types are the most tactically relevant for the controllers and should be awarded entry into the highly prized real estate of the target symbol and (b) what is the most optimal way to code this information.

When creating target symbols that code tactically relevant information, system designers must ensure that they are easy to interpret and that they take advantage of both perceptually salient and culturally learned coding conventions. Our assessment is a starting point that designers can use when making decisions about how to create new symbols for future ATC terminal displays and how to prioritize the information to be coded in these symbols. Furthermore, we have included a set of recommendations that designers and researchers should refer to when developing and testing new candidate symbols for terminal ATC displays (see Appendix F). It should also help system designers identify relevant situational factors to use when creating and selecting optimally designed symbols not only for aviation-related displays but also for displays in other safety-critical domains.

References

Ahlstrom, V., & Longo, K. (Eds.). (2003). *Human factors design standard for acquisition of commercial-off-the-shelf subsystems, non-developmental items, and developmental systems* (DOT/FAA/CT-03/05/HF-STD-001). Atlantic City International Airport, NJ: Federal Aviation Administration, William J. Hughes Technical Center.

American National Standards Institute. (1988). *American National Standard for human factors engineering of visual display terminal workstations* (ANSI/HFS 100-1988). New York: ANSI.

Bravo, M. J., & Farid, H. (2008). A scale invariant measure of image clutter. *Journal of Vision, 8*, 1-9.

Cardosi, K., & Hannon, D. (1999). *Guidelines for the use of color in ATC displays* (DOT/FAA/AR-99/52, DOT-VNTSC-FAA-98-5). Washington, DC: U.S. Department of Transportation, Office of Aviation Research, Federal Aviation Administration.

Cardosi, K. M., & Murphy, E. D. (Eds.). (1995). *Human factors in the design and evaluation of air traffic control systems* (DOT/FAA/RD-95-3, DOT-VNTSC-FAA-95-3). Washington, DC: U.S. Department of Transportation, Office of Aviation Research, Federal Aviation Administration.

Carney, C., Campbell, J. L., & Mitchell, E. A. (1998). *In-vehicle display icons and other information elements. Task A: Literature review* (FHWA-RD-98-164). Washington, DC: U.S. Department of Transportation, Federal Highway Administration.

Chandra, D. C., & Yeh, M. (2007). *Pilot identification of symbols and an exploration of symbol design issues for electronic displays of aeronautical charting information* (DOT/FAA/AR-07/37, DOT-VNTSC-FAA-07-07). Cambridge, MA: U.S. DOT Volpe National Transportation Systems Center.

Chandra, D. C., Zuschlag, M., Helleberg, J., & Estes, S. (2009). Symbols for cockpit displays of traffic information. In *Proceedings of the 28th Digital Avionics Systems Conference* (pp. 25-29). Orlando, FL.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology, 1,* 75-78.

Delucia, P. R. (1995). Effects of pictorial relative size and ground-intercept information on judgments about potential collisions in perspective displays. *Human Factors, 37,* 528-538.

Federal Aviation Administration. (2008). *Federal Aviation Administration Flight Plan 2009–2013.* Washington, DC.

Friedman-Berg, F., Allendoerfer, K., & Deshmukh, A. (2010). Developing target symbicons for the future terminal air traffic control environment. In *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting,* Santa Monica, CA.

Friedman-Berg, F., Allendoerfer, K., & Pai, S. (2009). *Moving toward an air traffic control display standard: Creating a standardized color palette for terminal situation displays* (DOT/FAA/ TC-08/15). Atlantic City International Airport, NJ: FAA William J. Hughes Technical Center.

Green, P., & Pew, R. W. (1978). Evaluating pictographic symbols: An automotive application. *Human Factors, 20,* 103-114.

International Civil Aviation Organization. (2001). *Aeronautical charts: Annex 4 to the Convention on International Civil Aviation* (10th ed.).

Johnston, J., Horlitz, K., & Edmiston, W. (1993). Improving situation awareness displays for air traffic controllers. *Proceedings of the Seventh International Symposium on Aviation Psychology,* Columbus, OH.

Lindberg, T., & Näsänen, R. (2003). The effect of icon spacing and size on the speed of icon processing. *Displays, 24*, 111-120.

