CHARACTERIZING SCAN PATTERNS IN A SPACECRAFT COCKPIT SIMULATOR: EXPERT VS. NOVICE PERFORMANCE

Valerie A. Huemer M.S., Miwa Hayashi, Ph.D.
San Jose State University Foundation
Moffett Field, CA

Fritz Renema, Steve Elkins
QSS Group
Moffett Field, CA

Jeffrey W. McCandless, Ph.D., Robert S. McCann, Ph.D.
NASA Ames Research Center
Moffett Field, CA

Operating a spacecraft is a complex and demanding task that requires years of training and constant monitoring of both navigation and systems parameters. By examining differences in scanning between "expert" and "novice" operators, we can develop cognitive models of scanning behavior or enhance training. In the Intelligent Spacecraft Interface Systems (ISIS) laboratory, we measure eye movements and record performance parameters in a part-task space shuttle cockpit simulator. We trained airline transport pilots (as our "novice" group) on fundamentals of flying an ascent ("launch-to-orbit") in the space shuttle. We tested three levels of malfunctions occurring during a trial—none (nominal), one malfunction, or three malfunctions—on both pilots and astronauts (our "expert" group). Astronauts had fewer errors and faster reaction times. Eye movement analyses showed that both astronauts and pilots similarly modified their scan strategies depending on the flight segment and how many malfunctions occurred during a trial.

INTRODUCTION

The space shuttle was developed in the 1970's, when the human factors field was still in its infancy. Researchers now have a better understanding of the importance of effective human-machine interfaces and the usefulness of cognitive models in developing automation and enhancing training. Future space vehicles will travel farther from Earth and on longer missions, and thus will need to have more human-centered interfaces and increased automation to enable greater onboard capabilities. To design effective man-machine interfaces, and to develop cognitive models for collaborative automation and enhanced training, we need to have a preliminary understanding of effective visual scanning behavior in a spacecraft cockpit.

Visual scanning behavior has been characterized in several different environments, but not yet in a spacecraft cockpit. One area that has benefited considerably from several studies of scanning behavior is aviation. Researchers have characterized pilots' scan patterns in a variety of ways, such as examining distribution of visual resource (Flemisch & Onken, 2000; Anders, 2001) and applying models to eye movements (Hayashi, 2003; Wickens, Helleberg, Goh, et al., 2001).

In aviation, eye movements have also been used to compare performance between expert and novice operators. Here, also, eye movements have been used in a large range of ways, from enhancing training (Wetzel, Anderson, & Barelka, 1998), to comparing quantitative differences between experts and novices (Ottati, Hickox, & Richter, 1999).

Comparisons between expert and novice operators can be particularly helpful in identifying what scanning strategies experts use effectively, and how experts' strategies and mental models differ from those of novices. Bellenkes, Wickens, & Kramer (1997) examined visual scanning and attentional flexibility in expert and novice pilots to explore performance differences and how mental models relate to performance. They measured performance and eye movements of expert and novice pilots over several segments of flight with varying task demands. Experts performed better in almost every metric. Bellenkes et al. found that this better performance could be explained in the context of resource theory; that is, experts gathered information more efficiently and thus had more attentional resources available to monitor tasks of lower priority and deal with varying task demands. Experts also had greater attentional flexibility than novices and, unlike novices, would differentially allocate their scanning capacity depending on the tasks demands of different segments of flight. Experts also demonstrated more automated skill in extracting information from flight instruments and a more refined mental model of flight dynamics. This study went beyond quantifying expert and novice scanning differences, as many other eye movement studies have done, by showing that eye movements can be used to examine attentional strategies and mental models, and that these strategies and mental models provide clues as to why performance differs between experts and novices. They suggested that such information could be used to develop targeted training of expert strategies.

Although eye movement research has shed considerable light on pilots' real-time behavior in aircraft cockpits, we know very little about real-time information acquisition strategies in a spacecraft cockpit. Simple extrapolations from an aircraft environment may be inappropriate, as the task mix on a spacecraft is substantially different. Spacecraft contain extremely complex propulsion, power, and other engineering systems that must operate in a much harsher environment than the systems onboard an aircraft (McCann & McCandless, 2003). Whereas today's aircraft have benefited from a century of systems and operations refinements, many displays in today's spacecraft are the products of only first or second generation engineering. When systems malfunctions occur, dealing with them is a major task requirement. To attain proficiency in these activities, crewmembers undergo a minimum of two years intense training on the vehicle and its systems.

