INTEGRATED MODELING OF COGNITION AND THE INFORMATION ENVIRONMENT:

A Closed-Loop, ACT-R Approach to Modeling Approach and Landing With and Without Synthetic Vision System (SVS) Technology

Final Report for July 31, 2004

Submitted to:

David Foyle, Ph.D.
Level 3 Manager
AvSP SWAP Human Performance Modeling Element
NASA Ames Research Center (MS 262-4)
Human-Automation Integration Research Branch
Human Factors Research & Technology Division
Moffett Field, CA 94035-1000

Submitted by:

Michael D. Byrne, Ph.D.
Department of Psychology
Rice University, MS-25
Houston, TX 77005-1892
Email: byrne@rice.edu
Phone: 713-348-3770

Alex Kirlik, Ph.D.
Human Factors Division,
Depts. of Psychology, Mechanical &
Industrial Eng., & Beckman Institute
Univ. of Illinois at Urbana-Champaign
Email: kirlik@uiuc.edu
Phone: 217-351-0857

Table of Contents

Acknowledgments

The authors would like to acknowledge and sincerely thank the following people who played an essential role in conducting the research reported here. They include Michael Fleetwood, David Huss, and Chris Fick of Rice University, and Sarah Miller, Alex Kosorukoff, Ruei-Sung Lin, Sam Zheng, Ahmad Zakaria and Piotr Adamczyk of the University of Illinois at Urbana-Champaign. We could never have performed this research without their hard work and dedication to this project, in the realms of ACT-R modeling, compuer networking and data communications, laboratory task design and programming, reviewing the literature, experimental design, data collection, eye-movement data coding and analysis, statistical modeling and analysis, and decision modeling.

Executive Summary

Detailed analysis of the NASA-provided eye-tracking data from approach and landing scenarios using (a simulation of) Santa Barbara airspace, airport, aircraft, and 3 human pilots indicate that a Synthetic Vision System (SVS) display in aviation is not simply a proxy for looking out the window in conditions of reduced visibility. Instead, the SVS appears to take on the additional role of a redundant instrument cluster which the pilots used extensively. Eye movement analysis revealed that visual attention allocation was substantially altered by the presence of the SVS, even in conditions when the window was not being used as an information source by the pilots. We constructed a computational process model of the pilot-aircraft-environment system in ACT-R. The model was able to capture many high-level features of the observed, and SVS-induced, impact on visual attention allocation. In addition, the model was also able to capture much of the variance in attention allocation at an even finer grain of analysis, namely, the frequency of transitions between the various sources of visually displayed information in this task environment. Additionally, the model was able to combine high-level task factors with low-level visual factors, and thus provided insight into why the SVS display appeared to have the effects that it did on pilot attention allocation behavior in the experiment conducted by NASA.

To date, however, work remains to be done to refine and elaborate the model in various ways. Early on in this effort, we stated that we would guide our modeling, not only by the empirical data provided by NASA, but also by the state-of-the-art in visual attention allocation research. In reviewing that research, most notably the pioneering work of Senders introducing the notion that visual attention allocation is strongly influenced by the bandwidth of displayed information, and of Wickens and his colleagues on their recent SEEV (Salience, Expectancy, Effort, Value) model of visual attention allocation, we concluded that it would prove very useful to run an additional experiment in order to clarify the relationship between the *product* (i.e., SEEV) and *process* (i.e., ACT-R) approaches to modeling attention allocation. The main reason we thought this additional experimentation was essential came from the nature of the NASA SVS scenarios themselves. In some of those scenarios, a lack of a veridical relationship between the displayed scene and the actual environment was deliberately introduced to assess the impact of imperfect automation. Our literature review revealed that the possible presence of probabilistic information sources has never been central to either product- or process-oriented modeling of visual attention allocation.

This fact led us to the conclusion that any theoretical account (or model) of attention allocation relevant to behavioral situations such as the NASA SVS scenarios had to address, head-on, the fact that the proximal (the most readily available or salient perceptual environment, and the information sources it contains), may bear only a probabilistic, rather than fully informative, relationship to the variables, objects and events which are the true objects of successful pilot or operator adaptation. To date, however, our understanding of the visual attention allocation literature is that researchers have largely framed the functional role of human attention allocation as one of obtaining an veridical internal model of the proximal environment. The literature reviewed in this report provides strong support for this conclusion, and as such, we found that perhaps even the best extant theories of visual attention allocation largely leave aside the question of understanding and modeling the probabilistic relations between features of the readily available, proximal information environment (signal bandwidths, values, etc.) and the

distal environment serving as the criterion of successful adaptation. We thus performed an experiment, modeled after Senders (1964) classic visual scanning study, in which we introduced uncertainty between the most salient, readily available proximal variable in the task (signal bandwidth), and the task criterion (alarm detection). Our experiment, also in contrast to Senders' original study, also varied the values of the various information sources present on the display; one central motivation for this manipulation was the attempt to clarify how bandwidth and value combine in influencing visual attention allocation. Our results demonstrated that, in this task, visual dwell time percentages were better predicted by models combining value and bandwidth *additively*, rather than *multiplicatively*, the latter being the case in the most recent formulation of the SEEV model of Wickens and his colleagues. Perhaps even more importantly, we found that varying levels of predictiveness, or ecological validity, of the proximally displayed information (i.e., bandwidth) influenced both the rate of skill acquisition in this task and also the manner in which the visual displays were visually sampled. This may be a surprising result if one assumes that the goal of visual attention allocation is to maintain a veridical internal model of the proximal world. In contrast, it is not a surprising result under the assumption that the goal of visual attention allocation is largely to achieve a functionally adaptive relationship with a (perhaps distal) task criterion, using whatever proximal resources are available to do so.

We look forward to bringing these and other empirical findings to bear in our forthcoming refinements of the visual attention allocation mechanisms of our ACT-R model of the NASA SVS scenarios. The new experimental findings actually may be able to potentially inform future versions, or at least applications, of ACT-R itself, to the extent that they provide a data set well suited to ACT-R modeling, and especially for evaluating the learning and visual attention mechanisms in that model. We also hope that our work along this lines fosters an increasingly synergistic relationship between product-oriented (e.g., SEEV) and process-oriented (e.g., ACT-R) modeling of attention allocation. Finally, we hope that these developments will result in an even deeper, and potentially more useful, understanding of the influences of SVS systems and related forms of information automation in human-machine interaction in the broad range of human-machine systems of interest to NASA.

1. Introduction

This report provides an overview of the current status of our research evaluating the impact of Synthetic Vision System (SVS) technology on pilot performance in commercial aviation.

1.1 Problem Statement

Evaluation of new technology for the commercial aircraft cockpit is an expensive and time-consuming process. The pool of potential subjects is small and consists of individuals with extensive training who are both relatively difficult and expensive to access. Thus, the typical design-test-modify iteration cycle is generally both slow and costly. One approach with potential application in this domain is the substitution of computational cognitive modeling for at least some phases of empirical evaluation. While we are not suggesting entirely removing humans from the evaluation process, other engineering disciplines rely heavily on mathematical or computational simulation models as a routine part of design. (This has been argued in more detail elsewhere; e.g., Byrne & Gray, 2003).

The focus of this research is on the evaluation of a new technology for the commercial airline cockpit (Foyle, et al., 2003). One of the factors that has long limited aviation is visibility; poor visibility conditions can substantially change the task of piloting an aircraft. However, with extensive and accurate computer-based geographic information systems, it is possible to generate the view of known terrain as long as the location of the observer is known. Modern GPS systems make it possible to know the location of an airplane with high accuracy. Thus, the combination of the two systems makes it possible to render on a computer display the terrain that may not be visible due to adverse environmental conditions (e.g. fog, rain). This is the basis for NASA's Synthetic Vision System or SVS. That is, an SVS is essentially a computer generated display designed to provide the pilot with information that augments the out-the-window view, to better enable the pilot to fly safely, at low levels, through traffic, around terrain, and in low visibility conditions.

That is, a Synthetic Vision System (SVS) is essentially a computer generated display designed to provide the pilot with information that augments the out-the-window view, to better enable the pilot to fly safely, at low levels, through traffic, around terrain, and in low visibility conditions. Experiments conducted at NASA Ames Research Center by NASA and Monterey Technologies Inc. were performed to investigate the potential positive and negative effects of augmenting a cockpit with a prototype SVS display (Goodman, et al., 2003). A similar SVS is also under evaluation by standard field-trial methodology (Prinzel, et al., 2002); we see these two approaches as complimentary.

This report describes our efforts to date in applying computational cognitive modeling techniques to the problem of evaluating SVS technology.

1.2 Report Outline

We first provide a brief discussion in Section 2 of our theoretical and methodological perspective on this problem, and a brief discussion of lessons learned from our research prior to this effort modeling pilot performance in the T-NASA Taxi Navigation simulations. We present some of those lessons learned that have provided concrete implications for our current SVS modeling effort. Section 3 briefly summarizes the methods and results, including analyses of the eye movement data, that form the empirical basis for the modeling work. As the conclusion of this section demonstrates, these analyses were quite informative in terms of identifying the particular phenomenal that would serve as the focus for our closed-loop, ACT-R modeling.

Section 4 of the report describes the modeling effort. This includes the additional (top-down or theoretical) sources of information guiding our modeling approach, including task analyses, the ACT-R approach, subject matter expert (SME) input, and extant theory of visual attention allocation from both engineering (e.g., Senders, 1964) and psychological (e.g., Wickens, 2002), perspectives. It also includes a description of how that information led us to focus on particular aspects of modeling pilot performance, as well as our detailed implementation approach.

Section 5 describes the performance of the model and provides validation by comparing the model's performance and the empirical data at multiple levels of analysis. Section 6 describes [Alex's work on the Senders stuff]. We discuss how we intend to finalize this work under the current program in Section 7 as well as discussing how this work might be extended in the future. Finally, Section 8 provides a discussion of implications of the current research and lessons learned along the way. References follow this section.

2. Theoretical and Methodological Perspective

2.1 Theoretical Orientation

Aviation incident and accident investigations often find both cognitive and environmental sources of human error. Environmental sources include factors such as flawed interface design, confusing automation, and unexpected weather conditions. Improved environmental design, such as the use of the Synthetic Vision Systems (SVS) that are the subject of our current research, often provide important leverage for reducing error and improving human performance. On the other hand, cognitive sources underlying the effectiveness and efficiency of performance include factors such as situation awareness, procedural compliance or non-compliance, and crew coordination. Many if not most significant incidents and accidents result from some combination of both cognitive and environmental factors. In fact, in a highly proceduralized domain such as aviation, with pilots who are highly trained and well-motivated, accidents rarely result from either environmental or cognitive causes alone. Training and experience are often sufficient to overcome even the most confusing interface designs, and the environment is often sufficiently redundant, reversible, and forgiving so that the vast majority of cognitive slips and mistakes have no serious consequences. Most highly consequential incidents and accidents result only when both environmental and cognitive factors collectively conspire to produce disaster.

Introducing new technology is a common approach to trying to reduce either the frequency, severity, or consequences of less-than-perfect pilot performance. Human performance modeling associated with evaluating the impact of technological interventions therefore requires giving consideration to both cognitive and environmental issues. This report describes the progress made to date on a research project in which dynamic, closed loop cognitive-environmental modeling, or more specifically pilot-vehicle-airport modeling, is currently being performed in order to shed light on both the positive and potential negative effects on the introductions of SVS technology in the commercial airline cockpit. Our current modeling consists of integrating a pilot model developed within the ACT-R cognitive architecture (Anderson, et al., in press) with a commercial, off-the-shelf (COTS; Bowers and Jentsh, 2001) model of aircraft dynamics and the Santa Barbara airspace and airport which served as the basis for part-task experimentation. The overall objective of the NASA program in which we are participating is to develop computational human performance models with the predictive ability to aid designers and analysts in identifying likely vulnerabilities in human-machine performance in aviation.

Our current SVS modeling is actually the second stage in a longer term effort to meet this goal. Our research in the previous year focused on modeling to understand the causes of taxiing errors in a NASA simulation involving taxi-to-gate scenarios at a simulation of Chicago O'Hare (ORD) airport. To set the stage for the rest of this report, we briefly discuss the lessons learned from that effort which motivated our approach to the current problem. These lessons helped us get a bit of a head start regarding the selection of an initial modeling architecture for SVS modeling: a closed-loop, dynamically interacting dyad comprised of an ACT-R model of cognition and a commercially available aircraft-airport simulation package.

2.2 Lessons Learned From Phase 1

As we learned in our taxi modeling research, it is a nontrivial matter to apply scientific models of cognition, developed and validated primarily with psychological laboratory data, to applied contexts such as human-machine performance in aviation. Specific challenges include the following.

2.2.1 Communication Between Cognition and the Outside World

Experimental tasks are typically carefully designed in such a way that the inputs to, and outputs of, the cognitive system are readily identifiable. This is largely done by making the perceptual and motor demands associated with cognitive experimentation relatively trivial. Unfortunately, the perceptual and motor demands associated with aviation cognition can be extensive. This problem surfaced in Phase 1 research in the difficulties associated with coupling the ACT-R/PM model with the visual scene database, and also to some extent, with the aircraft model. The latter was not a severe problem because motor outputs could be considered relatively discrete in this instance, consisting of distinct settings of the throttle and brake. The SVS scenarios are similar in this regard although, as will be seen below, we have dedicated a good bit of effort to couple the perceptual mechanisms of ACT-R with both the cockpit and external scene provided by the flight simulator with which ACT-R is intended to interact.

2.2.2 Modeling Environmental Objects and Dynamics

As Phase 1 research clearly demonstrated, achieving a reasonable model of pilot cognition in dynamic, interactive contexts depends heavily on the availability of reasonable models of the visual, physical, and controlled environment, as well as its dynamics. The dynamics of human cognition and behavior is interleaved with, and occurs in concert with, the dynamics of environmental entities that also participate in the functioning of the integrated human-environment system. In our Phase 1 final report, we noted that cognitive modeling software packages can make better and more explicit provisions for representing objects and dynamics in the external environment so facilitate the task of modeling interactive behavior in contexts more complex than the desktop computer. As will be seen below, this issue has again resurfaced as a non-trivial issue in our current SVS modeling, if not theoretically, at least from a technical and data-communication perspective.

2.2.3 Timing Issues

Based on our Phase 1 research we concluded that modeling systems for running human-environment system models should provide separate clocks and processing resources for simulating cognitive and environmental dynamics. We noted that neither is subservient to, nor should be subsumed, by the other, nor should either model have to wait while the other is updating. We additionally noted that the current solution to this problem, in which processing resources are passed back and forth between cognitive and environmental components of the model, will be found to be increasingly unwieldy as more dynamic contexts are modeled. On the basis of this finding from our Phase 1 research, we made an early and explicit decision to implement our ACT-R pilot model and our aircraft/airport model on separate computers, connected by networked resources for 2-way data communication. With this approach, the processing demands associated with mimicking the dynamics of the two components of the total interactive systems do not compete. However, we have learned that neither of the two pieces of software we are using for modeling environments make explicit provisions for such intercomputer communication. As a result, we are currently running both in real-time mode, so that, in essence, the wall clock keeps time for both systems. This is inefficient for ACT-R, which is capable of running at many times real-time, but is perhaps less of an issue for the flight simulation package.