Mackett-Stout, J., & Dewar, R. (1981). Evaluation of symbolic public information signs. *Human Factors*, *23*, 139-151.

McFadden, S. M., Jeon, J., Li, A., & Minniti, A. (2008). *Evaluation of symbol sets for naval tactical displays* (DRDC-TORONTO-TR-2007-046). Waterloo ONT (CAN): Defense R&D Canada - Toronto, Toronto ONT (CAN): Waterloo University.

Military Standard 1787C. (2001). *Aircraft display symbology* (MIL-STD-1787C). Washington, DC: Department of Defense.

Military Standard 2525B. (2005). *Common warfighting symbology* (MIL-STD-2525B). Washington, DC: Department of Defense.

Military Standard 2525C. (2008). *Common warfighting symbology* (MIL-STD-2525C). Washington, DC: Department of Defense.

North Atlantic Treaty Organization. (1990). *North Atlantic Treaty Organization draft STANdardization AGreement: Display symbology and colors for NATO maritime units.* (STANAG-4420) Brussels, Belgium.

North Atlantic Treaty Organization. (2005). *North Atlantic Treaty Organization draft STANdardization AGreement: Aircraft display symbology* (STANAG-7161 ED.1). Brussels, Belgium.

Nunes, A., & Mogford, R. (2003). Identifying control strategies that support the picture. *Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society*, 71-75.

Olson, W., Kaliardos, B., Zuschlag, M., & Kendra, A. (2009). Impact of traffic symbol directional cues on pilot performance during TCAS events. *In Proceedings of the 28th Digital Avionics Systems Conference* (pp. 25-29). Orlando, FL.

Palmer, E. M., Clausner, T. C., & Kellman, P. J. (2008). Enhancing air traffic displays via perceptual cues. *ACM Transactions on Applied Perception*, *5,* 1-22.

Radio Technical Commission for Aeronautics, Inc. (2003). *Minimum operational performance standards for the depiction of navigational information on electronic maps* (DO-257A). Washington, DC: RTCA.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114,* 510-532.

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology, 4,* 328-350.

Smallman, H. S., St. John, M., Oonk, H. M., & Cowen, M. B. (2001a). Information availability in 2D and 3D displays. *IEEE Computer Graphics and Applications, 21*, 51-57.

Smallman, H. S., St. John, M., Oonk, H. M., & Cowen, M. B. (2001b). 'SYMBICONS': A hybrid symbology that combines the best elements of SYMBols and ICONS. *Proceedings of the 45th Annual Meeting of the Human Factors and Ergonomics Society*, 110-114.

Snodgrass, J. G., & McCullough, B. (1986). The role of visual similarity in picture categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 12*, 147-154.

Society of Automotive Engineers. (1997). *Electronic aeronautical symbols* (ARP 5289). Warrendale, PA: SAE.

Tam, L., Corker, K. M., & Duong, V. (2003). Information support for air traffic management in the SuperSector: Color-saturation coding of aircraft altitude and effects on air traffic controller performance. *Fifth Annual USA/Europe ATM R&D seminar.*

Thompson, G. L. (2006). An SPSS implementation of the non-recursive outlier deletion Procedure with shifting *z*-score criterion. *Behavior Research Methods*, *38*, 344-352.

Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size and skew on outlier elimination. *Quarterly Journal of Experimental Psychology, 47A,* 631-650.

Wolfe, J. M. (1998). What can 1,000,000 trials tell us about visual search? *Psychological Science, 9*.

Yeh, M., & Chandra, D. C. (2004). Issues in symbol design for electronic displays of navigation information. *Proceedings of the 23rd Digital Avionics Systems Conference* (pp. 24-28). Salt Lake City, UT.

Yeh, M., & Chandra, D. C. (2005). *Designing and evaluating symbols for electronic displays of navigation information: Symbol stereotypes and symbol-feature rules* (DOT/FAA/AR-05/48, DOT-VNTSC-FAA-05-16). Cambridge, MA: U.S. DOT Volpe National Transportation Systems Center.