Because the task mixes in a spacecraft and an aircraft are so different, the displays and acquisition strategies in the two vehicles also necessarily differ. In both spacecraft and aircraft, navigational parameters (such as attitude, altitude, and speed) must be constantly monitored. However, spacecraft have different systems and operational objectives than aircraft (for instance, the space shuttle accelerates from 0 mph to MACH 25 in the first 8 ½ minutes of flight). Thus, monitoring requirements to meet these operational objectives in these vehicles necessarily differ, as do the displays in each vehicle to meet these objectives. The fact that the shuttle cockpit has a different mix of flight versus systems status displays than an aircraft cockpit further complicates any effort to extrapolate aircraft scan patterns to shuttle scan patterns.

Consider the information processing requirements during ascent ("launch-to-orbit"), which lasts a mere 8 ½ minutes. Crew responsibilities during this highly dynamic phase of flight involve both specific time-dependant checks of various parameters (such as solid rocket booster separation and functional status of the Freon system at two minutes and three minutes into flight, respectively) and continuous monitoring of the vehicle's navigation and system states. During ascent, an astronaut must continuously monitor many navigational parameters, such as trajectory, velocity, vertical velocity, attitude, and current abort options, as well as system parameters, such as main engine ullage pressures and helium flows. In addition to these tasks, when a malfunction occurs, the astronaut must act quickly to assess the situation and solve the problem.

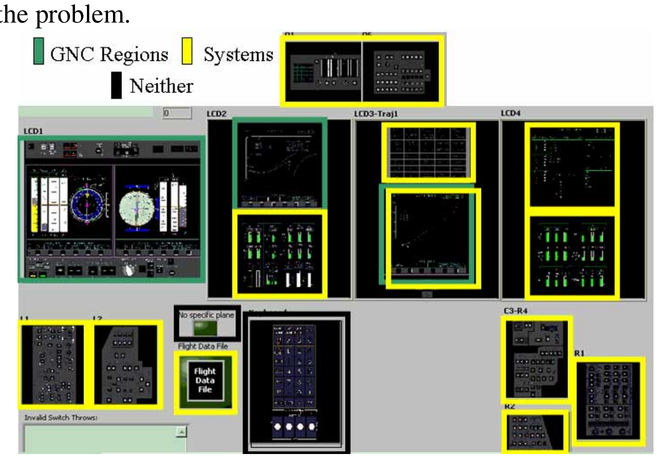

Figure 1. Classes of Information (GNC, Systems, Neither) across a space shuttle cockpit

To understand how this monitoring is accomplished, we can categorize crew members' fixations as either flight-related (that is, fixations on what shuttle operators define as guidance, navigation and control [GNC] displays) or as systems-related (fixations on systems status or systems summary displays). Figure 1 shows the categorization of what kind of information (flight-related versus systems-related) is shown by each cockpit display.

The relative importance of each type of information differs depending on the current flight segment of the ascent phase. For instance, from vehicle lift-off to about 30 seconds into flight, the primary monitoring activities involve checking various GNC parameters to verify the vehicle is rolling to a target attitude and maintaining the proper trajectory. As the flight continues, however, systems must be monitored, and GNC monitoring decreases to meet these new task requirements.

For the purpose of our analyses, the proportion of GNC monitoring varies during a nominal 8 ½ minute ascent phase of flight over the following functionally-divided segments (defined by how much mission elapsed time (MET) has elapsed):

- Segment 1: MET 0:00-0:30
  This segment involves mostly initial navigation (GNC) checks, but some checks of systems parameters.
- Segment 2: MET 0:31-2:20
  From initiation of the "thrust bucket"[1] to solid rocket booster (SRB) separation, this segment involves both navigation and systems monitoring.
- Segment 3: MET 2:21-5:40
  From SRB separation to the "roll-to-heads-up" check, this segment involves almost equal navigation and systems monitoring during nominal flight (i.e., when no malfunctions occur).
- Segment 4: MET 5:41-8:30
  From the "roll-to-heads-up" check to main engine cut-off (MECO), navigation checks increase in importance, yet important systems checks, such as 3-G throttling and MECO, also occur.