3. Empirical Methods and Results

3.1 Methods

We will provide only a brief overview of the empirical methods since they are covered in some detail in Goodman, Hooey, & Foyle (2003). Pilots were placed in a flight simulator that approximated the instruments and controls of a Boeing-757. The aircraft simulator was linked with a visual database modeling Santa Barbara Municipal Airport and its surrounding terrain. The purpose of these experiments was to collect data characterizing pilot performance and eye movement behavior during the approach and landing phase of flight using with both conventional and augmented displays under both Instrument Meteorological Conditions and Visual Meteorological Conditions. The purpose of these experiments was to collect data characterizing pilot performance and eye movement behavior during the approach and landing phase of flight using with both conventional and augmented displays under both Instrument Meteorological Conditions (IMC) and Visual Meteorological Conditions (VMC) conditions. The experimental plan, due to the cost and time required for studies of this complexity, focused on a limited number of pilots operating across a variety of conditions and treatments.

In the experimental configuration, the pilot could look at multiple displays—also termed sceneplanes, areas of interest (AOIs) and regions of interest (ROIs). These included looking out the window (OTW), the SVS, the primary flight display (PFD), the navigation display (NAV), the mode control panel (MCP) and a display for miscellaneous controls (DMC or simply Controls). The displays were configured as displayed in Figure 1.

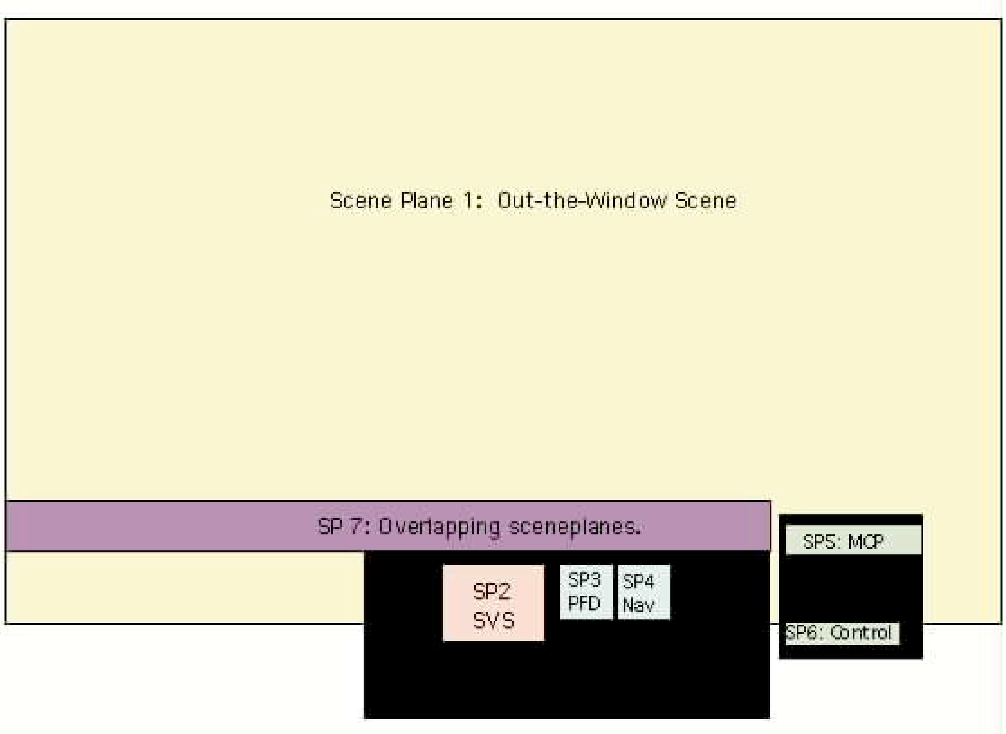

Figure 1. Overview of displays

Due to limitations in the eye-tracking technology, some gaze points were ambiguous and coded as "overlapping" one or more regions, as shown in Figure 1. Additionally, the tracker could fail to generate a valid reading either due to technical failure or the pilot looking somewhere off the defined displays (e.g., looking down at the joystick). These samples were coded as "off."

It is important to note that the SVS was designed to serve as a proxy for looking out the window during conditions where actually doing so would yield little information (e.g., when visibility is poor). Thus, it is reasonable to expect that the SVS would simply take on the role of the window when pilot would otherwise be looking out the window. However, the SVS contains additional overlaid symbology as depicted in Figure 2.

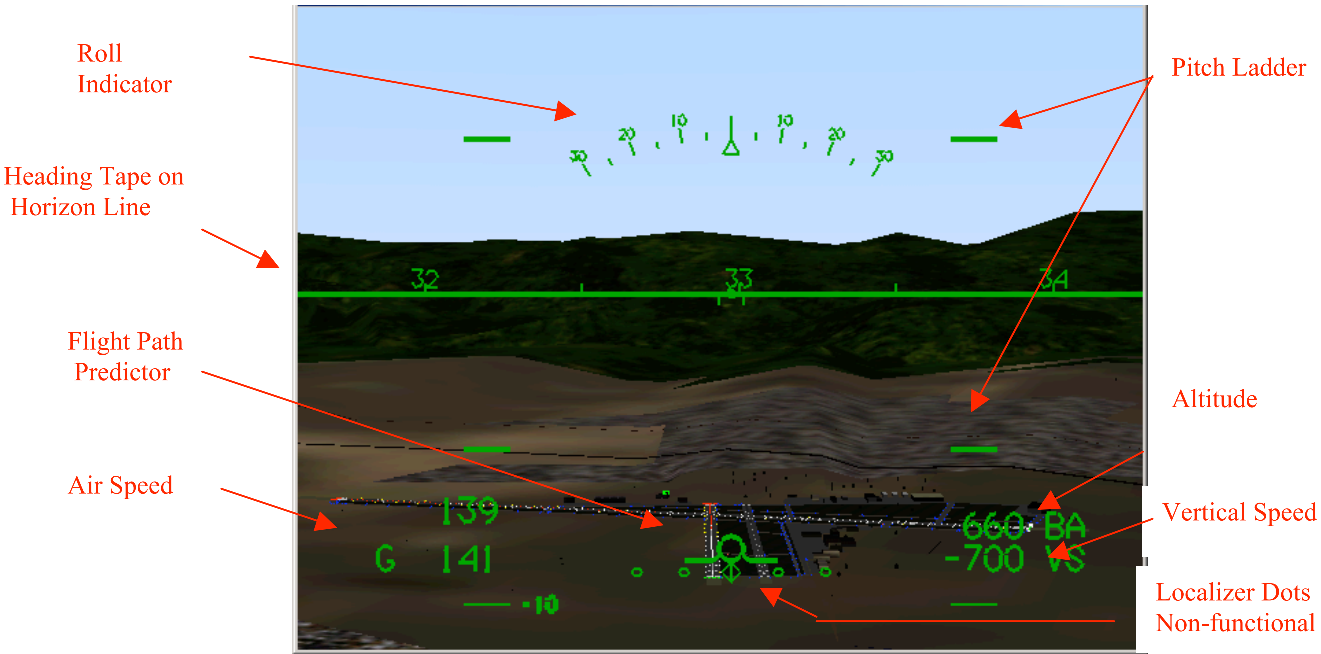

Figure 2. SVS display corresponding to the OTW scene with symbology identified in red.

Thus, one of the central questions to be answered by the empirical study is the extent to which ths SVS actually does serve primarily as a proxy for looking out the window or whether it may sreve additional functions.

The first issue to be considered in this modeling effort is what to model. That is, what variance is there to be explained? In these scenarios and with the limited number of subjects available, there is not a rich corpus of speech or control manipulations, let alone errors, to model. Indeed, even if there were more errors, it is not clear how representative the small sample would be. The sample is also limited in terms of airports, weather conditions, scenarios, etc.

However, there is one extremely rich source of data: the eye movement record. Not only is there a great deal of data with which to work, there is substantial variance to be explained. Furthermore, we believe this eye movement record is the most generalizable component of the data. What we want to know is how the SVS, a visual display, affects pilots' visual behavior. For instance, if some higher-level metric of task performance (such as number of errors made) is not sensitive enough to show an effect of the SVS, this may be because the SVS is ignored by the pilots or because the difference it makes is compensated for by other factors. We can distinguish those cases via the eye-tracking record. Thus, we have focused our empirical analysis on the eye data.

3.2 Eye Movement Analysis and Results

The most striking result of the simulator study is that there was very little impact of the SVS on pilot performance in terms of errors or the quality of the observable decisions made by the pilots. We believe this is due to the fact that for a well-trained and highly-motivated commercial pilot, the approach and landing scenarios, flown primarily by the autopilot, were well within their competence. However, there was one aspect of the pilots' behavior that was significantly impacted, which was their allocation of gaze across the various available displays. Thus, the data analysis will focus entirely on this aspect of the pilots' behavior.

Fixations are distinguished from saccades (rapid voluntary eye movements used to move from one fixation to another) and very small involuntary eye movements of several types that occur during fixation. A "dwell" is defined as the time period during which a fixation or series of continuous fixations remain within a sceneplane or ROI.

Selected Data Analysis
To focus on SVS versus non-SVS cases in similar conditions, our initial data analysis focused solely on the conditions listed in Table 2.

Table 1. Selected conditions for data analysis

Display Configuration		Baseline	SVS
Visibility		IMC	IMC
Approach Event	Nominal Approach (nominal landing)	*Scenario #4*	*Scenario #7*
	Missed Approach (go-around)	*Scenario #5*	*Scenario #9*
	Terrain Mismatch (go-around)	*Scenario #6*	*Scenario #10*

Because we are primarily concerned with the allocation of visual attention, the primary variables of interest are those that index the amount of attention given to each sceneplane. Allocations can be counted in two ways, by count (number) of fixations or by total duration of dwell. Of course, if all fixations have the same duration, the relative proportions spent on each sceneplane will be identical, but this is an empirical question.

Additionally, there are a potentially large number of ways to segment the data. In this presentation we will limit our analysis to the following: we consider only four categories, namely, pilots' baseline eye movements, pilots' SVS versus baseline eye movements, pilots' eye movement by different flight phases, and pilots' eye movements by different approach scenarios.

3.2.1 Baseline Attention Allocation
The first thing to be established is the baseline allocation of attention. If the SVS is to serve as a proxy for looking out the window, it is useful to know how often pilots look out the window in the first place. In addition, since the SVS contains information that can be found on other displays (e.g., altitude, which is also on the PFD), it is important to know how often those other displays are accessed.

Figure 3 shows that overall pilots spent 40.2% of their total fixation dwell time on the PFD, and about 40% fixation duration on NAV. They only spent about 3% looking out the window, about 4.2% at the MCP, and about 6.5% at the Controls. This result suggests that the PFD and NAV displays are the major targets of attention, as they account for about 80% of the fixation dwell time.

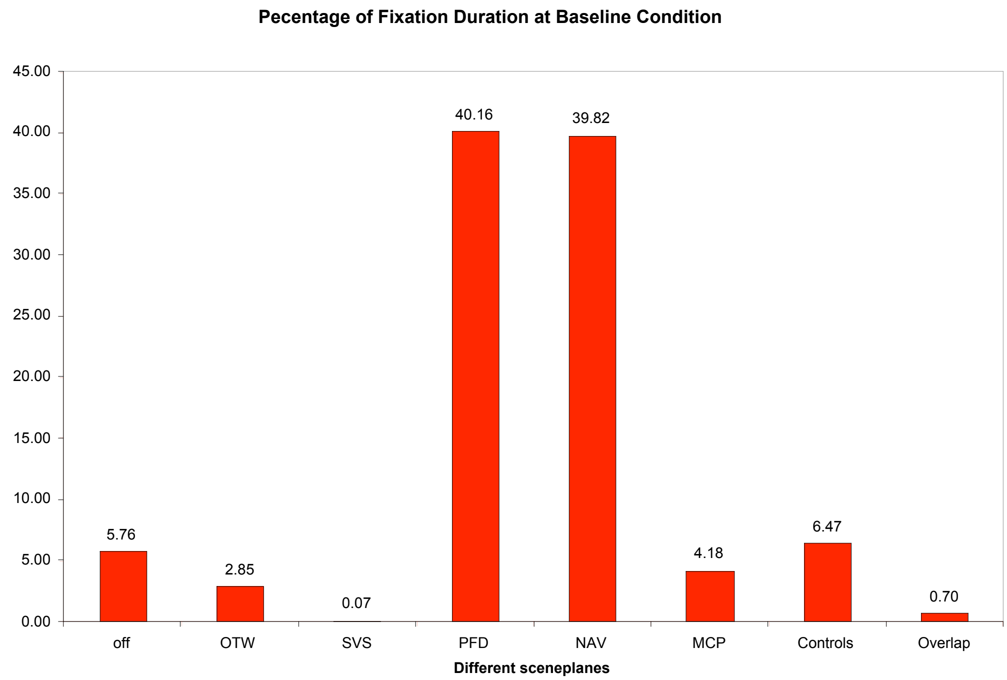

Figure 3. Percentage of fixation dwell time in the baseline condition.

The fixation count measure shows a similar distribution, as depicted in Figure 4. Pilots spent 38.3% of their fixations on the PFD, and another 39.6% of their fixations on NAV, and relatively small amounts of time looking at other displays, most notably less than 3% out the window.

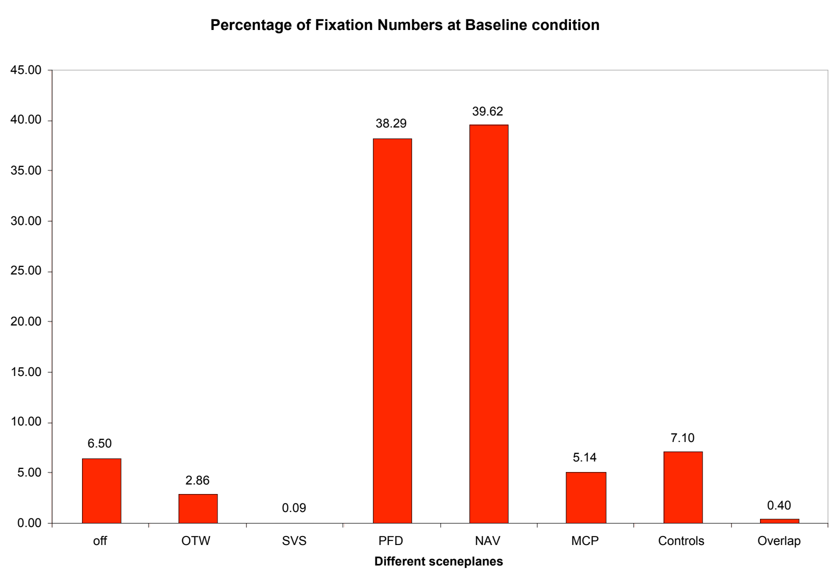

Figure 4. Percentage of fixation count in the baseline condition.

3.2.2 Pilots Eye Movements at SVS versus Baseline Conditions

The obvious next question is how the SVS affects this distribution. Figure 5 shows that the distributions of pilots' fixation dwell time on different sceneplanes are similar for both the SVS and the baseline (No SVS) condition. In the SVS condition, pilots also spent most of their fixation time on the PFD (35.7%) and NAV (29.6%). Pilots also spent 2.6% fixation time looking out the window in the SVS condition. However, one notable feature in the SVS condition is that pilots displayed a large number of fixations (20.2%) on the SVS. This clearly shows that pilots did look at the SVS display quite frequently.