Yeh, M., & Chandra, D. C. (2006). Pilot stereotypes for navigation symbols on electronic displays. In R. Reuzeau, K. Corker, & G. Boy (Eds.), *Proceedings of the International Conference on Human-Computer Interaction in Aeronautics* (pp. 66-73). Toulouse, France: Cépaduès-Éditions.

Acronyms

ADS-B	Automatic Dependent Surveillance-Broadcast
ARTS	Automated Radar Terminal System
ATC	Air Traffic Control
ATO	Air Traffic Organization
CARTS	Common Automated Radar Terminal System
CPC	Certified Professional Controller
Data Comm	Data Communications
FAA	Federal Aviation Administration
FLM	Front Line Manager
GPS	Global Positioning System
NAS	National Airspace System
NextGen	Next Generation Air Transportation System
NNEW	NextGen Network Enabled Weather
NVS	National Airspace System Voice Switch
RT	Reaction Time
STARS	Standard Terminal Automation Replacement System
SWIM	System-Wide Information Management
TCAS	Traffic Alert and Collision Avoidance System
TMU	Traffic Management Unit
TRACON	Terminal Radar Approach Control

Appendix A

Statement of Ethics and Informed Consent

Target Symbology Study

Statement of Ethics & Informed Consent

Nature and Purpose of Activity

Thank you for volunteering to participate in the Target Symbology Study. The study has 4 primary objectives: 1) investigate whether different coding schemes for aircraft category are intuitive, 2) investigate whether coding aircraft symbols by aircraft category aids controller performance, 3) investigate whether color coding for altitude aids the controllers, and 4) investigate whether using directional symbols helps the controllers identify an aircraft's direction of flight. We believe the results of this study will help system developers design target symbols that aid controller performance.

Who is Eligible to Participate?

Any FAA Air Traffic Terminal Controller who is participating in another Human Factors study conducted at the William J. Hughes Technical Center Research and Development Human Factors Laboratory facility or field personnel.

Experimental Procedures

We will explain the procedures and your rights and responsibilities before you begin. The session will last approximately 60 minutes. At the beginning of the study, we will ask you to complete a background questionnaire. You will perform four types of tasks. During the first few tasks, we will ask you to identify the aircraft symbol that best represents each aircraft category. For the other types of tasks, we will measure how quickly and accurately you can identify novel target symbols on the basis of certain criteria. We may ask you to identify aircraft category, heading, altitude or some combination of these. When you are finished, we would like you to provide us with feedback on the novel symbol coding. A researcher will be available to answer any questions you may have.

Discomforts and Risks

We anticipate no discomfort or risk in performing the experiment. In this study, you will view a simulated STARS display and will identify specific aircraft.

Benefits

Participation in this study provides no direct benefit to you. Your participation benefits the FAA and its workforce. The FAA expects to use this data to aid in the development of aircraft symbology that may be used on future terminal systems.

Participant's Responsibilities

The results of this effort depend greatly on your attention to the task and on your honest responses to the questionnaire. If there is something you do not understand, please ask a researcher. In addition, to avoid biasing the results, please do not discuss the study with other potential participants until the study is completed.

Participant's Assurances

Researchers from the Human Factors Team–Atlantic City maintain strict standards regarding participant confidentiality and informed consent. We base our standards on the *Ethical Principles in the Conduct of Research with Human Participants* by the American Psychological Association (APA) and FAA Order 9500/25. The APA and FAA standards conform to the following principles.

- Your participation in this study is voluntary. You may withdraw from this assessment at any time, for any reason, without consequence. If you feel you must withdraw, please inform a researcher immediately. There is no penalty whatsoever for ending the session early. In addition, the researchers may terminate your participation if they believe this to be in your best interest.
- Your responsibilities will be clear. The researchers will explain all procedures thoroughly, and will answer all questions.
- We will keep your identity anonymous. We will identify your responses using a code known only to you and the researchers. We will keep your identity separate from your data. We will not associate your name with data contained in any report or briefing. To facilitate this, please do not write your name or any other identifying marks on your questionnaires. Please do not share your participant code with anyone other than the researchers.
- We will keep your data confidential. The *raw* data collected in this assessment will become the property of the Human Factors Team–Atlantic City. Specialists from this organization and its contractor employees will analyze the data. We will not make any individual data available to other organizations without your permission. We will only present *aggregate* statistics in briefings and reports. These will include averages, standard deviations, and other summary statistics.