Goals of the Present Study

The present study uses eye movement analyses to compare the scanning strategies of an experienced group of astronauts with a less-highly trained (in terms of "astronaut" training) group of airline pilots over the different task requirements of the above flight segments. To examine adaptive aspects of scanning strategies, we introduced varying levels of difficulty by inserting various numbers of malfunctions during the runs. The goal was to achieve some preliminary understanding of how "novices" and "experts" acquire visual information and what visual scanning strategies they use in the context of a spacecraft cockpit, specifically, how the proportion of GNC monitoring is affected when malfunctions are introduced, and over the varying task requirements of different flight segments. We used the proportion (of all fixations) that was navigation-related (i.e.,

[1] Power of the shuttle main engines is momentarily reduced when the shuttle is in the beginning of its flight in order to reduce the aerodynamic force to tolerable limits. This is known by astronauts as going through the "thrust bucket."

were directed to information on GNC displays) as a dependent variable for our analyses. We hypothesized that because the malfunctions which we simulated were all systems-related, the participant must draw his attention from GNC checks due to the increased information processing requirements needed to process systems information, and that the change in GNC fixations would be a sensitive measure of the disruption of the malfunctions and the varying task requirements. Our ultimate goal in this characterization of spacecraft cockpit behavior is to develop cognitive models for effective automation in next-generation spacecraft, and to enhance training through better understanding of effective expert scanning strategies.

METHOD

Participants

Five astronauts, with a minimum of two years of astronaut training, formed the "expert" group in our experiment. Six airline transport pilots, with an average of 15,000 flight-hours on various aircraft, participated as our "novice" group. Experienced airline pilots were chosen as the novice group because they already had familiarity with flight dynamics and effective aviation scan techniques (although not for the specific task of flying a spacecraft).

Apparatus

A one-person part-task simulator at the Intelligent Spacecraft Interface Systems (ISIS) laboratory at NASA Ames Research Center, partially replicating the commander (left) side of the space shuttle cockpit, was used for the experiment. The fixed-base simulator consists of a wood apparatus (to avoid interference with the magnetic head-tracker) that holds 12 computer monitors, which are used to represent many of the cockpit displays and switch panels in a layout similar to that of the actual space shuttle. Touch-panel liquid crystal display (LCD) monitors are used to allow the subject to manipulate switches, as required. An audio system provides background engine noise, SRB separation noise, and alarm annunciation.

The entire system is driven by a distributed, multi-platform (SGI and PC) set of computers, and is controlled and monitored outside of the simulator room by an experimenter at an experiment operator station.

Eye movements were measured with a head-mounted ISCAN ETL-500 eye-tracking system (ISCAN Inc., Burlington, MA) and a magnetic head-tracker (FasTRAK, Polhemus, Colchester, VT).

Procedure

Prior to testing, the airline pilots ("novice" group) were trained during a week-long training course on the basic shuttle systems, nominal monitoring tasks during ascent, and resolution procedures for specific possible malfunctions. Each pilot was also given a 2-hour simulator familiarization session. Each astronaut ("expert" group) was given three 10-minute simulator familiarization runs before testing.

Participants in each group completed 4 ascent scenario trials, representing vehicle flight from launch to MECO. Each trial lasted 8 ½ minutes of simulator time. The trials consisted of two nominal runs (the first and the last run), a run with a single malfunction occurring, and a run with three malfunctions. Simulator parameters, switch throws, and eye movements were recorded during each trial.

During the single-malfunction run, a leak in the external tank holding the fuel (liquid hydrogen) for the main engines (annunciated to the participants as a low ullage pressure problem) was inserted at 1:50 MET (a period of high GNC monitoring, right before SRB separation). During the three-malfunction run, the following systems malfunctions were inserted: 1) a malfunction involving a regulator in the helium supply subsystem for one of the shuttle's three main engines (annunciated to the participants as a main engine helium pressure problem) at 1:50 MET; 2) a failure of one of the four onboard general purpose computers to maintain synchronous operations with the remaining three machines (annunciated to the participants as a computer fail-to-synch problem) at 2:00 MET, and 3) a failure in the vehicle's thermal management system responsible for cooling the freon loops during ascent (annunciated to the participants as an evaporator out temperature high problem) at 3:05 MET.

The independent variables are spacecraft expertise level, number of malfunctions occurring during a trial, and segment of flight. To measure the affect that malfunctions had on nominal scanning, we used the proportion of GNC fixations as a dependent variable for all runs, as this measure represents required scanning of flight parameters during flight. For runs with malfunctions, additional dependent variables are malfunction resolution procedure accuracy and response time.