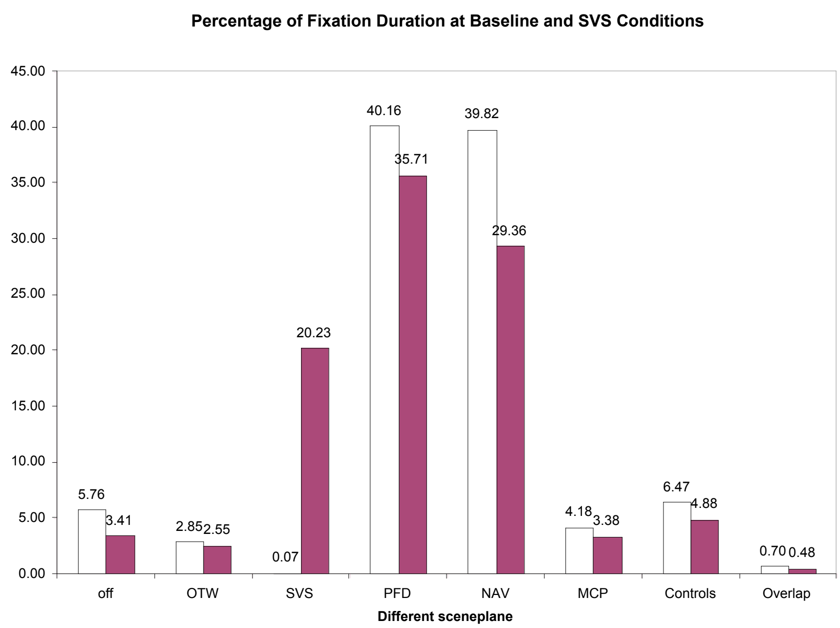

Figure 5. Percentage of total fixation dwell time on each sceneplane in the baseline (light bars) and the SVS (dark bars) conditions.

Again, the distributions for fixation count follow those for time. Figure 6 shows the graph for fixation count for SVS and baseline conditions, and yields about the same results.

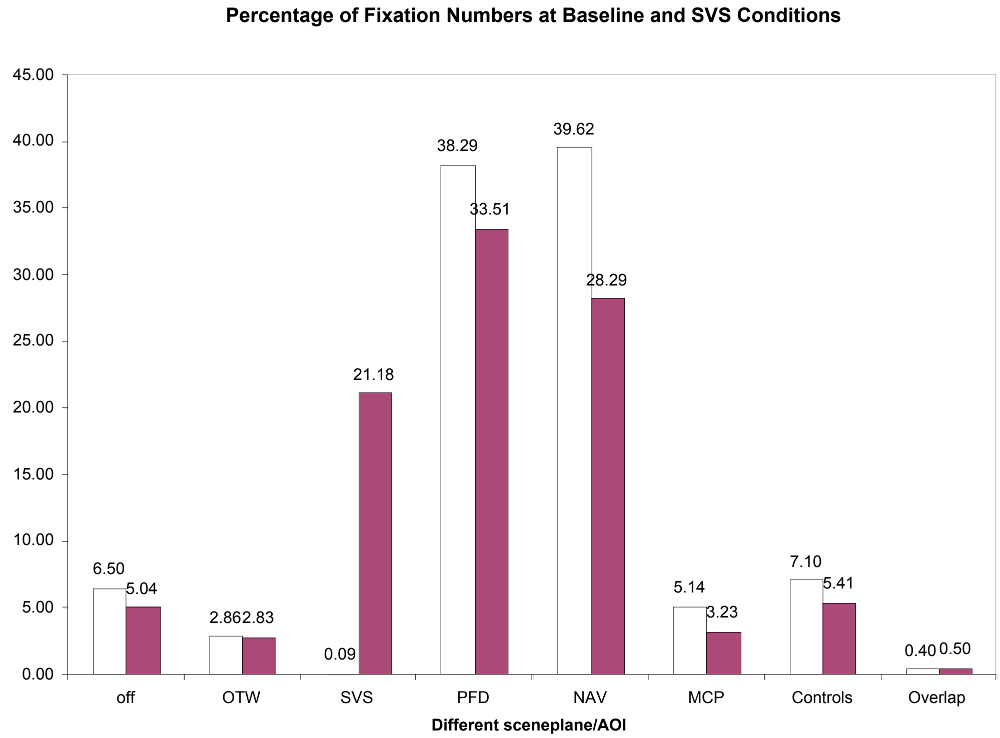

Figure 6. The percentage of fixation count on each sceneplane in the baseline (light bars) and the SVS (dark bars) conditions.

Clearly, pilots spend a fair proportion of their gaze on the SVS. An interesting question we wanted to address is from where they "stole" these gazes. Namely, they must reduce some portion of fixations associated with other sceneplanes. A natural suspect for the location from which fixations would be stolen was the OTW display, due to the inherent redundancy between the perceptual information gained from the SVS and OTW displays.

Surprisingly, however, the fixation dwell time and frequencies on different sceneplanes under SVS versus baseline (No SVS) conditions, as shown in the above two figures, shows no obvious difference in the amount of gaze directed out the window. Instead, it appears the SVS is associated with a reduction in the amount of gaze directed at the PFD and NAV displays. Thus, the SVS display was drawing attention away from other sources of information from within the cockpit, it was not acting as a "substitute" for the information provided by the OTW display.

We find this result both counterintuitive and quite interesting. The underlying rationale behind the SVS display is that it will act as a substitute for the OTW display when the information obtainable from the latter is degraded. Instead, these data seem to indicate that it did not act as a substitute source of environmental information (at least for these pilots). And as a possibly unintended result of the presence of the SVS, less attention was paid to other displays.

3.2.3 Analysis by Phase of Flight
All the above analyses combined all the flight phases together. During different flight phases, however, pilots may have different needs for different information. We therefore decided to break down the analysis of data by flight phase. The fight phases were defined as follows:

Phase 1. Start to Initial Approach Fix (IAF)
Phase 2. Initial Approach Fix (IAF) to Final Approach Fix (FAF)
Phase 3. Final Approach Fix (FAF) to Decision Height (DH)
Phase 4. Decision Height (DH) to End

Because the percentage of fixation dwell time and fixation counts on different sceneplanes provides similar information, to simplify the analysis we only dealt with the percentage of fixation dwell time on different sceneplanes in all subsequent analysis. Again, the natural starting point is the baseline condition, which is presented in Figure 7. Note how the use of the PFD increases as the flight moves on, and the sharp drop in the use of NAV and sharp increase in OTW gazes in phase 4.

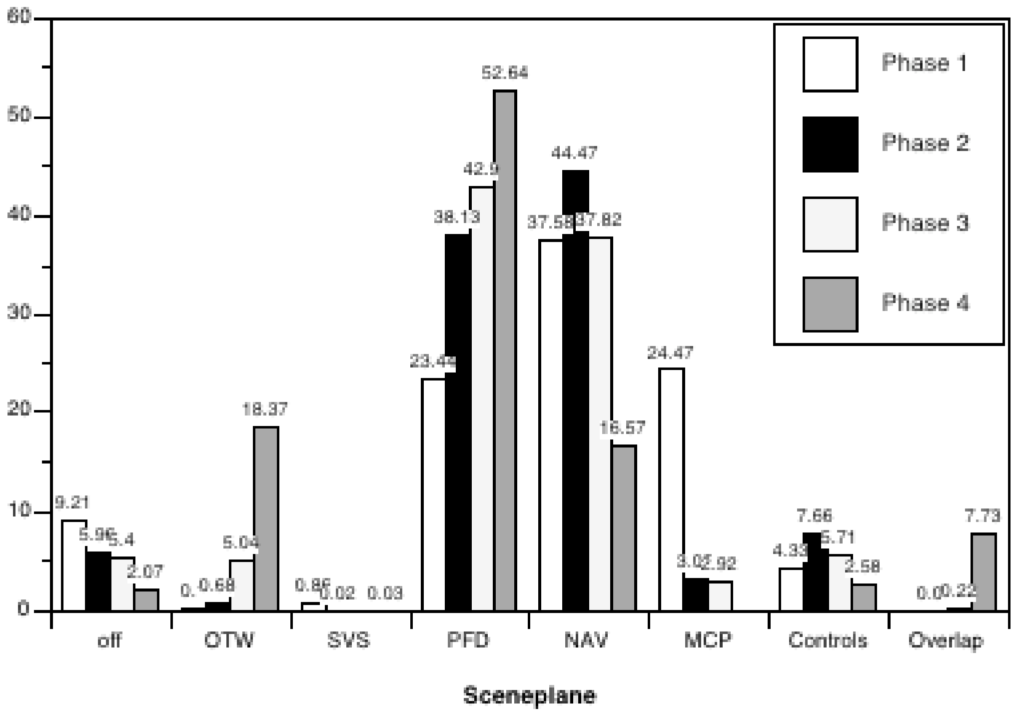

Figure 7. Percentage of fixation dwell time in the baseline condition by different flight phases.

Because pilots so rarely look out the window in phase 1, one might expect little use of the SVS in this phase. In fact, pilots do look a little at the SVS in this phase but overall the allocation of gaze is not substantially different in baseline vs. SVS in this condition, as shown in Figure 8.

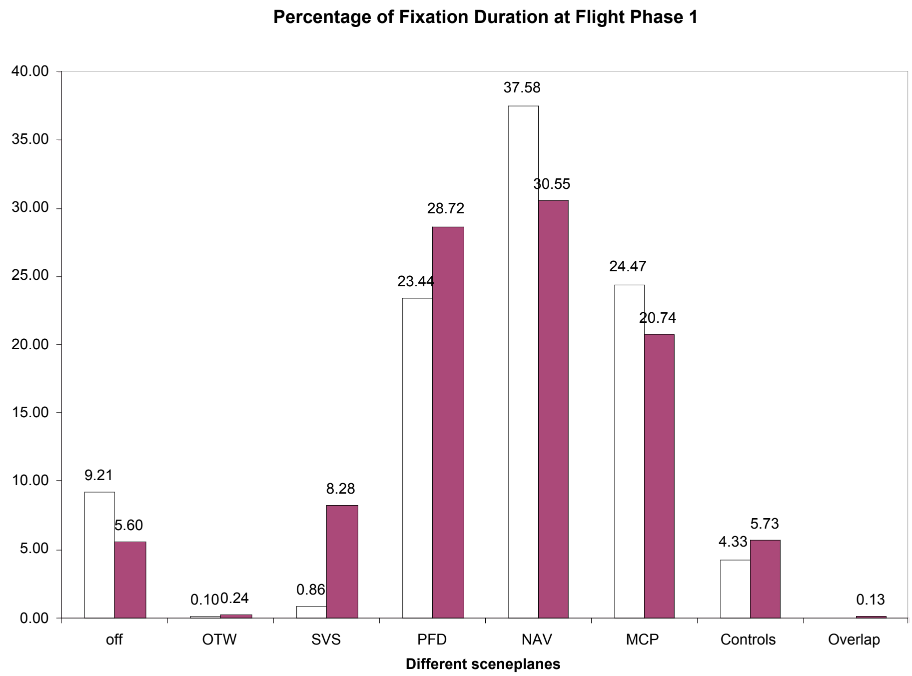

Figure 8. Percentage of fixation dwell time in the baseline (light bars) and the SVS (dark bars) conditions for flight phase 1.

Phase 2 shows an increase in SVS use over phase 1, with most of the gaze time being "stolen" from PFD and NAV. (See Figure 9.) Note that in this phase of flight, there is virtually no use of OTW in either baseline or SVS conditions, but the SVS is still used when present. We believe

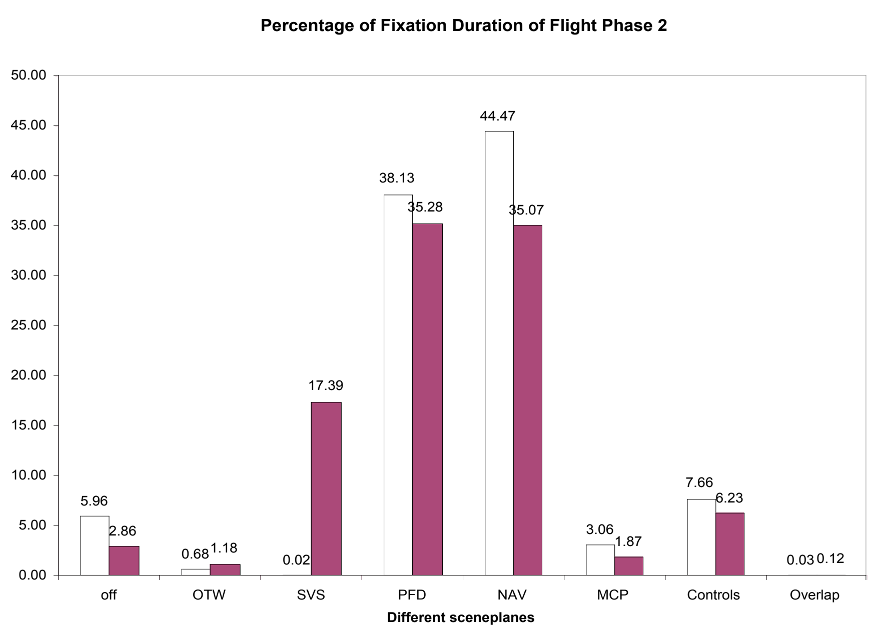

Figure 9. Percentage of fixation dwell time in the baseline (light bars) and the SVS (dark bars) conditions for flight phase 2.

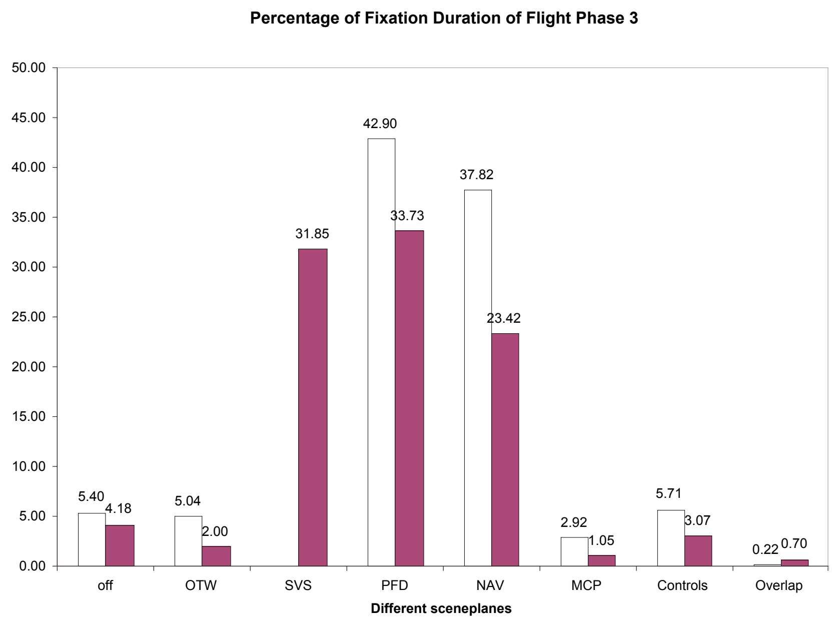

Figure 10*.* Percentage of fixation dwell time in the baseline (light bars) and the SVS (dark bars) conditions for flight phase 2.

this is due primarily to the symbology overlaid on the SVS. This trend continues through flight phase 3, shown in Figure 10. Note substantial use of SVS and decrease in use of PFD and NAV. Flight phase 4 shows the largest reduction of OTW looking as a result of the SVS (Figure 11), but this reduction in OTW gaze only accounts for about a third of the total SVS time, which appears to come primarily from the PFD. Overall, SVS gaze accounts for a fairly substantial proportion of total gaze time in phases 3 and 4, but this is not by simple reduction of OTW gaze. Instead, pilots seem to borrow gaze from the PFD.