If you any have questions about this study or need to report any adverse conditions, you may contact the researchers conducting the study, the Primary Investigators Dr. Ferne Friedman-Berg or Kenneth Allendoerfer. You may also contact Dr. Earl Stein (609) 485-6389, the Human Factors Team – Atlantic City Manager, at any time with questions or concerns.

I have read this consent document. I understand its contents, and I freely consent to participate in this study under the conditions described.

Participant:	(print) ______________________________		
	(sign) ______________________________	Date: __________	
Investigator:	(sign) ______________________________	Date: __________	
Witness:	(sign) ______________________________	Date: __________	

Appendix B

Background Questionnaire

Participants responded to these questions on the computer before starting the experiment.

1. Please enter the subject's age:
2. Please enter the subjects' sex:

    Male

    Female
3. What is your current job?

    CPC

    Front Line Manager

    TMC

    Staff

    Other
4. How many years have you worked as a Certified Professional Controller (or equivalent) or Front Line Manager (or equivalent) in TRACON?
5. How many years have you worked as a Certified Professional Controller (or equivalent) or Front Line Manager (or equivalent) in En Route?
6. How many years have you worked as a Certified Professional Controller (or equivalent) or Front Line Manager (or equivalent) in the tower?
7. How many years have you worked as a Certified Professional Controller (or equivalent) or Front Line Manager (or equivalent) in TMU?
8. How many years have you worked as a Certified Professional Controller (or equivalent) or Front Line Manager (or equivalent) in the ATC System Command Center?
9. How many years experience do you have using STARS?
10. How many years experience do you have using CARTS?
11. How many years experience do you have using ARTS IIIA?
12. How many years experience do you have using ARTS IIIe?
13. How many years experience do you have using STARS?
14. How many years experience do you have using other systems?
15. Is there anything else we should know about your participation in this study (e.g., prior study experience, system design, or design background)?

Appendix C

Task Instructions

Task 1 Instructions

Introduction: Designing appropriate symbology for safety critical systems is an extremely important component of good display design. Poorly designed symbology can lead to errors and can make a system unsuitable for performing certain functions.

We would like to apply human factors design practices to the design of an enhanced target symbol set for terminal displays. To accomplish this, we identified types of information that might be operationally useful for TRACON controllers if coded into the target. The four types of information we coded include:

1. category:
    - small
    - large
    - heavy
    - super heavy
2. heading
3. altitude
4. conformance

You will complete seven tasks to help us evaluate comprehension, preference, and reaction times for both existing coding schemes and novel coding schemes. If you have any questions, please ask the experimenter now.

Task 2 Instructions

Instructions: For this task, you will classify 12 of the target symbols from the previous set. However, in this task you will assign EACH symbol to one of the four aircraft type categories. Pick the category you believe provides the best match for that symbol. You will need to reuse categories to complete this task.

Task 3 Instructions

Instructions: For this set of trials, you will see mock ups of a radar display with aircraft targets. The target symbols on this display are from a typical STARS terminal display operating in single sensor mode & the target and beacon symbols will be oriented toward the closest radar. Before each trial, a + will appear in the center of the screen. After the +, you will be given a cue telling you which aircraft to find. The cue will indicate one of the following:

category
heading
altitude
category + heading

The cue will then be replaced by the image of a radar screen. When the image appears, you should click on the aircraft that matches the cue. Please respond as quickly and accurately as possible. Once you make your selection, a new trial will begin. There are 64 trials in this set.

Task 4 Instructions

Instructions: For this task, the targets are oriented toward the direction of flight and are shaped like TRIANGLES. Before each trial, a + will appear in the center of the screen. After the +, you will be given a cue. The cue will contain one or more of the following:

category
heading
altitude
category + heading

The cue will then be replaced by an image of a radar screen. When the image appears, click on the aircraft that matches the cue. Please respond as quickly and accurately as possible. Once you make your selection, a new trial will begin. There are 64 trials in this set.

Task 5 Instructions

Instructions: For this task the targets provide information about the four aircraft categories:

small
large
heavy
super heavy

The placement or absence of lines indicates category membership.