RESULTS

Performance during Malfunctions

Single malfunction trial. Astronauts performed the correct procedure 100% of the time with a mean reaction time (time from alarm annunciation to correct resolution) of 22 seconds during single malfunction trials. Pilots performed the correct procedure 80% of the time with a mean reaction time of 58 seconds. Astronauts performed the procedure significantly faster than the pilots ($t(9) = -5.07$, $p < 0.001$).

Multiple malfunction trial. Astronauts performed 87% of procedures correctly, significantly more ($t(5) = 4.08$, $p < 0.01$) than pilots, who performed 28% of procedures correctly. Astronauts had a mean reaction time of 2:07 minutes to completion of each procedure during the multiple malfunction trials, while pilots had a mean reaction time of 2:58 minutes to completion of each procedure.

Eye Movement Analyses

Eye movement data were categorized into fixations, with a fixation defined as gaze in the same area (within 1 degree of visual angle) for at least 150 msec. Fixations on

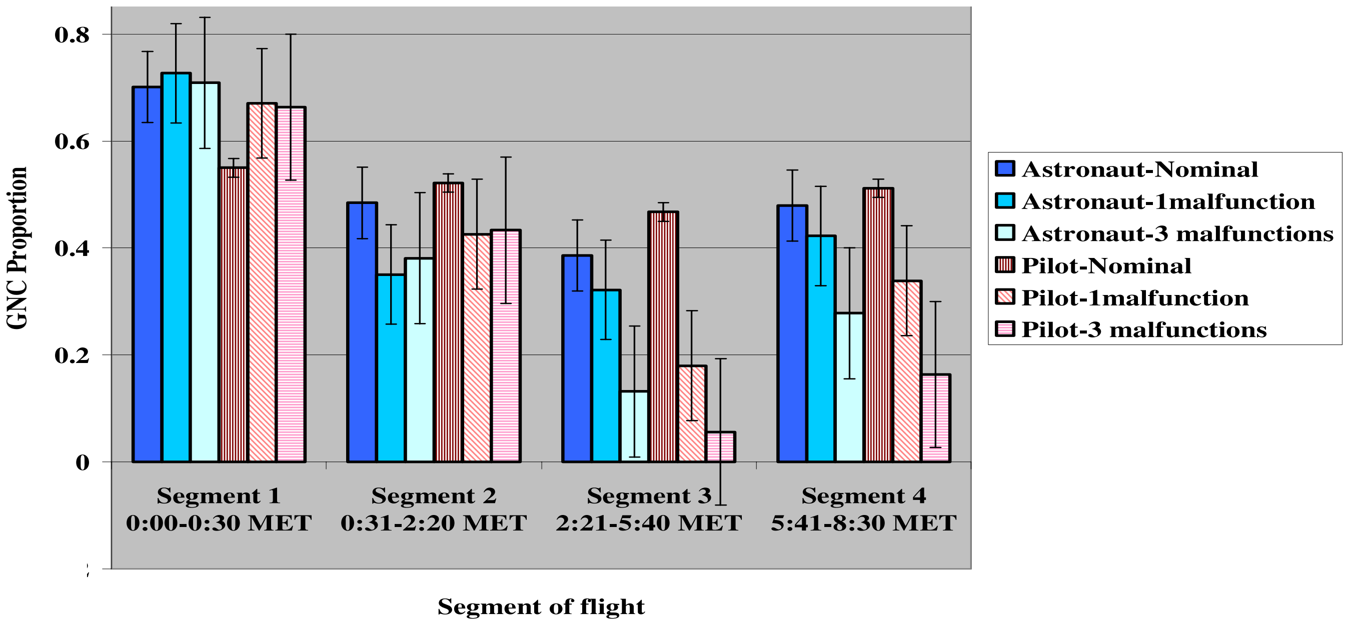

Figure 3. Percentage of time fixating on GNC displays

viewed display provided—GNC, systems, or neither (see Figure 2).

We used the proportion (of all fixations) that was navigation-related (i.e., were directed to information on GNC displays) as a dependent variable for our analyses of the following independent variables:

- Expertise - Astronaut vs. Pilot (2 levels)
- Level of Malfunctions (0, 1, 3) during the trial (3 levels)
- Segment of Flight (4 levels-described in the introduction)

A 3-way ANOVA performed on these independent variables showed a significant effect of Level of Malfunctions, F (2, 96) = 16.66, $p < 0.001$. A post-hoc analysis indicated that the proportion of GNC fixations were significantly higher during nominal runs than during runs with three malfunctions ($t(12)=2.26$, $p<0.05$).