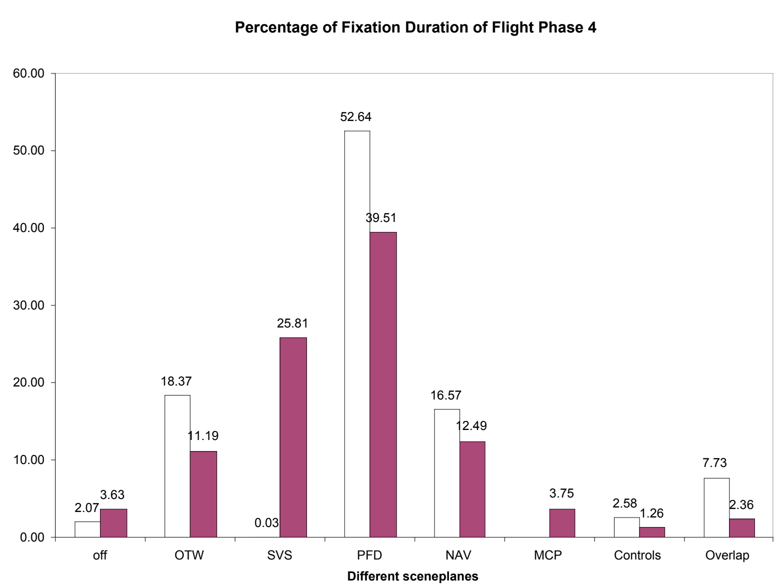

Figure 11*.* Percentage of fixation duration at the baseline and the SVS condition of flight phase4.

3.2.4 Late Phases of Flight Broken Down by Approach Events

Since the SVS versus non-SVS differences showed up most prominently in the final phases of flight, the next step in the data analysis focused on phase 3 and phase 4, further breaking down the data by different approach scenarios, namely, Nominal Landing, Missed Approach and Terrain Mismatch. This is depicted in Figures 12–14.

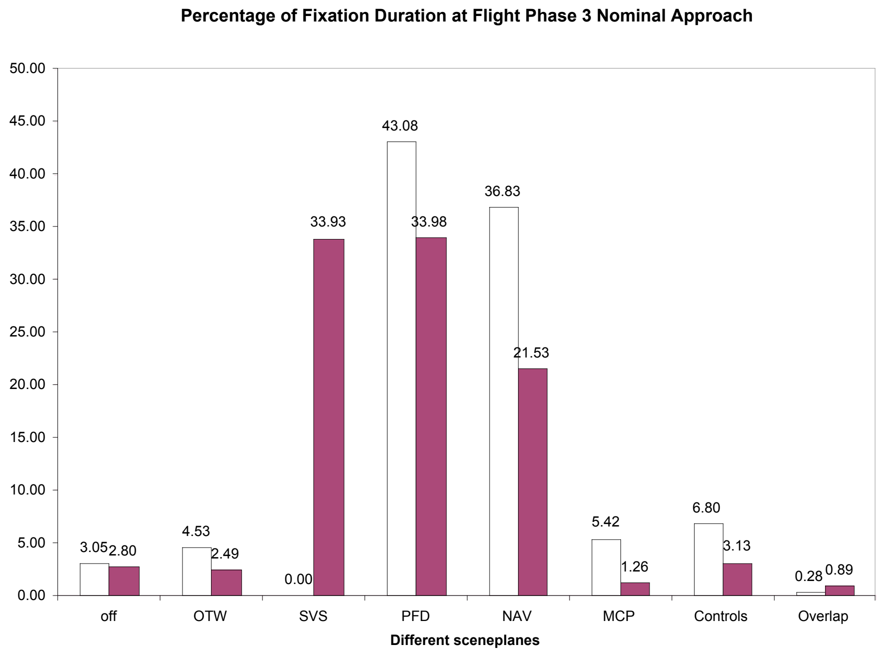

Figure 12. Percentage of fixation dwell time in baseline (light bars) and SVS (dark bars) conditions during flight phase 3 for the Nominal Approach scenarios.

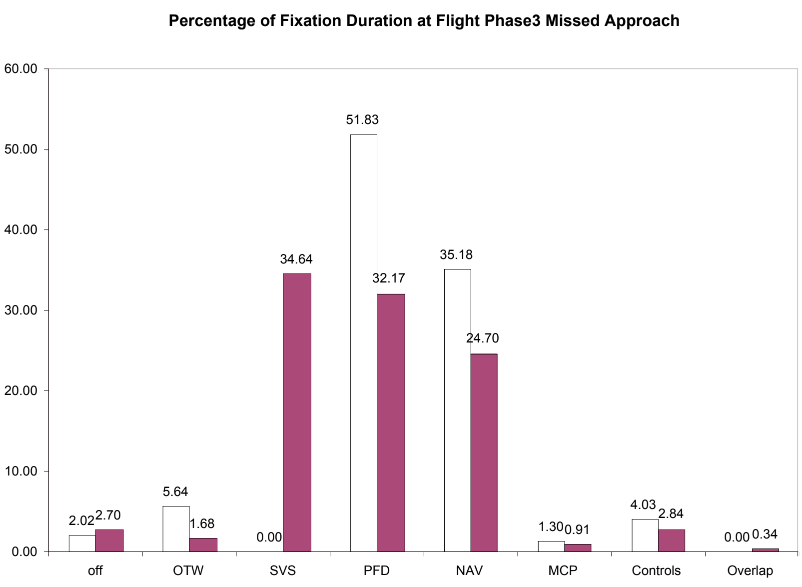

Figure 13. Percentage of fixation dwell time in baseline (light bars) and SVS (dark bars) conditions during flight phase 3 for the Missed Approach scenarios.

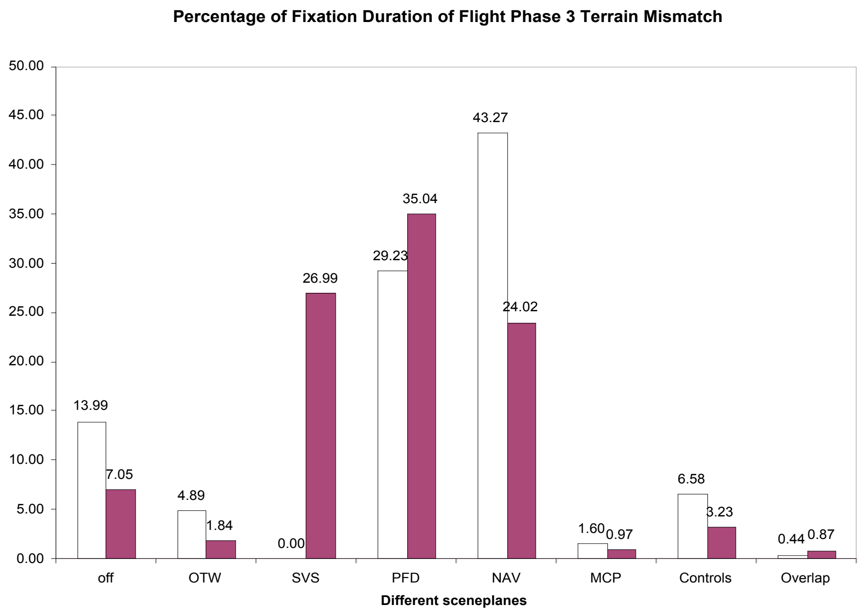

Figure 14. Percentage of fixation dwell time in baseline (light bars) and SVS (dark bars) conditions during flight phase 3 for the Terrain Mismatch scenarios.

Note how similar the OTW and SVS usage is for the three scenarios during phase 3. This is in stark contrast to phase 4, presented in Figures 15–17.

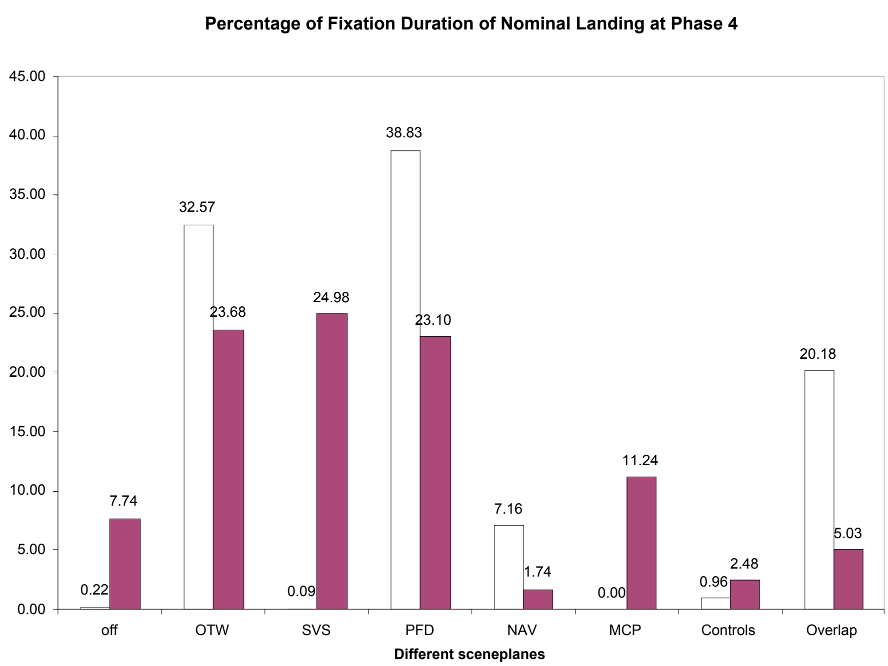

Figure 15. Percentage of fixation dwell time in baseline (light bars) and SVS (dark bars) conditions during flight phase 4 for the Nominal Approach scenarios.

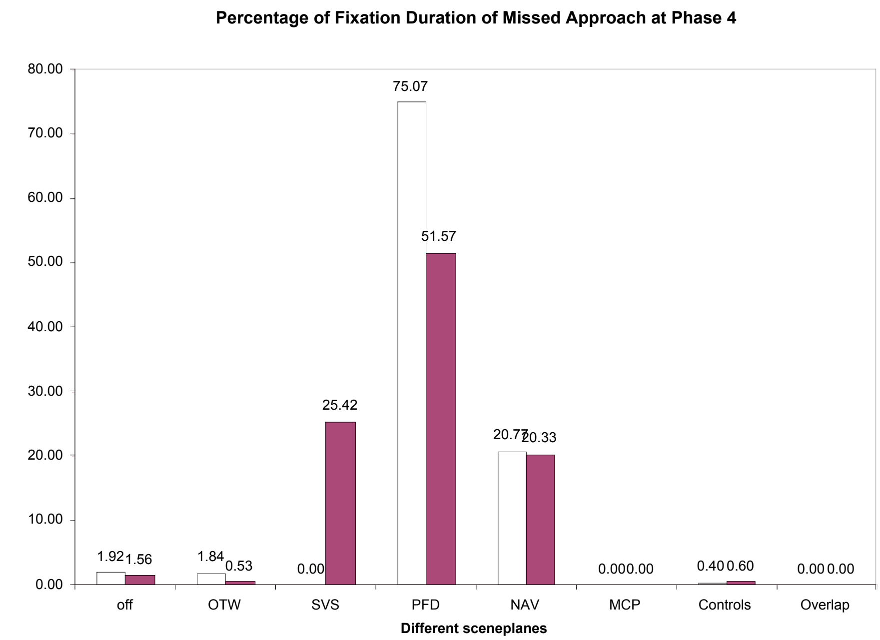

Figure 16. Percentage of fixation dwell time in baseline (light bars) and SVS (dark bars) conditions during flight phase 4 for the Missed Approach scenarios.

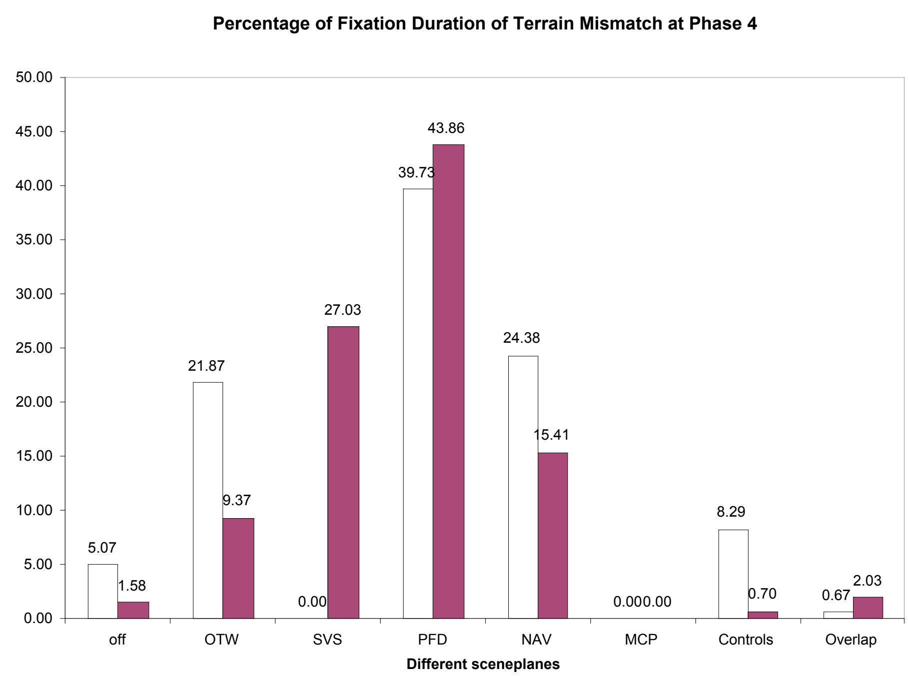

Figure 17. Percentage of fixation dwell time in baseline (light bars) and SVS (dark bars) conditions during flight phase 4 for the Terrain Mismatch scenarios.

The differences here are striking. In the nominal landing, pilots looked out the window quite a lot, even when the SVS was available. SVS use was certainly evident for the nominal approach in phase 4, but as seen in other graphs, this seems to come at least in part through less allocation of gaze to the PFD and possibly NAV. However, in the missed approach, OTW looking fell to virtually zero, both with and without SVS. Note that SVS use was roughly similar in nominal and missed approach, but in missed approach, this appears to have come almost entirely at the expense of gaze at the PFD.

For terrain mismatch conditions, the SVS use was again about the same (about a quarter of the total time), but this time it was *not* due to a reduction in looking at the PFD. Thus, use of the SVS is clearly conditioned on phase of flight and scenario. In addition, the SVS is evidently not simply a proxy for looking out the window. In particular, the SVS appears to serve many of the informational functions of the PFD, and perhaps also NAV, though to a lesser extent. This has other implications that will be discussed later.

3.2.5 Analysis of Gaze Transitions

The presence of the SVS may change not only how much the pilots look at various displays, but the strategies they use for acquiring information from their visual world. This should be reflected in changes to the order in which various pieces of information are acquired. That is, if fixation *n* is on the PFD, where is the most likely location for fixation *n*+1? Is this affected by the presence of the SVS, and if so, how?

We analyzed the data provided by NASA to in order to answer such questions. This is difficult data to visualize so it will be presented in tabular form. The values in each cell of all such tables is computed by counting each transition of fixation, which can be from one display to another or can be from a display to itself, and then dividing by the total number of such transitions, yielding a proportion (or probability) of occurrence of each transition. (Thus, the sum of all table values will be 1.) Values that would generate table entries of less than 0.01 have been omitted for clarity. Consider Table 2, which displays the relevant table aggregating across all phases of flight for the no-SVS condition. What this table shows is that transitions from the PFD to the PFD accounted for 25% of all transitions. This is certainly reasonable since a fairly high percentage of the fixations were on the PFD, so many transitions would involve the PFD.