1. The triangle with NO LINES indicates that the aircraft is a SMALL aircraft.

   - Small aircrafts are the lightest, so this is the most basic of the symbols.

2. The triangle with a SINGLE LINE in the FRONT indicates that the aircraft is a LARGE aircraft.

   - Note that the line in the front of the symbol adds “weight” to the symbol, and so this symbol is “heavier” than the basic symbol.

3. The triangle with a SINGLE LINE in the BACK indicates that the aircraft is a HEAVY aircraft.

- The line in the back makes the symbol look like "weight" has been added to the rear, and so this symbol is "heavier."

4. The triangle with TWO LINES in the BACK indicates that the aircraft is a SUPER HEAVY aircraft.

   - The two lines make it look like even more "weight" has been added to the rear, and so this symbol is the "heaviest"

There are 64 trials in this set.

Task 6 Instructions

Instructions: For the next set of trials, we have come up with a coding scheme for aircraft altitude using symbols you encountered in previous sets. We have coded aircraft altitude in these targets using dark green to indicate the lowest altitude and lighter greens to indicate higher altitudes. You can use this coding to help you remember which aircraft is at which altitude.

For the upcoming round of trials, you will see a static radar display with aircraft targets that look like triangles that are oriented to the direction of flight. Category membership is indicated by the placement or absence of lines and altitude is indicated by color coding. All other information in the data blocks remains the same.

Before each trial, a fixation cross will appear in the center of the screen. An aircraft category, a heading, an altitude, or a category + heading will replace the fixation cross. This will indicate the type of aircraft you will be looking for in the next slide. An image of a radar screen will then replace the search cue and your task will be to find, as quickly and as accurately as possible, the aircraft that matches the given cue. Once you click on the correct aircraft with your mouse, the next trial will begin.

Task 7 Instructions

Instructions: In this task, you will see a set of four possible aircraft targets symbols that could be used to indicate conformance. We would like you to assign a symbol to each of the following four conformance categories: Not Monitored, In Compliance, Out of Compliance –Projected, and Out of Compliance – Actual. There is no right answer. We would simply like your opinion on which four symbols you believe best represent those conformance categories.

Appendix D

Tasks 3 through 6: Sample Set of Symbology Questions

Target Symbology Assessment

Table D1. Tasks 3-6: Sample set of symbology questions

No.	Question type	Question
1	heading	Click on the aircraft that is heading: North
2	category	Click on the aircraft that is a category: Small
3	altitude	Click on the aircraft that has: the highest altitude
4	category + heading	Click on the aircraft that is: Small, Heading North
5	heading	Click on the aircraft that is heading: North
6	category	Click on the aircraft that is a category: Large
7	altitude	Click on the aircraft that has: the lowest altitude
8	category + heading	Click on the aircraft that is: Large, Heading North
9	heading	Click on the aircraft that is heading: North
10	category	Click on the aircraft that is a category: Heavy
11	altitude	Click on the aircraft that is between altitudes: 6000'-9000'
12	category + heading	Click on the aircraft that is: Heavy, Heading North
13	heading	Click on the aircraft that is heading: North
14	category	Click on the aircraft that is a category: Super heavy
15	altitude	Click on the aircraft that is between altitudes: 9000'-12000'
16	category + heading	Click on the aircraft that is: Super Heavy, Heading North
17	heading	Click on the aircraft that is heading: South
18	category	Click on the aircraft that is a category: Small
19	altitude	Click on the aircraft that is between altitudes: 3000'-6000'
20	category + heading	Click on the aircraft that is: Small, Heading South
21	heading	Click on the aircraft that is heading: South
22	category	Click on the aircraft that is a category: Large
23	altitude	Click on the aircraft that has: the lowest altitude
24	category + heading	Click on the aircraft that is: Large, Heading South
25	heading	Click on the aircraft that is heading: South
26	category	Click on the aircraft that is a category: Heavy
27	altitude	Click on the aircraft that has: the highest altitude
28	category + heading	Click on the aircraft that is: Heavy, Heading South
29	heading	Click on the aircraft that is heading: South
30	category	Click on the aircraft that is a category: Super Heavy

(*table continues*)

Table D1. (continued)