There was also a significant effect of Segment of Flight, F (3, 96) = 59.18, $p < 0.001$. Post-hoc analyses indicated that the proportion of GNC fixations were significantly higher in Segment 1 than in Segment 2 ($t(8)=5.7$, $p<0.001$), Segment 3 ($t(8)=5.45$, $p<0.001$), or Segment 4 ($t(8)=4.79$, $p<0.001$). Also, GNC fixations in Segment 2 were significantly higher than those in Segment 3 ($t(8)=2.43$, $p<0.05$).

Additionally, there was a significant Level of Malfunctions by Segment of Flight interaction, F (6, 96) = 5.64, $p < 0.001$. Level of Malfunctions had greater effect on GNC monitoring in Segments 3 and 4 (the Segments after malfunctions occurred) than in Segments 1 and 2 (before malfunctions occurred). There was no significant effect of Expertise (astronaut vs. pilot).

Figure 3 shows the proportion of astronaut and pilot GNC fixations as a function of flight segment for the three levels of malfunction trials (nominal, one malfunction, three malfunctions). Astronaut proportions are shown by the solid leftmost bars of each segment group, and the pilot proportions are shown by the striped rightmost bars of each segment group.

As expected (by the GNC monitoring requirements mentioned in the introduction), both astronauts and pilots consistently had a high proportion of GNC fixations during the first Segment of Flight. During nominal runs, astronauts maintained at least 38% GNC monitoring, although balancing it with systems monitoring in the varying proportions (based on the monitoring requirements) for each Segment of Flight. Pilots also had a high proportion (at least 46%) of GNC fixations during nominal runs, although their GNC monitoring tended not to vary as much by Segment of Flight as the astronauts' GNC fixations did.

GNC fixations for both astronauts and pilots remained high (at least 48%) during the first and second Segments of Flight. When a malfunction was introduced (near the end of flight Segment 2), GNC fixations for both astronauts and pilots decreased on subsequent flight segments, especially during the three-malfunction trial.

DISCUSSION

In both the single malfunction and multiple malfunction trials, the astronauts correctly completed more procedures and performed the procedures faster than the pilots. This is not surprising, as astronauts are much more familiar with the relevant procedures and are less likely to make mistakes (and are thus able to complete more procedures in the limited time of ascent).

Since the malfunctions which we simulated were all systems-related, we used the change in proportion of GNC monitoring to measure disruption of nominal flight monitoring. We ran a 3-way ANOVA on GNC fixation proportions of eye movements, and found a significant main effect of Level of Malfunctions. Both astronauts and pilots significantly reduced their GNC scans when three malfunctions occurred. This can be explained in the context of resource theory: the malfunctions involved the shuttle systems, and resulted in more focus on those systems and less on GNC parameters during malfunction runs, given finite attentional resources.

We also found a significant main effect of Segment of Flight, which would be expected from the varying monitoring requirements during the four flight segments. Both astronauts and pilots focused on GNC fixations during the first segment of flight--a segment which requires several navigational monitoring tasks. Also both groups focused more on GNC monitoring in Segment 2 than in Segment 3. Since malfunctions began occurring (in non-nominal runs) at the end of Segment 2, it makes sense that GNC monitoring decreased in Segment 3, while participants were busy dealing with the systems-related malfunctions.

Finally, we found a significant Level of Malfunctions by Segment of Flight interaction, showing that the Segments were disproportionately affected by the Level of Malfunctions. GNC fixations varied to a greater degree by Level of Malfunctions in Segments 3 and 4 than they did in Segments 1 and 2. Since the malfunctions in the non-nominal conditions were introduced at the end of Segment 2, it is reasonable that both groups reduced their GNC fixations during Segments 3 and 4, which were the Segments occurring after the malfunction annunciation and the Segments where they had to sacrifice GNC monitoring in order to work to resolve the systems malfunctions.

Interestingly, we did not find a significant main effect of Expertise level ("expert vs. novice" or astronaut vs. pilot) on proportion of GNC monitoring. The lack of significance in this main effect could be due to several reasons. First, we had small sample sizes (five astronauts, six pilots). Secondly, we had an arbitrary definition of "novice." In their field of aviation, our pilots were experts, and thus may have had many of the information processing qualities of "experts," such as attentional flexibility (to vary scanning strategies by task requirements), and adaptability. Lastly, we used GNC monitoring as our dependent variable. Since navigational monitoring is also performed on an aircraft, we may have been measuring a skill common to both astronauts and pilots.