Table 2. Transition matrix for no-SVS condition across all phases of flight, showing proportion of gaze transitions from (vertical) a display region to (horizontal) a display region.

	off	**OTW**	**SVS**	**PFD**	**NAV**	**MCP**	**DMC**	**overlap**
off	0.03	–	–	0.01	0.02	0.01	0.01	-
OTW	–	0.02	–	–	–	–	–	-
SVS	–	–	–	–	–	–	–	-
PFD	0.01	–	–	0.25	0.1	–	–	-
NAV	0.02	–	–	0.1	0.27	–	0.01	-
MCP	–	–	–	–	–	0.04	–	-
DMC	0.01	–	–	–	0.01	–	0.05	-
overlap	–	–	–	–	–	–	–	–

Note that in general, the table is dominated by the diagonal; that is, most fixations are followed by another fixation on the same display. This suggests a kind of inertia in the pilots's attention allocation—once a display is fixated, it tends to be fixated again. Consider Table 3, which presents the same data but this time with the SVS present.

Table 3. Transition matrix for SVS condition across all phases of flight, showing proportion of gaze transitions from (vertical) a display region to (horizontal) a display region

	off	OTW	SVS	PFD	NAV	MCP	DMC	overlap
off	0.02	–	–	0.01	0.01	–	0.01	–
OTW	–	0.02	–	–	–	–	–	–
SVS	–	–	0.17	0.02	0.01	–	–	–
PFD	0.01	–	0.02	0.22	0.08	–	–	–
NAV	0.01	–	0.01	0.08	0.18	–	–	–
MCP	–	–	–	–	–	0.02	–	–
DMC	–	–	–	–	–	–	0.04	–
overlap	–	–	–	–	–	–	–	–

Again, the diagonal dominates, but now the SVS figures in substantially. Not only are there many SVS-to-SVS transitions, but there are also transitions in and out of the SVS, primarily to and from the PFD and NAV. Of course, this is collapsing across all four phases of flight. Since total attention allocation is so different across phases of flight, it makes sense to break this down by phase of flight as well. Table 4 presents the transition matrix for flight Phase 1 without SVS, and Table 5 presents it with SVS. Again, the diagonal dominates in both cases. The presence of the SVS clearly causes a decrease in repeat visitations to the NAV display in particular.

Table 4. Transition matrix for no-SVS condition for Phase 1, showing proportion of gaze transitions from (vertical) a display region to (horizontal) a display region

	off	OTW	SVS	PFD	NAV	MCP	DMC	overlap
off	0.06	–	–	0.01	0.02	0.02	–	–
OTW	–	–	–	–	–	–	–	–
SVS	–	–	0.01	–	–	–	–	–
PFD	0.01	–	–	0.15	0.07	–	–	–
NAV	0.02	–	–	0.08	0.24	0.02	–	–
MCP	0.02	–	–	–	0.02	0.19	0.01	–
DMC	–	–	–	–	0.01	–	0.02	–
overlap	–	–	–	–	–	–	–	–

Table 5. Transition matrix for SVS condition for Phase 1, showing proportion of gaze transitions from (vertical) a display region to (horizontal) a display region

	off	OTW	SVS	PFD	NAV	MCP	DMC	overlap
off	0.03	–	–	0.01	0.01	0.02	–	–
OTW	–	–	–	–	–	–	–	–
SVS	–	–	0.06	0.01	0.01	–	–	–
PFD	0.01	–	0.01	0.18	0.07	0.01	–	–
NAV	0.01	–	0.01	0.07	0.16	0.01	–	–
MCP	0.02	–	–	0.01	0.01	0.18	–	–
DMC	–	–	–	0.01	0.01	–	0.04	–
overlap	–	–	–	–	–	–	–	–

Tables 6 and 7 present the transition matrices for flight Phase 2. Here again, transitions to and from the NAV display are reduced with the addition of the SVS. The PFD, however, is largely unaffected. Consistent with the dwell time numbers, OTW is essentially unused by the pilots both with and without the SVS, but the SVS is used extensively when present. This clearly indicates that the SVS is not simply a proxy for OTW.

Table 6. Transition matrix for no-SVS condition for Phase 2, showing proportion of gaze transitions from (vertical) a display region to (horizontal) a display region

	off	OTW	SVS	PFD	NAV	MCP	DMC	overlap
off	0.03	–	–	0.01	0.02	0.01	0.01	–
OTW	–	–	–	–	–	–	–	–
SVS	–	–	–	–	–	–	–	–
PFD	0.01	–	–	0.22	0.11	–	–	–
NAV	0.02	–	–	0.11	0.32	–	0.01	–
MCP	–	–	–	–	–	0.03	–	–
DMC	0.01	–	–	0.01	0.01	–	0.05	–
overlap	–	–	–	–	–	–	–	–

Table 7. Transition matrix for SVS condition for Phase 2, showing proportion of gaze transitions from (vertical) a display region to (horizontal) a display region

	off	OTW	SVS	PFD	NAV	MCP	DMC	overlap
off	0.02	–	–	0.01	0.01	–	0.01	–
OTW	–	0.01	–	–	–	–	–	–
SVS	–	–	0.14	0.02	0.01	–	–	–
PFD	0.01	–	0.02	0.22	0.08	–	–	–
NAV	0.01	–	0.01	0.08	0.22	–	–	–
MCP	–	–	–	–	–	0.02	–	–
DMC	0.01	–	–	0.01	–	–	0.05	–
overlap	–	–	–	–	–	–	–	–

Tables 8 and 9 present the transition matrices for Phase 3. In this phase, pilots without SVS start to make more transitions to and from the window. However, with the SVS, this is reduced and SVS usage becomes more prominent. Again, this seems to be at the expense of the NAV display.

Table 8. Transition matrix for no-SVS condition for Phase 3, showing proportion of gaze transitions from (vertical) a display region to (horizontal) a display region

	off	OTW	SVS	PFD	NAV	MCP	DMC	overlap
off	0.02	–	–	–	0.02	–	0.01	–
OTW	–	0.04	–	0.01	0.01	–	–	–
SVS	–	–	–	–	–	–	–	–
PFD	0.01	0.01	–	0.26	0.11	–	–	–
NAV	0.02	–	–	0.11	0.25	–	–	–
MCP	–	–	–	–	–	0.03	–	–
DMC	0.01	–	–	–	0.01	–	0.04	–
overlap	–	–	–	–	–	–	–	–

Table 9. Transition matrix for SVS condition for Phase 3, showing proportion of gaze transitions from (vertical) a display region to (horizontal) a display region

	off	OTW	SVS	PFD	NAV	MCP	DMC	overlap
off	0.02	–	0.01	0.01	0.01	–	–	–
OTW	–	0.01	–	–	–	–	–	–
SVS	0.01	–	0.26	0.03	0.01	–	–	–
PFD	0.01	–	0.03	0.21	0.06	–	–	–
NAV	0.01	–	0.02	0.06	0.13	–	–	–
MCP	–	–	–	–	–	0.01	–	–
DMC	0.01	–	–	–	–	–	0.03	–
overlap	–	0.01	–	–	–	–	–	–

Finally, Tables 10 and 11 present the transition matrices for Phase 4. These are somewhat similar to Phase 3 but the OTW region is even more involved. The SVS is again clearly used if present, but seems to take a lot of action away from both the PFD and NAV in this case.

Table 10. Transition matrix for no-SVS condition for Phase 4, showing proportion of gaze transitions from (vertical) a display region to (horizontal) a display region

	off	OTW	SVS	PFD	NAV	MCP	DMC	overlap
off	0.01	–	–	–	0.01	–	–	–
OTW	–	0.07	–	0.03	–	–	–	0.01
SVS	–	–	–	–	–	–	–	–
PFD	–	0.03	–	0.43	0.09	–	–	0.01
NAV	0.01	–	–	0.09	0.14	–	–	–
MCP	–	–	–	–	–	–	–	–
DMC	–	–	–	–	–	–	0.01	–
overlap	–	0.01	–	0.02	–	–	–	0.01

Table 11. Transition matrix for SVS condition for Phase 4, showing proportion of gaze transitions from (vertical) a display region to (horizontal) a display region

	off	OTW	SVS	PFD	NAV	MCP	DMC	overlap
off	0.01	–	0.01	0.01	–	0.01	–	–
OTW	0.01	0.07	0.01	0.01	–	–	–	0.01
SVS	0.01	0.01	0.21	0.03	0.01	–	–	0.01
PFD	0.01	0.01	0.02	0.28	0.06	–	–	0.01
NAV	0.01	–	0.01	0.05	0.05	–	–	–
MCP	–	–	–	–	–	0.01	–	–
DMC	–	–	–	–	–	–	0.01	–
overlap	–	0.02	–	–	–	–	–	–

One thing to notice is that across all phases of flight, the matrices tend to be highly symmetrical. This is not what would be produced by a simple systematic strategy such as scanning the displays left-to-right. And in all phases of flight, the displays that receive the bulk of the fixations are the PFD and the NAV, and the SVS when present. All three of these appear to act as if they have "gravity"; once a transition is made to one of these regions odds are the next several transitions will be made to the same region. This suggests that pilots are not simply glancing at these displays, but are making repeated fixations on them. The MCP and DMC, however, seem to have this property somewhat less.

3.3 Discussion of Empirical Results

The central results in the empirical study do not concern pilot performance as measured by more traditional measures like time to complete tasks or errors, since these were few errors and other performance was generally within accepted parameters. Based on measures like those, the effects of the SVS were negligible. However, the adding the SVS is a significant change in the visual environment of the pilots. So, while by some measures, performance was not changed by the SVS, but by measures of visual performance, the SVS had considerable impact on pilot performance.

These effects were not necessarily what one might have anticipated from the stated goals of the SVS technology, aiding pilots in situations where the actual out-the-window view is obscured. If that were the case, one might expect attention to be allocated to the SVS only in place of where it had previously been allocated to looking out the window. However, this was quite clearly not what happened. The SVS was used extensively by pilots even in cases where they almost never looked out the window; in this case, this is Phases 1 and 2. It should be noted that there is very little to see out the window that is task-relevant in Phases 1 and 2. Similarly, there is little task-relevant information in the SVS rendering of what should be visible out the window. However, there is a great deal of task-relevant information on the SVS in the symbology that is overlaid on the display.

Pilots appear to be making use of this information with little change in overall scanning strategy, as revealed by the transition matrices. The most task-relevant and information-rich displays, the PFD and the NAV, tend to hold gaze for multiple fixations. Also, there is little evidence for some kind of top-down direction to the pilots' scanning; transitions from region A to region B tend to be mirrored by B to A transitions. The addition of the SVS changes this very little, other than that it gets a substantial proportion of the fixations and also acts as an attractor. Pilots are not using the SVS for quick glances and they are not using the SVS in a way which is substantially different than how they are using the PFD or the NAV displays.

What clearly needs explanation is how or why this might occur. We have attempted to address this question through construction of an ACT-R-based computational simulation.

4. Description of the Modeling Effort

A major focus of our modeling effort is to reproduce the trends observed on the basis of the detailed analyses presented in Section 4. We believe that to do so in a manner consistent with one of the major lessons learned from our Phase 1 research on taxi modeling, we must focus on the specification of cognition, the environment, and their interaction at a very fine grain of detail. We believe that this is especially crucial for modeling performance in a dynamic, interactive task. Highly detailed data and task analyses are critical, particularly for a cognitive architecture that operates at a very fine grain of temporal resolution, such as ACT-R. Thus, our focus has been on laying an appropriate foundation for the modeling effort. Our preference has been to eschew shortcuts for the promise of higher fidelity. While this has slowed certain aspects of our progress, we believe this is now paying off.

4.1 Sources of Knowledge and Constraint

Our approach has been to try to understand the major sources of both insight and constraint in generating our models. In addition to the detailed eye-movement analyses presented above, we have identified three additional sources of information and constraint. These have included:

4.1.1 Task Analysis

Besides data analysis, our first order of business was to try to understand the task at a detailed level. This is relatively challenging for this task because there is little overt action taken by the pilots in these scenarios; it appears on the surface to be primarily a supervisory control task, at least until the pilot takes manual control. However, the task is more complex than just that. To understand it, we have relied on three primary sources of information: the task analysis information collected and supplied by NASA Ames (Keller, Leiden, & Small, 2003); other related work in the human factors of aviation; and conversations with our subject matter expert (SME). We have synthesized these into the ACT-R formalism. An example of some of the resulting control structure appears in Figure 28. It should be noted that this control structure is more refined than the one which was described in the 2003 progress report, and we believe better reflects the actual organization of pilot knowledge in this domain.

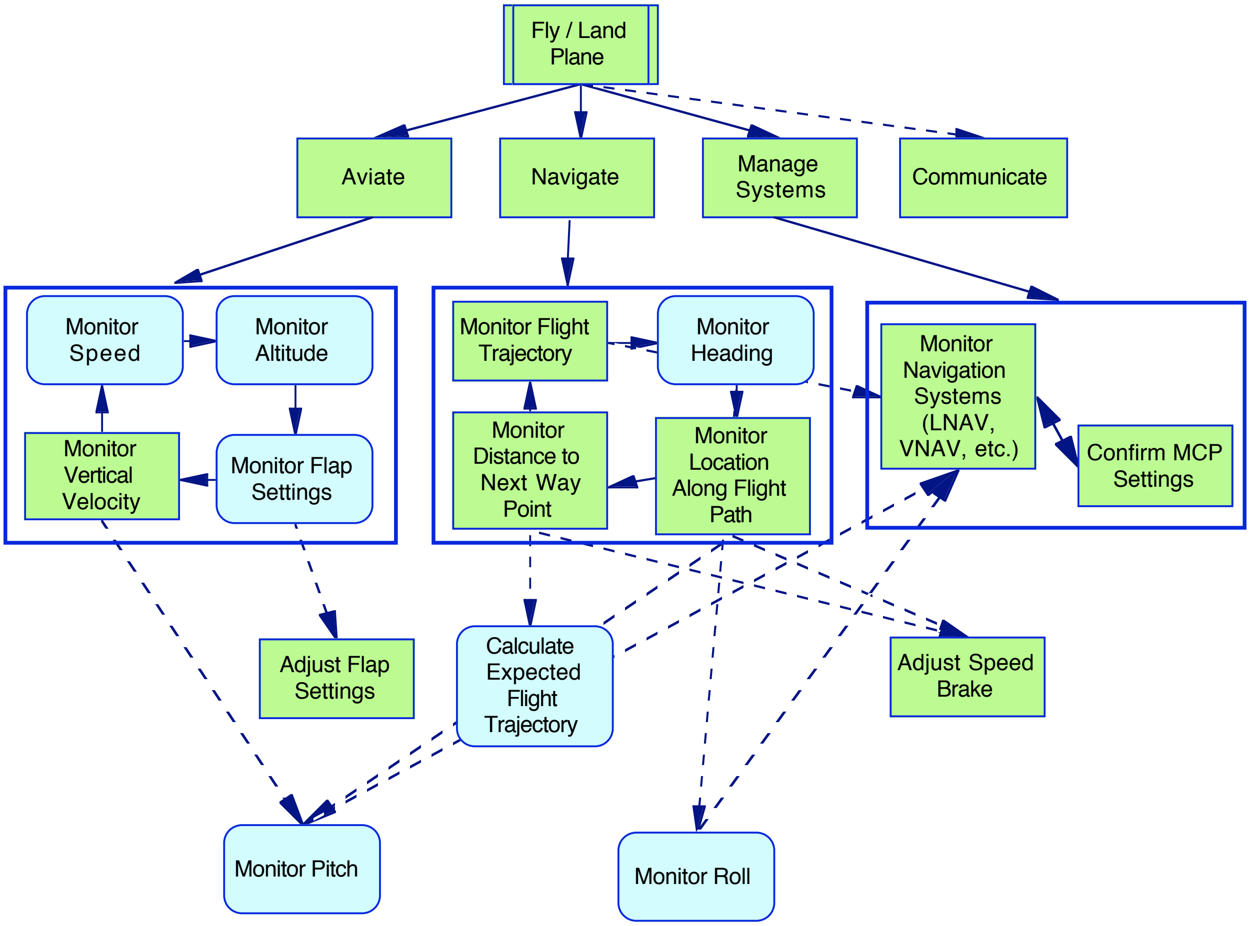

Figure 18. Flow of control resulting from task analyses. Dashed lines represent conditional subgoals that may not be executed every time; blue rounded boxes indicate that some information required by that subgoal may be found on the SVS, if present.