No.	Question type	Question
31	altitude	Click on the aircraft that is between altitudes: 0000'-3000'
32	category + heading	Click on the aircraft that is: Super Heavy, Heading South
33	heading	Click on the aircraft that is heading: Northwest
34	category	Click on the aircraft that is a category: Small
35	altitude	Click on the aircraft that is between altitudes: 6000'-9000'
36	category + heading	Click on the aircraft that is: Small, Heading Northwest
37	heading	Click on the aircraft that is heading: Northwest
38	category	Click on the aircraft that is a category: Large
39	altitude	Click on the aircraft that is between altitudes: 9000'-12000'
40	category + heading	Click on the aircraft that is: Large, Heading Northwest
41	heading	Click on the aircraft that is heading: Northwest
42	category	Click on the aircraft that is a category: Heavy
43	altitude	Click on the aircraft that has: the lowest altitude
44	category + heading	Click on the aircraft that is: Heavy, Heading Northwest
45	heading	Click on the aircraft that is heading: Northwest
46	category	Click on the aircraft that is a category: Super Heavy
47	altitude	Click on the aircraft that has: the highest altitude
48	category + heading	Click on the aircraft that is: Super Heavy, Heading Northwest
49	heading	Click on the aircraft that is heading: Southwest
50	category	Click on the aircraft that is a category: Small
51	altitude	Click on the aircraft that has: the highest altitude
52	category + heading	Click on the aircraft that is: Small, Heading Southwest
53	heading	Click on the aircraft that is heading: Southwest
54	category	Click on the aircraft that is a category: Large
55	altitude	Click on the aircraft that is between altitudes: 0000'-3000'
56	category + heading	Click on the aircraft that is: Large, Heading Southwest
57	heading	Click on the aircraft that is heading: Southwest
58	category	Click on the aircraft that is a category: Heavy
59	altitude	Click on the aircraft that is between altitudes: 3000'-6000'
60	category + heading	Click on the aircraft that is: Heavy, Heading Southwest
61	heading	Click on the aircraft that is heading: Southwest
62	category	Click on the aircraft that is a category: Super Heavy
63	altitude	Click on the aircraft that has: the lowest altitude
64	category + heading	Click on the aircraft that is: Super Heavy, Heading Southwest

Appendix E

Exit Questionnaire

Target Symbology Assessment

Exit Questionnaire

Instructions. This questionnaire asks for your opinion about the symbols you have seen in this study. Please be as specific as possible.

1. What type of impact would the heading information in the enhanced target symbol (i.e., the triangles) have on your ability to control traffic?

1	2	3	4	5	6	7
Very Negative	Somewhat Negative	Slightly Negative	No Impact	Slightly Positive	Somewhat Positive	Very Positive

Explain your answer. ____

____

____

____

____

____

2. What type of impact would the enhanced target symbol with the aircraft category (i.e., the triangle with the category markings) have on your ability to control traffic?

1	2	3	4	5	6	7
Very Negative	Somewhat Negative	Slightly Negative	No Impact	Slightly Positive	Somewhat Positive	Very Positive

Explain your answer. ____

____

____

____

____

____

3. What type of impact would the enhanced target symbol with the aircraft category and altitude (i.e., the triangle with the category markings and color coded altitude) have on your ability to control traffic?

1	2	3	4	5	6	7
Very Negative	Somewhat Negative	Slightly Negative	No Impact	Slightly Positive	Somewhat Positive	Very Positive

Explain your answer. ____
____
____
____
____
____

4. How helpful would the flight plan conformance symbology presented in this study be for controlling traffic in an operational environment?

1	2	3	4	5	6	7
Very Unhelpful	Somewhat Unhelpful	Slightly Unhelpful	Neutral	Slightly Helpful	Somewhat Helpful	Very Helpful

Explain your answer. ____
____
____
____
____
____

5. Do you have any additional comments or recommendations regarding the use of enhanced target symbology on the STARS display in the operational environment?