Although our pilots were not as adept as the astronauts in malfunction resolution, they did have effective scans and a proper understanding of monitoring task requirements. Thus, they can be useful in preliminary testing of new concepts for further study of monitoring tasks and automation in the next generation of spacecraft cockpits.

Further study is needed to find what other aspects of each group's eye movements differed. In the meantime, characterization of astronaut scan pattern can help improve training techniques. Measurement of eye-movements has the potential for training novice pilots (or astronauts) on monitoring of the vehicle state during nominal periods as well as fault recognition and identification during off-nominal conditions. Use of eye movement measurement has been used in aviation training, where flight instructors found it useful to be able to review novice's eye movements (Wetzel, Anderson, & Barelka, 1998). Another possible implementation of eye movement measurement is for scan pattern characteristics of expert astronauts to be recorded. When novice astronauts monitor nominal and off-nominal conditions, they can be alerted if critical aspects of their scan pattern (such as dwell time on a particular display) are markedly different from similar aspects of an expert's scan pattern.

Ultimately, we would like to use eye-movement data to develop a cognitive model of an expert astronaut's scan pattern, to understand how the expert astronaut is effectively able to monitor many parameters at once. Such a model can also be used to enhance training. A preliminary model of a subset of our data has been developed by Matessa and Remington (2005). Future studies will concentrate on developing a cognitive model of the supervisory monitoring behavior used by astronauts, and applying this model to enhance training and develop efficient automation for the next generation of space vehicles.

ACKNOWLEDGEMENTS

This project was funded by an NRA 01-OBPR-07 Award, an Engineering for Complex Systems: RSO Program Element 4.1.4, and the Shuttle Cockpit Avionics Upgrade Project. We would like to acknowledge Mike Matessa and Brent Beutter for their valuable input. And last (but not least) we would like to acknowledge the astronauts and pilots who participated in this project.

REFERENCES

Anders, G. (2001). Pilot's attention allocation during approach and landing – Eye- and head-tracking research in an A330 full flight simulator. *Presentation at the 11th International Symposium on Aviation Psychology*, Columbus, OH.

Bellenkes, A., Wickens, C., & Kramer, A. (1997, July). Visual scanning and pilot expertise: The role of attentional flexibility and mental model development. *Aviation, Space, and Environmental Medicine, 68* (7). 569-579.

Flemisch, O.F. and Onken, R., (2000). Detecting usability problems with eye tracking in airborne battle management support. In *Proceedings of the RTO HFM Symposium on "Usability of Information in Battle Management Operations,"* Oslo, Norway.

Hayashi, M. (2003) Hidden markov models to identify pilot instrument scanning and attention patterns. In Proc. IEEE Int. Conf. Systems, Man, and Cybernetics, pp. 2889-2896.

Ottati, W.L., Hickox, J.C., and Richter, J. (1999). Eye scan patterns of experienced and novice pilots during visual flight rules (VFR) navigation. In *Proceedings of the Human Factors and Ergonomics Society, 43rd Annual Meeting*. (Vol. 1; pgs 66-70). Santa Monica, CA: Human Factors and Ergonomics Society.

Matessa, M. and Remington, R. (2005). Eye movements in human performance modeling of space shuttle operations. In *Proceedings of the Human Factors and Ergonomics Society, 49th Annual Meeting*, Orlando, FL, Sept. 26-30 (in press). Santa Monica, CA: Human Factors and Ergonomics Society.

McCann, R.S. and McCandless, J. (2003). Human-machine teaming for dynamic fault management in next-generation launch vehicles. In *Proceedings of the Joint Army-Navy-NASA-Air Force 3rd Modeling and Simulation Meeting*, Colorado Springs, FL, Dec 1-5.

Wetzel, P.A., Anderson, G.M., Barelka, B.A. (1998). Instructor use of eye position feedback for pilot training. In *Proceedings of the Human Factors and Ergonomics Society, 42nd Annual Meeting*. (Vol. 2; A99-14126 02-54). Santa Monica, CA: Human Factors and Ergonomics Society.

Wickens, C. D., Helleberg, J., Goh, J., Xu, X., and Horrey, W. J. (2001). Pilot task management: Testing an attentional expected value model of visual scanning (Technical Report ARL-01-14/NASA-01-7). Moffett Field, CA: NASA Ames Research Center