The first insight from the knowledge engineering process is that the bulk of the task, particularly for the first two phases of flight, is primarily a monitoring task in which the pilot is engaged in maintaining his or her representation of the state of the aircraft. Additionally, we learned that pilots very actively check for a number of events and conditions which do not occur in the scenarios, such as late changes of wind direction that might lead to wind shear. Thus, for a lot of the time during the experimental trials, there is the appearance of little workload while in fact the pilots do still have a lot to do. This is fairly realistic for most landings which are, in fact, routine. However, pilots do have to monitor for non-routine conditions. In order to simulate the true workload accurately, we have included checks for many of these things in the model even though they do not occur in the scenarios.

4.1.2 ACT-R

The ACT-R architecture provides a great deal of constraint as well. Working within the parameters of the architecture sets certain boundaries and delimits scope. In particular, it means that we are modeling the task at a highly detailed level of analysis. ACT-R provides end-to-end modeling of the human operator side of the human-in-the-loop, from basic visual and auditory attentional operators to complex cognition and back down to basic motor movements. This impacts the strategies that are even possible and the way in which knowledge about dynamic state has to be updated to be maintained. A thorough review of ACT-R is far beyond the scope of this presentation. However, it should be noted that we are now using the most recent version of ACT-R, termed ACT-R 5.0 (see Anderson, et al., in press for a detailed description). ACT-R 5.0 incorporates the perceptual-motor extensions found in ACT-R/PM and provides for even more aggressive parallel execution of cognitive, perceptual, and motor operations than did the ACT-R/PM version of the system used for the taxi model work.

4.1.3 Extant Accounts of Relevant Phenomena

Because the eye-movement data are the primary focus of the modeling effort, we have examined other data and models in the "allocation of attention" domain in the human factors literature (e.g., Senders, 1964, Wickens, 2002). These are high-level (relative to ACT-R) accounts of how operators choose which objects to visually sample and at what frequency. The basic findings are that the rate at which particular displays are sampled depends jointly on the task importance of the displayed information as well as the rate of change of the information. As one might expect, more important information is sampled more often, and more dynamic information is sampled more often. We believe that these accounts provide a useful high-level starting point; we hope to provide the explanation for how these high-level phenomena emerge from a combination of task and environmental constraints and relatively low-level cognitive-perceptual capabilities. In other words, we recognize and have learned from theories that predict the percent of fixation time on particular information sources through mathematical modeling (e.g., see Section 6). Our goal, in contrast, is to create a process model from which such higher-level descriptions emerge as a function of the lower-level mechanisms in the model-environment system.

4.2 Key Considerations

Here we lay out some of the critical considerations which shaped the modeling effort.

4.2.1 A Dynamic, Closed-Loop Approach

One of the things which distinguishes an analysis at the level of a cognitive architecture such as ACT-R is that it is possible to "close the loop" of the human-machine system. That is, both the human and the evaluated system are modeled dynamically and in detail, and the two sub-models are coupled, yielding a model of the complete dynamic system. Work on the taxiing model revealed that fidelity of the machine/environment model was critical in understanding the performance of the human model; in particular, many of the "higher-level" decisions ultimately depended on "low-level" properties of the human-environment system. For example, the decision of "which strategy should I use to choose which direction to go?" often depended on things like the distance between the sign and the intersection as well as when the cognitive system was free to sample that part of the visual environment. Because ACT-R is fundamentally a non-linear system, small perturbations in the dynamic state of the human-environment system at one time can often lead to large differences in state or behavior further down the road.

Thus, we felt it was critical to continue with this rather complete, closed-loop approach. As previously mentioned, this meant we had to contend with a great deal of detail in modeling the pilots' behavior, but ultimately we believe that path will lead to the best model.

4.2.2 An Adapted Pilot
Present efforts are based on modeling a pilot who is both knowledgeable about the task and well-adapted to it. We are neither modeling novice pilots nor the acquisition/development of piloting expertise. This limits the scope of the model but has other implications as well.

In particular, this means the task analysis information is, in some sense, “contaminated” by the fact that the pilots come into the task with a pre-existing strategy for how to sample the relevant displays. Because they know which information is most important and have a clear model of which information will be most dynamic, their strategies reflect this knowledge. That is, the relevance and rate of change for properties like altitude are known in advance by the pilots, so the pilot does not have to figure out how often to sample that information, he or she already knows how often it needs to be sampled. However, we believe that this has certain implications which we may want to relax later, see the section on later efforts for more details.

4.2.3 An Attention Allocation Focus
As mentioned previously, we believe the primary phenomenon to be explained here is how the pilots deploy their visual attention across the visual array and how this is (or is not) affected by the SVS. While this appears straightforward, there are some subtle issues here which we are exploring. For example, the ACT-R model produces time stamped individual shifts of visual attention (saccades) to small targets; we believe it is a mistake to attempt to map these directly to the individual saccades made by the pilots. Rather, such data can be analyzed at different levels of abstraction. For example, one could reasonably be interested only in more gross performance measures, such as the proportion of fixations on each scene plane, for which we have human data. We can run the model, which produces data at a much finer level of detail, but then extract these higher-level measures from the model run. In fact, this extraction can be performed with more or less the same set of analysis tools that were used on the human data.

An important research question is What level of analysis is appropriate to guide design decisions? Do we want only the more gross measures such as proportion of fixations on each scene plane, or is it worthwhile to attempt to match the exact sequence of fixations generated by a model run with the exact sequence generated by one human trial? We believe that the answer is probably somewhere in between, but this is still an empirical question. Because ACT-R produces behavior at such a fine grain size, we are in the advantageous position that we can at least potentially examine multiple levels.

4.3 Implementation Approach
Many of the details of the implementation have already been discussed. The primary inputs to the cognitive model come from the task analysis; this is the source of the procedural knowledge and the bulk of the initial declarative knowledge given to ACT-R. The output of the model is a time stamped series of behaviors including individual attention shifts, speech output, button presses, and the like. The primary point of comparison for the model output is the human eye-tracking data, which can be examined at various levels of abstraction. One piece that has not been described in much detail thus far is the other half of the simulation: the simulation of the aircraft.

We have mocked up the primary displays (NAV, PFD, MCP, etc.) in the language of ACT-R so that it can directly "view" those pieces of the display. However, this is not enough; ACT-R requires a dynamic environment with which to interact. For instance, if the model changes the flap setting, there are certain expectations about downstream effects on flight performance. To make those happen properly, a simulation of the airplane is required. We have purchased the commercial software package X-Plane for this purpose and are in the process of linking X-Plane to ACT-R (note that X-Plane has been certified by the FAA for training pilots, see http://www.x-plane.com/FTD.html). Figure 19 presents a picture of X-Plane in action.

Figure 19. The X-Plane flight simulation package.

This linkage process is not trivial; we wrote a network interface (based on the UDP protocol) between the two programs from the ground up. X-Plane natively supports sending certain kinds of information such as altitude and heading via the network interface, but other things cannot be sent, including the view out the window. This represents something of a problem since the ACT-R model needs something to "see" out the window (and on the SVS). However, we believed this problem could be solved relatively straightforwardly by abstracting out only what the model would need to look for when it looks. For example, because we know the plane's absolute position and orientation with respect to the airport, we can determine whether whatever piece of information the model was seeking would be available. In the case of our model, this turns out to simply be the four-sided polygons representing the two runways during Phases 3 and 4 only. This task-oriented solution may have uses in other domains as well.

In addition, we had to supply X-Plane with the aircraft specifications (a 757) and the appropriate approach/navigation and FMC programming (e.g., fix points) for Santa Barbara. Fortunately, the 757 specifications and the airport and geography for Santa Barbara were freely available and could simply be plugged in. Figure 20 presents a diagram describing the system. System runs involved initializing both ACT-R and X-Plane appropriately, running them, and collecting a trace of the output. X-Plane is designed to run in real time, so generating multiple simulation runs is time-consuming.

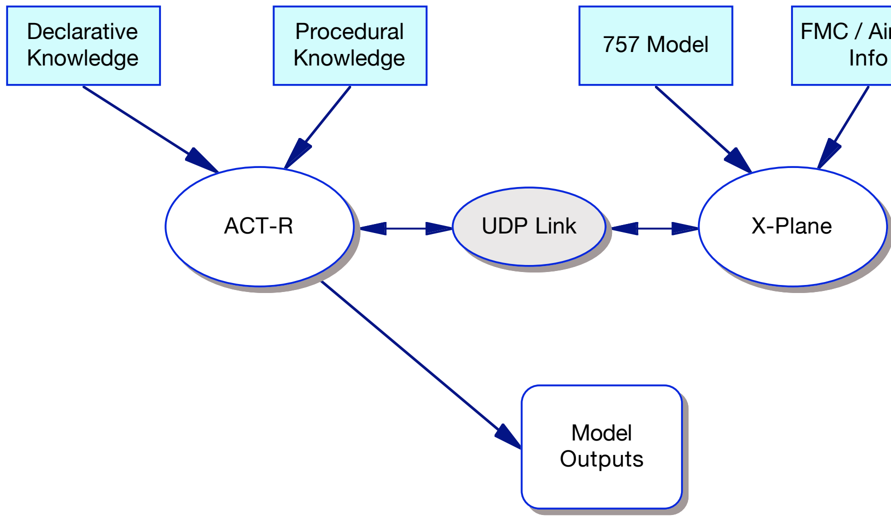

Figure 20. System overview

4.4 Modeling the Land/Go-Around Decision

Special considerations had to be made for modeling one particular aspect of pilot performance, which is the Land/Go-Around (LGA) decision (i.e., should the pilot land the plane or execute a "go around"?) At this stage in the project, we have completed the design and coditng of this aspect of the model, but have not yet implemented it within the ACT-R model itself, but intend to do so shortly. We assumed that five environmental or system variables would provide likely candidate inputs when making this decision; altitude, pitch, offset angle, alignment angle, and approach angle (see Figure 21 for diagrams of the three angles). None of these five variables conforms to any standard statistical distributions, nor are the successive samples of these variable values statistically independent. These factors, along with the dynamic rather than static nature of the decision task (i.e., the *time* at which the decision is made as well as the decision itself are required outputs), to the best of our knowledge at least, ruled out every off-the-shelf decision model in the psychological literature.

In order to evaluate our candidate designs for various models of the LGA decision, we generated a simulated stream of input data values for each of the variables. Successive values of these perceptual streams are governed by basic but informed models of how each of the variables fluctuates during an approach, and account for common patterns along with occasional dramatic changes in value. For example, altitude decreases by a small amount with each sampling, but sometimes the drop in altitude is more pronounced. We introduced variation in the perceptual streams to test our logic against more realistic data, and also to generate cases where an approach can degrade quickly. The five variable streams taken together constitute our simulated approach environment. This first phase of approach simulation is defined in the various *updateVariable* methods in the decision algorithm code contained in Appendix 1.

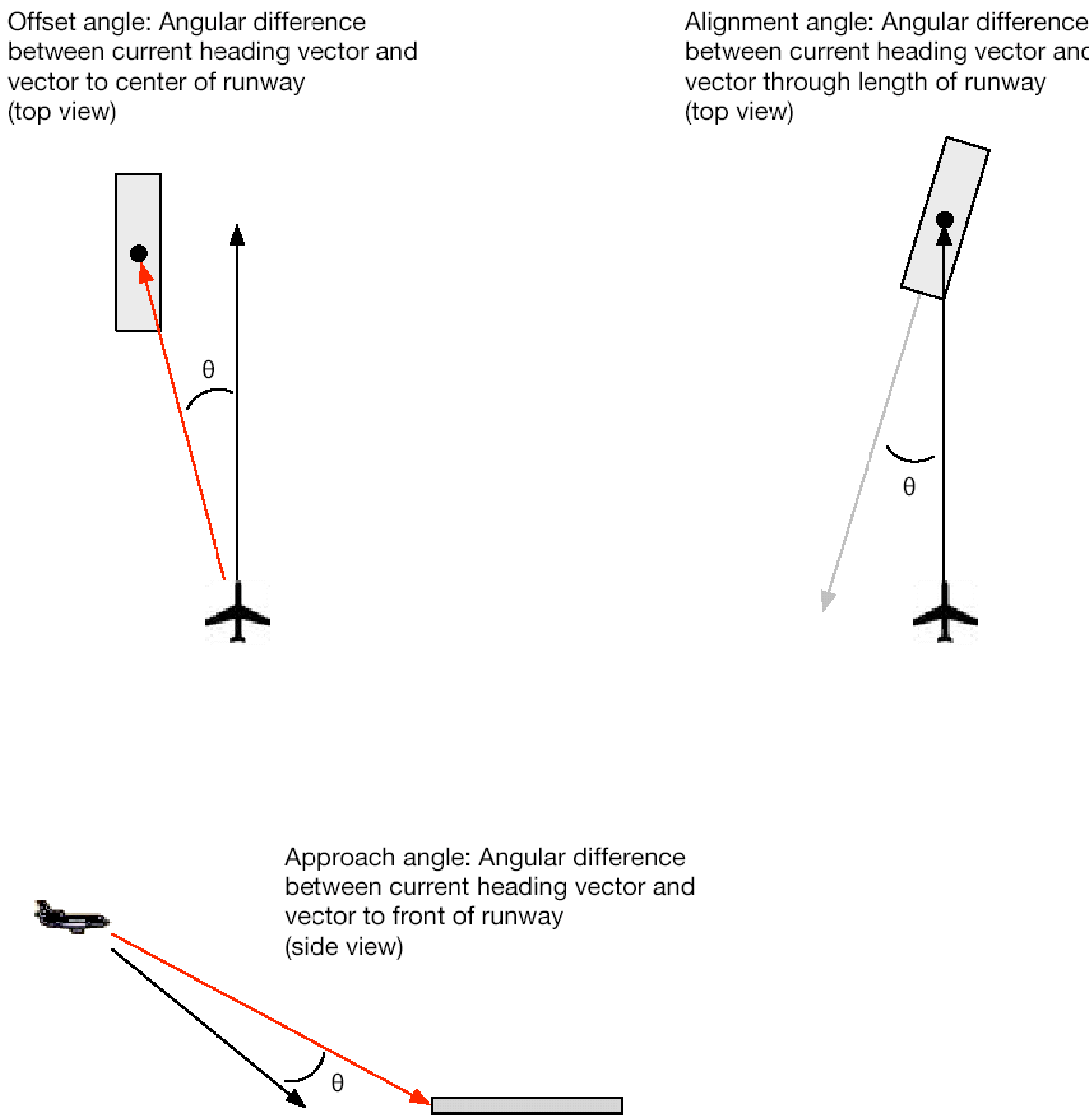

Figure 21. The 3 Angles Considered in the Land-Go Around Decision Logic

The decision to land or go-around depends largely on the quality of the approach. To judge the quality of the approach we calculated an ideal approach pattern for our scenario; i.e., one best (prototype) approach defined by a sequence of values for each of the variables. We then compare any actual stream of input variables constituting a simulated approach to assess its deviation from the prototype, and thus its quality.. *The convertVariable* methods are another set of basic models that convert the deviation in our simulated approach to a measure of fit against the ideal. Variable values that align well with the ideal approach receive better scores, providing a quantitative estimate of the objective quality of the approach.