____
____
____
____
____
____
____

Appendix F

Recommendations: Designing Symbols for ATC Terminal Displays

One factor designers must consider when developing new symbols is the end user. The primary users of terminal displays are terminal controllers, and the primary responsibility of terminal controllers is to separate traffic while guiding aircraft through the airspace and to their destination. The purpose of the terminal displays is to provide all of the relevant air traffic information that the controllers need to perform these duties. This information must be in a format that communicates all of the information that is necessary for the controllers to maintain an appropriate level of situational awareness regarding the activities of aircraft both in and around the terminal airspace. This section provides guidelines that designers, vendors, and researchers should use when developing and evaluating new or existing symbology for future ATC terminal displays.

A. Symbol design considerations.
   1. Design symbols for the user (Cardosi & Hannon, 1999).
      a. Understand the environment where the symbol will be used.
         i. Lighting considerations.
         ii. Viewing distance.
         iii. Existing symbol sets.
      b. Understand the tasks being performed by those who will use the symbols.
         i. Define relevant information to be communicated by the symbol(s).
         ii. Prioritize information.
         iii. Code only the most tactically useful information.
   2. Maximize positive transfer (Cardosi & Murphy, 1995).
      a. Make use of existing symbol sets unless there is a defined advantage to creating novel symbols.
   3. Ensure that symbols are legible (Yeh & Chandra, 2005).
      a. Measure the relevant physical qualities of the display, viewing factors related to the location of the user of the display, and the physical environment where the display will be used. Consider the following factors when determining the legibility of a candidate symbol (Ahlstrom & Longo, 2003):
         i. Display resolution.
         ii. Contrast.
         iii. Display size.
         iv. Viewing distances.
         v. Viewing angles.
         vi. The degree of visual arc subtended by the symbol.
         vii. Lighting conditions.
         viii. Display background.

4. Define the minimum and maximum target size (Ahlstrom & Longo, 2003; Lindberg & Näsänen, 2003).
    a. The minimum size of the symbol must preserve the key features that define it.
    b. The maximum size of the symbols must allow the controllers to display multiple targets at the same time.
    c. Ensure that you do not compromise legibility and discriminability at either the minimum or the maximum size.
    d. Use a single size to represent all aircraft on the display.
        i. Ensure that if the users can adjust the size of symbols, that all symbols of a type are adjusted simultaneously.
        ii. Observers had difficulty detecting the collision of two objects that differed in size, but their errors decreased with equal-sized objects (Delucia, 1995).
5. Ensure that a symbol is distinctive (Yeh & Chandra, 2005).
    a. Make sure the symbol can be discriminated from other symbols.
    b. Evaluate distinctiveness both within and across symbol sets.
        i. Will end users confuse the symbol with any other symbols currently on the display?
        ii. Is the symbol confusable with any other symbols end users may encounter on other tools or displays?
        iii. If two symbols are similar, determine whether their meanings are consistent or inconsistent. If the meanings are inconsistent, develop an alternative symbol so as to avoid negative transfer.
6. Ensure that a symbol is interpretable.
    a. Design symbols so that the meaning of the symbol can be easily identified (Yeh & Chandra, 2004).
        i. Determine whether end users can readily discern the meaning of the symbol.
        ii. When designing a set of symbols that convey a similar type of information, make them similar enough to convey the notion that they are part of a set, without making them so similar that they become hard to differentiate.
        iii. If a symbol uses a feature or attribute to represent an information parameter, ensure that the critical feature is legible for the typical end user.
    b. Design symbols using shape and symbol coding that users can recognize and easily identify.
        i. For example, represent heading using the point of an arrow (e.g., triangle, chevron).

c. Avoid using similar features to represent different information parameters (Chandra, Zuschlag, Helleberg, & Estes, 2009).

d. Avoid using a combination of features to represent a single information parameter.

7. Ensure that a symbol is salient (Yeh & Chandra, 2005).

    a. Define how much attention a particular symbol should be given on the display.

        i. Determine whether the symbol is equal in salience to other symbols.

        ii. Determine whether the symbol needs to stand out from other symbols (e.g., alerts).

    b. If a symbol needs to have a greater salience than other items on the display, measure whether the color of the symbol meets existing requirements for attention (Friedman-Berg, Allendoerfer, & Pai, 2009).

    c. Conduct performance evaluations to determine whether and how much a symbol stands out in a given context and determine whether the degree of salience is appropriate for that symbol.

        i. For example, conduct a target search experiment to measure how quickly (reaction time) and how accurately (% correct identifications) a target can be found on typical display configurations.