The other component in the decision to land or go-around is the opinion of the pilot. The pilot in our model makes judgments about the values from the *convertVariable* methods (the actual objective quality of the approach) to determine his or her own opinion. This third phase of opinion formation is outlined in the *opineVariable* methods. Qualitatively, these calculate the opinion of the pilot about whether a particular value of a variable reflects an approach in which they would land the plane. This is accomplished by transforming the objective quality of the approach by using a function that takes the risk aversion of the pilot into account. A pilot who is more risk averse is less likely to rely on the objective quality of the approach, so their opinions (*opineVariable* values) would be consistently lower than the objective quality (*convertVariable* values). This aversion variable can be tuned to simulate a range of opinion formation.

There are three kinds of opinion models built into our decision logic. The first is a simple Boolean opinion. If the pilot holds a strong enough opinion about any particular variable then the opinion about that variable is simply flagged as land, otherwise as go-around. A more refined version of opinion keeps a record of the strength of the opinion directly from the *opineVariable* method. This is referred to as Fuzzy, rather than Boolean, opinion. An additional construct is a set of weights associated with each of the opinions. This applies to all of the opinion models and represents the significance of a particular variable to the pilot. For example, for one pilot altitude may be a stronger deciding factor than pitch.

In addition to the two opinion models that look at the evidence directly, we include a model of decision based on probabilistic inference. As a first step, we define the pilot's predisposition to land. This reflects how determined a pilot would be to land the plane, or how likely they are to be convinced by the evidence that it would be prudent to go-around. When this probability, p(land), approaches 100% the pilot always attempts to land the plane, while a pilot with p(land) of 50% will only land half the time. So p(land) represents the likelihood of landing the plane with a particular pilot before any additional evidence is presented (of course, the vast majority of pilots are predisposed to want to land the airplane).

In addition to this *prior probability*, we have the results from the *opineVariable* set of methods. Again, these measure the opinion of the pilot about whether a particular value of a variable reflects a situation in which they would land the plane. An ideal set of opinions would be 100% for each. Using Bayes' Rule, we combine these pieces of evidence with the pilot's predisposition to determine the probability that they will make the decision to land. The comments in the *bayes method* in the code contain further mathematical details, but put succinctly Bayes' rule shows how to update our probability of landing given the new evidence of the pilot's opinion of the situation.

We use this newly updated probability in two different decision styles, *decideBayesian*. The simpler of the two makes an instantaneous decision based only on the available evidence, always working from the pilot's initial predisposition to land. The second uses a technique known as temporal difference learning to incorporate a memory for how the approach has been evolving. With *every update* of the variables, the pilot's moment to moment opinions shape their subsequent prior probabilities to land. We incorporate a learning rate, the strength of the memory for previous opinions, to determine how quickly the probability of landing will change. For example, a pilot with an initial 80% predisposition to land is presented with a situation in which they only have 60% confidence. The learning rate determines how much the initial 80% predisposition of landing is influenced by the less than ideal evidence. This same method assures that opinion is not sensitive to momentary changes in any one of the variables in our model but is still discounted quickly when a pilot is confronted with a bad approach, or one that is getting worse.

In addition to the *decideBayesian* approach, we have included three additional decision making styles. Each can be applied to any model of pilot opinion, Boolean, fuzzy, or Bayesian. The *decideDisjunction* is the most restrictive of the decision methods. It considers the pilot's opinions and if any one is below a threshold, the decision is made to go-around. The *decideMajority* method looks at all of the pilot's opinions and makes the final decision based on a pure majority decision – tie votes are go-around decisions. The method looks at each of the opinions in turn and determines if they are above or below a threshold to place them into the land or go-around categories. *decideThreshold* takes into account the weights and strengths of all of the opinions together. The decision to land or go-around is made based on whether the overall strength is above a specified threshold.

Decisions can be made at any point in the simulation. If an approach is poor enough, the go-around decision can occur as soon as the pilot is of that opinion. If the simulation reaches the decision threshold – the time where the pilot needs to commit to landing or go-around – the same decision process is executed and will the output at that time will determine whether the pilot will land or go-around.

We would like to emphasize that this is an initial attempt at a decision logic that reflects a coarse approximation of the complex judgments made in this dynamic problem domain. More experimental data is required to make the models more robust, and representative of pilot decision making.

5. Evaluating and Validating the ACT-R Model

5.1 Evaluation and Validation Criteria

Evaluating and/or validating a model this complex with this level of detail is not trivial. In fact, it is not even clear what the best criteria are for doing so. In an ideal world, the model would be constructed based on the current data, evaluated on how well it fit those data, and then validated by seeing how well it predicts the behavior of new subjects, perhaps under slightly different conditions. This presents an interesting conundrum, since one of the reasons for modeling is that empirical evaluation with complex experimental cockpit systems using expert pilots is terrifically expensive in itself. Thus, the possibility of validating our model in the ideal a way is infeasible (at best). The simple question of evaluation by fit-to-data is complex one as well: which data, and how to judge "good" fit?

There are several options we considered but rejected. While the Wickens data was available to us, the situation is not the same (RNAV landing), the aircraft and therefore the flight dynamics and pilot knowledge is not the same (i.e., neither a 757 or 757 pilots), and so what we would be validating would be a model which would be substantially different from the current model. We did not feel this would be the best approach. Another approach which might be taken would be to do some kind of split-half validation. Split-half validation methods are popular with researchers who use regression methods and collect data from large numbers of individuals, such as questionnaire data. However, with only three pilots (who clearly demonstrated individual differences) it was not clear to us that this was a wise idea. It also raised the issue for us of how the data might be split; we felt we could not randomly sample fixations because that destroys the sequential nature of the data.

In the end, what we decided to do was capitalize on the strengths of the supplied data. While there are few subjects, there is a great deal of data which can be considered at multiple levels. We chose to evaluate the model on the basis of fit to data at one level of abstraction and then validate against more difficult, lower-level criteria. That is, we decided to attempt to fit the more global attention allocation data at the level of models like SEEV; this is the question of what percentage of the time did the model look at each display vs. how often the human pilots did. We validated by examining the performance of the same model at a the more fine-grained level of transitions. That is, we ask how well the model, with parameters selected to fit the more global data, fits the more detailed data in the transition matrices.

5.2 Phases 1 and 2

In some sense these are the key phases because anything that changes in these phases as a function of the presence of the SVS is, if not unintended, then not easily anticipated. The primary goal of the SVS was not to change these phases of flight. However, as seen in the previous section, fairly dramatic changes did occur. The primary question, then, is to what degree does the model capture those changes? At the qualitative level, watching the model do the task does have the same general feel as watching video of the actual pilots doing the task (though without the eye-tracker-induced noise). However, we believe the bar should be somewhat higher than that.

There are a few parameters that can be tweaked in this model which affect how well it captures the data. These generally involve strategy selection (“conflict resolution” in ACT-R parlance) parameters that affect the model’s behavior in two key circumstances: first, when choosing which high-level task (e.g., aviate, navigate) to pursue next; and second, when choosing which display to look at for a particular datum (e.g., look for altitude on the SVS or on the PFD?). Many of these parameters can be set on an *a priori* basis without searching for best-fitting values, and thus made to be non-free parameters. The parameters that control the second type of decision are of this type, the values can be estimated by considering the distance to be traveled by the saccade, which factors in how long it takes to successfully complete. Note that these parameters will have a large impact on how often the model re-fixates in the same region, for it is this cost difference which drives the model to prefer to re-fixate in the same sceneplane if the needed datum in available in multiple locations. However, we did not change these parameters in order to achieve a better fit.

Thus, we hand-optimized (purely through trial-and-error) only the parameters that control how often the model selects among the available high-level (e.g., aviate, navigate) goals. This is effectively two parameters that were kept constant through all conditions. Values were selected to produce a good fit to the proportion of fixations in each region for Phase 2 with SVS. Thus, the behavior of the model in all other conditions, and on the transitions, is essentially a parameter-free prediction.

Finally, in assessing fit-to-data here certain adjustments had to be made to the raw data values. Since the model does not produce “off” or “overlap” fixations—those are really artifacts of the eye-tracker—all data for such comparisons was re-normalized with those fixations removed. Also, due to the time-intensive nature of the simulations, model predictions are based on a single run of the model in each condition. We hope to run larger-scale simulations in the near future but have encountered some technical issues with automating the X-Plane application.

Obviously, the model’s best fit is to the total fixations in each region for Phase 2 with SVS, as depicted in Figure 22. The model-to-data r-squared here is 0.978. While the model very slightly under-predicts the PFD and MCP proportions and over-predicts the SVS and DMC proportions, overall the model does a good job of capturing the pilots’ attention allocation performance.

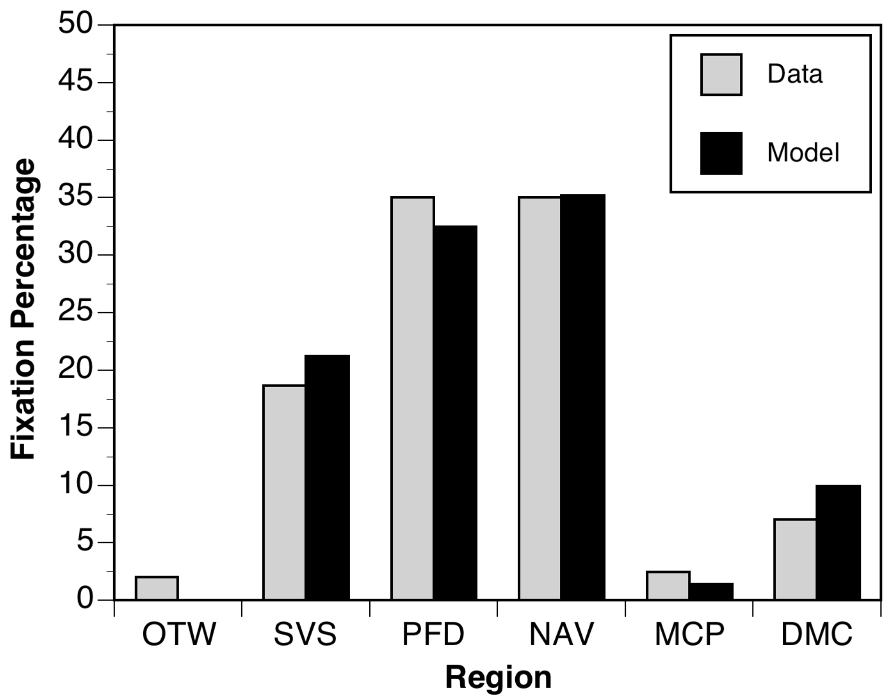

Figure 22. Model vs. data overall attention allocation for Phase 2 with SVS

So, while the model captures the performance with the SVS, how does it fare when the SVS is not present? This is a prediction without parameter tweaking, and is shown in Figure 23. The model is somewhat off in that it ended up allocating slightly too many of the SVS fixations to the PFD and slightly too few to NAV. However, the prediction is by no means poor, with an r-squared of 0.932.

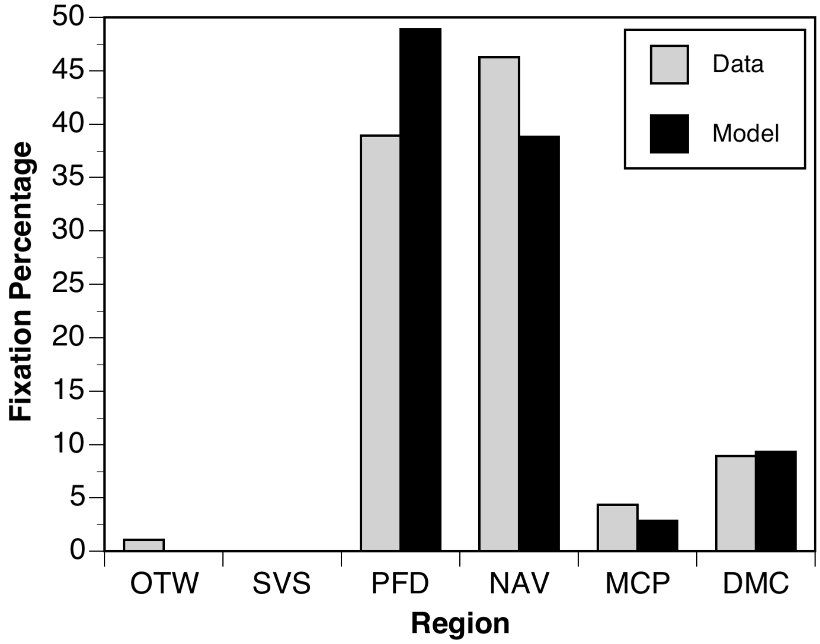

Figure 23. Model vs. data overall attention allocation for Phase 2 without SVS

The next question is how the model does with Phase 1. Phase 1 is probably in general somewhat less important than Phase 2 because it is somewhat less realistic (pilots in the real world do not begin a flight at that point) and also fairly short. However, it is important to see if the model can capture the differences between the two phases. Figure 24 presents the model-data comparison for the SVS condition for Phase 1. This is not a great fit but at least the general trends are captured, explaining almost 70% of the variance (r-squared of 0.678). The model is a little too focused on navigating at this point and does not spend enough fixations assessing the state of the FMC (which is displayed on the MCP).

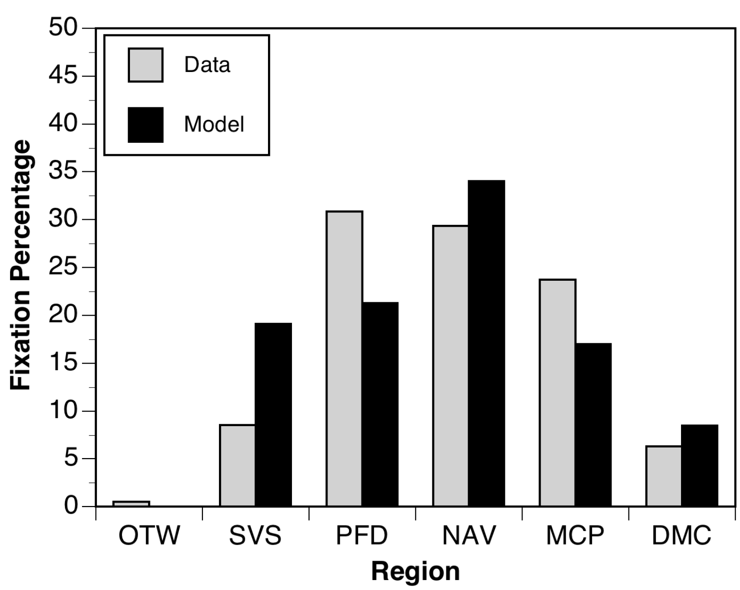

Figure 24. Model vs. data overall attention allocation for Phase 1 with SVS.

However, the situation is somewhat better in the no-SVS condition. The model again does not spend enough of its fixations on the MCP and still over-predicts NAV (and PFD as well), but the fit is somewhat better, r-squared of 0.849. The fit is shown in Figure 25.