            1. Does the symbol stand out enough?

            2. Does the symbol stand out too much?

        ii. Obtain measures of display clutter.

            1. Global density.

            2. Local density.

            3. Other clutter metrics (Bravo & Farid, 2008).

        iii. Identify features of the symbol that may influence salience.

            1. Color.

            2. Intensity.

8. Ensure that color coding is used appropriately.

    a. Use colors that are consistent with existing color coding conventions (Cardosi & Hannon, 1999; Friedman-Berg, Allendoerfer, & Pai, 2009).

        i. For instance, if purple has an existing meaning, ensure that new uses of purple are consistent with this meaning.

    b. Use colors that are consistent with cultural color coding conventions (Cardosi & Hannon, 1999; Friedman-Berg, Allendoerfer, & Pai, 2009).

i. Red is typically used for alerts.

ii. Yellow is typically used for warnings.

c. Use color to highlight or to capture attention (Ahlstrom & Longo, 2003).

d. Use colors that are distinguishable and that conform to recommendations for color use on ATC displays (Friedman-Berg, Allendoerfer, & Pai, 2009).

B. Steps for evaluating symbols (Yeh & Chandra, 2005).

1. Collect and catalog existing symbols and symbol sets.

a. Identify symbol sets currently in use in the end users workspace.

b. Identify symbol sets in similar domains (McFadden, Jeon, Li, & Minniti, 2008).

i. Naval Tactical Data System (NTDS) symbol set.

ii. North Atlantic Treaty Organization draft STANdardization AGreement (STANAG) 7161.

iii. North Atlantic Treaty Organization draft STANdardization AGreement (STANAG) 4420.

iv. Military Standard 1787C (MIL-STD-1787C).

v. Military Standard 2525B (MIL-STD-2525B).

vi. Military Standard 2525C (MIL-STD-2525C).

vii. International Civil Aviation Organization (ICAO) standards document, Aeronautical Charts Standards and Recommended Practices (SARPS) Annex 4.

viii. RTCA DO-257A, Minimum Operational Performance Standards for the Depiction of Navigational Information on Electronic Maps.

ix. Society of Automotive Engineers (SAE) Aerospace Recommended Practice (ARP) for electronic aeronautical symbols.

2. Develop new symbols.

a. Determine whether there is an informational need for new symbols.

i. Determine whether there are enough current symbols.

ii. Determine whether current symbols meet future system and informational requirements.

b. Determine whether new symbols conflict with existing symbols.

    i. Determine whether end users can interpret the new symbols.

    ii. Determine whether end users can integrate the information conveyed by the new symbols with the information conveyed by the existing symbols. (Yeh & Chandra, 2006).

    iii. Determine how much similarity with existing symbols is desirable.

    iv. Determine whether the coding conveys the desired meaning (Snodgrass & McCullough, 1986).

3. Identify the end users who will evaluate the symbols.

    a. Test and evaluate the symbols using the end users.

        i. Although nonexperts, retired users, or Subject Matter Experts (SMEs) may provide valuable preliminary input, their input cannot take the place of the experienced end user.

    b. Recruit participants from a variety of backgrounds to ensure that your user group represents all potential end users.

4. Determine the appropriate methodology for evaluating symbols. It may be desirable to use more than one method to evaluate symbols.

    a. Visual search tasks.

        i. Measure reaction times.

        ii. Measure accuracy.

    b. Ease of learning.

        i. Measure time to proficiency.

        ii. Measure symbol confusions.

        iii. Measure asymptotic accuracy.

    c. Symbol comprehension.

        i. Symbol identification tasks.

            1. Match a symbol with its meaning.

        ii. Symbol recognition tasks (McFadden, Jeon, Li, & Minniti, 2008).

1. Measure accuracy.
2. Measure reaction times.
3. Generate confusion matrices.

iii. Symbol categorization tasks.

1. Measure category preference.

5. Measure symbol acceptability.
   a. Collect acceptability ratings.
   b. Collect subjective feedback.
      i. Ease of learning.
      ii. Intuitiveness.
      iii. Impact on performance on standard tasks.