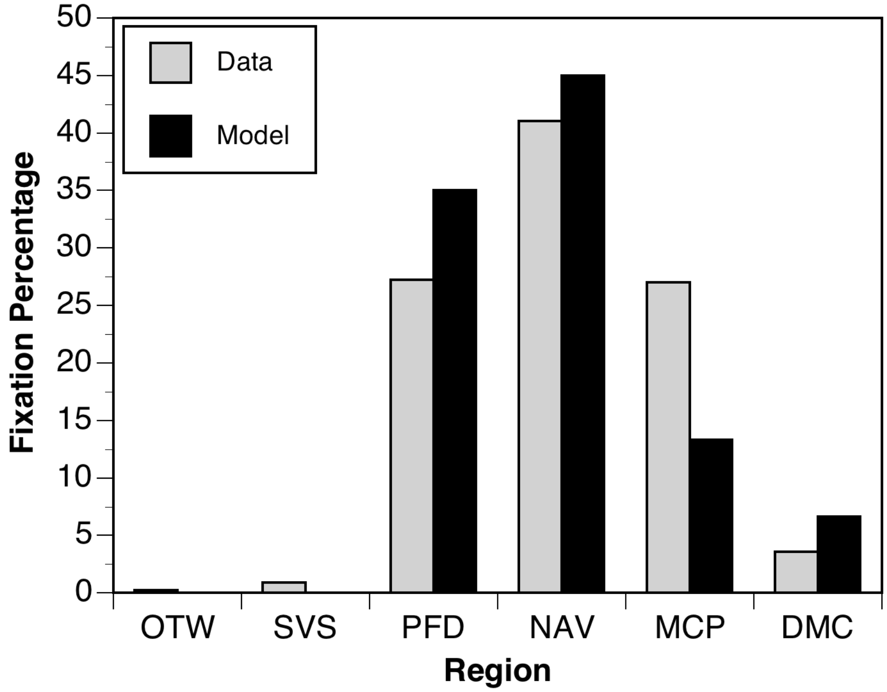

Figure 25. Model vs. data overall attention allocation for Phase 1 without SVS

Taken across these four conditions, the model averages explaining about 85% of the variance in allocating attention to various regions. Given the relatively low N here and the high inter-individual variance, we believe this is about as well as can be done on this data set. However, there is a higher bar here, which is validating the model at a more fine-grained level: the transition matrices. This is a fairly stringent test of the model's ability to match the human pilots at a fairly fine grain of analysis; we believe a process-oriented cognitive model is the only kind of model likely to fare well in such an evaluation, as many other modeling approaches do not even produce such data.

For the sake of brevity, the matrices presented in Section 4 will not be repeated here, only the relevant model-produced matrices.. Note again that the model does not produce "off" and "overlap" fixations and thus produces no transitions to or from those regions. Again, when computing fit metrics, the human data has been re-normalized with those removed.

Table 12 shows the model-produced transition matrix for the no-SVS Phase 1 (the relevant comparison is Table 4). Overall, the match is fairly reasonable. The model also favors the diagonal and matches the NAV and PFD entries quite well. The model explains almost 80% of the variance in the transition matrix, r-squared 0.792.

Table 12. Model-generated transition matrix for no-SVS condition for Phase 1, showing proportion of gaze transitions from (vertical) a display region to (horizontal) a display region

	OTW	SVS	PFD	NAV	MCP	DMC
OTW	–	–	0.02	–	–	–
SVS	–	–	–	–	–	–
PFD	–	–	0.15	0.1	0.05	0.05
NAV	–	–	0.13	0.28	–	0.02
MCP	–	–	0.03	0.02	0.08	–
DMC	–	–	0.02	0.05	–	–

Just as with the overall attention allocation, the model does not do quite as well in the SVS condition with Phase 1, which is presented in Table 13. The model does not even explain half the variance in the data, r-squared 0.419. We suspect this is either an aberrant model run or that one of the pilots did something unusual in one of the scenarios here because this is somewhat inconsistent with the other results we have obtained with the model. Or perhaps the model does require some revision here.

Table 13. Model-generated transition matrix for SVS condition for Phase 1, showing proportion of gaze transitions from (vertical) a display region to (horizontal) a display region

	OTW	SVS	PFD	NAV	MCP	DMC
OTW	–	–	0.02	–	–	–
SVS	–	0.06	0.02	0.04	0.02	0.04
PFD	–	0.06	0.06	–	0.06	0.02
NAV	–	0.06	0.02	0.21	–	0.02
MCP	–	–	0.04	0.04	0.09	–
DMC	–	–	0.04	0.04	–	–

The model fares slightly better with the Phase 2 data. While the model over-allocates attention to the PFD and DMC and under-allocates to NAV, the major trends in the data are captured well by the model, producing an r-squared of 0.772. The relevant transition matrix is presented in Table 14.

Table 14. Model-generated transition matrix for no-SVS condition for Phase 2, showing proportion of gaze transitions from (vertical) a display region to (horizontal) a display region

	OTW	SVS	PFD	NAV	MCP	DMC
OTW	–	–	–	–	–	–
SVS	–	–	–	–	–	–
PFD	–	–	0.31	0.09	0.01	0.11
NAV	–	–	0.13	0.22	–	–
MCP	–	–	0.01	–	–	–
DMC	–	–	0.07	0.04	–	–

For Phase 2 with SVS, the model is approximately as good, as can be seen in Table 15. The model does a fairly good job with the NAV and PFD displays, but is a bit off on the SVS. In particular, the SVS tends to send too many fixations off to other displays and the other displays tend to feed too many fixations to the SVS. However, the model is still in the right part of the space, producing an r-squared of 0.690.

Table 15. Model-generated transition matrix for SVS condition for Phase 2, showing proportion of gaze transitions from (vertical) a display region to (horizontal) a display region

	OTW	SVS	PFD	NAV	MCP	DMC
OTW	–	–	–	–	–	–
SVS	–	0.07	0.05	0.06	–	0.04
PFD	–	0.07	0.12	0.05	0.01	0.06
NAV	–	0.04	0.09	0.22	–	–
MCP	–	–	0.01	–	–	–
DMC	–	0.03	0.05	0.02	–	–

Overall, the model's ability to predict pilots' attention allocation at both the molar level and the detailed level is good, though there is certainly still room for improvement. Most of the major trends are captured, and

5.3 Discussion of Model Results

In general, model results were satisfactory. While the fits are not perfect, the majority of the variance is explained not only in overall attention allocation, but also in terms of the transitions which underlie the more global behaviors. This was done with a bare minimum of numerical parameter-tweaking, meaning these fits have credibility as predictions.

Just as importantly, the model provides some explanation for why the data are as they are. Pilots use the SVS at the rate they do because it contains task-relevant information, and once the pilot chooses to look at the SVS for one of those pieces of information, then the local costs of access—that is, more rapid access to nearby visual items—dictate that if the next needed piece of information is also available on the SVS, it is likely to be sought there.

Thus, the model is sensitive to both top-down factors, which in this case are the information needs of the pilot as determined by the task analysis and the memory system of the pilot, and bottom-up factors, including the layout and redundancy of the available displays and low-level parameters of the human visual system, such as saccade latency and accuracy. This is an important insight in and of itself; one cannot predict performance by looking just at the display or considering just the task structure.

There is still some important integration work to be done to fully complete modeling of Phase 3. The Land/Go-Around logic needs to be integrated with the ACT-R model so we can predict not just when pilots will choose when to attempt a landing or not, but also their patterns of information access in this stage of the landing process. We expect this integration to be relatively straightforward, given we already have prototype code for this logic written (see Apppendix 1)

6. Process and Product Modeling of Attention: Toward an Integration

In addition to the previously described ACT-R modeling of the Santa Barbara SVS scenarios, we have also completed a related empirical investigation that we hope will contribute toward the validation and verification, or at least continued refinement, of our process modeling of visual attention allocation in contexts such as aviation. In the research described below, the validation goal is not one of achieving correspondence with the NASA SVS data set *per se*, but rather one of achieving correspondence with what is known, in a larger sense, about visual attention allocation in dynamic, technological work environments. We begin with a review of the relevant literature.

6.1 What Factors Mediate Visual Attention Allocation?

Several factors are known to influence scanning behavior by guiding visual attention allocation to various areas of interest in the aviation environment. Models of monitoring have been incorporated, in various ways, into several models of monitoring behavior. Monitoring, or visual sampling, is a form of selective attention where the operator scans a display, without actively seeking to change the system state (Moray, 1986), in order to maximize value or minimize cost (Wickens & Hollands, 2000). These models are often found in the supervisory control literature. Monitoring behavior and supervisory control, although different, are often highly related because appropriate control requires a veridical understanding of the supervised system (Moray, 1986). Monitoring is a necessary requirement for supervisory control, but it is only one of the components involved. Since visual scanning while piloting is a not a supervisory control task, but simply a monitoring task, this review will not cover supervisory control. Instead, this section will discuss the models that have been developed to predict attention allocation strategies in monitoring behavior.

6.1 Initial Research

Some of the earliest studies to address monitoring behavior were conduced by Fitts and colleagues in the late 1940s and early 1950s (see Fitts et al., 1950 or Moray, 1986, for a summary). These studies involved recording the eye movements of pilots during actual flights in order to develop improved cockpit instrument layouts. The Fitts studies were also instrumental in establishing a correlation between fixation frequency and importance, the length of fixations and difficulty in encoding a display, and the limitations of eye movement fixation durations (Senders, 1983).

6.1.2 Senders' Model (Bandwidth as a Primary Mediator)

Senders (1964) provided one of the first models of operator visual sampling behavior (but see Craik, 1947). Senders approach was based on an information theoretic treatment of continuous functions that represent the signals that human-machine system operators often must monitor. The model created by Senders and his colleagues assumes that operators sample to reduce uncertainty about the state of the system, which provides continuous information to the observer. Operator uncertainty stems from the rate of change of information, or bandwidth, presented in the display. Senders (1964) applied the Nyquist sampling theorem, which implies that, to effectively monitor a display, it is necessary to sample a signal at two times its bandwidth (a bandwidth of W Hz would need to be sampled at a rate of 2W Hz).

Senders (1964) tested the model by asking operators (N=4) to monitor a display with four instruments, one mounted at each corner of a square viewing area, each instrument separated by 60 degrees of visual angle. Each of the instruments had a different bandwidth, ranging from 0.5 to 4.0 radians per second. These bandwidth values were chosen so that operators would be required to continuously monitor the displays, but would not be overloaded by the required observations. Operators were asked to monitor the four displays and indicate when any one of the instruments went out of range (a pointer moving into an "alarm" area).

As shown in Figure 26, Senders found that much of the variance in operators' visual scanning is explained by the bandwidth of the instrument. However, operators have a tendency to undersample high bandwidths and oversample low bandwidths. This is a consistent finding across many fields (see the discussion of "sluggish beta" in Wickens & Hollands, 2000). When reanalyzing the data, Senders (1983) was able to account for the undersampling by suggesting that, at high bandwidths, observers are able to detect bandwidth instantaneously. This means that it would only be necessary to sample at a rate of W instead of 2W Hz (Moray, 1986). Moray (1981) suggested that the oversampling was caused by forgetting. Forgetting may be an even more important issue when dealing with pilot distraction because conversation may stress already limited working memory capacity. If that is the case, then the optimal model of pilot visual scanning should perhaps be designed to accommodate forgetting.

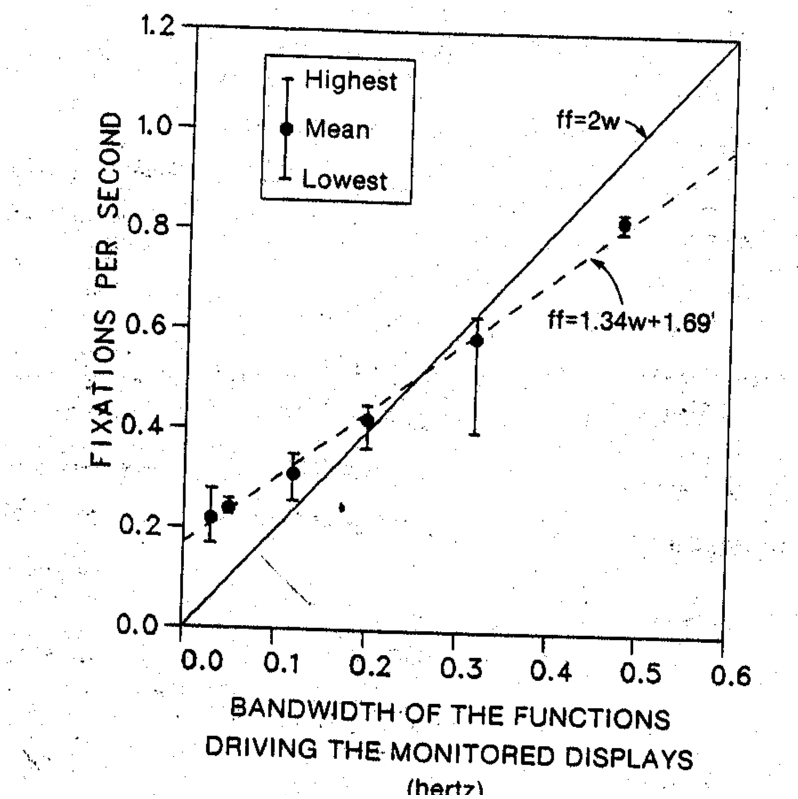

Figure 26: Results from Senders, 1964 (From Moray, 1986)

Senders successfully applied his model of sampling behavior in a more basic monitoring task to the more applied driving domain (Senders et al., 1967). In this study it was proposed that in driving, scanning behavior is influenced by the curvature of the road, the density of obstacles, the ability of the vehicle to maintain a straight line without correction, and the driver's estimation of the probability that other traffic will impede into the path of his or her vehicle. The authors briefly mention that due to the dynamic nature of the driving task (e.g. changing traffic, density of road signs, etc.) bandwidth is variable, leading to the need for drivers to adapt their sampling strategies to accommodate changing bandwidth rates. However, the study did not suggest how changing bandwidth rates will affect sampling strategies, or how much (if any) attention is required to update those sampling strategies.

Moray (1990) suggested that signal bandwidth is the primary factor guiding selective attention in transportation. He suggested that, in driving, the environment is highly structured and that a glance the state of one object or event can inform one about the state of many additional the objects and events in the neighborhood. These causal dependencies substitutes for observation. Additionally, Moray (1990) suggested that signal bandwidth is relatively low while driving; i.e., that although a vehicle is moving at a much higher speed than during manual locomotion (e.g., walking), this speed is still slow relative to the limitations of the perceptual system.

Although Senders et al. (1964, 1967, 1983) provided a good starting point for understanding what drives attention in visual monitoring, there are some noted limitations of the model. The first criticism that has been levied against the Senders Model is that it is not representative of "real-life" tasks (Moray, 1986). In the Senders et al. experiments, the instruments were not given any meaning, therefore the operators (most likely) did not prescribe any value to any of the displays (but see Senders, 1968, for an application). It is not often that all displays are of equal importance (e.g. a fuel gage has a high cost of a missed signal, a speedometer often does not). Also, it is not realistic to assume that the operator will never be overloaded. Senders does not account for how operators will react to potential overload (Carbonell, 1966).

Another limitation of the original Senders Model is that it does not account for situations where the instrument is near the "out-of-bounds" area when viewed (Moray, 1986). Intuition would suggest that if an operator has noticed a control is close to the limit, their next glance at that control should be made sooner than the model would suggest. An example of this in driving would be noticing that the gas gauge is close to empty would create a situation where it would be looked at more often. A complementary model was later proposed to account for this situation (Senders, 1983), but no studies were found that validated the model.

6.1.3 Carbonell's (Costs) and Sheridan's Model (Rewards): Value as a Mediator

Several studies have proposed modifications to Senders early model to make it more representative of operational tasks by adding a value-based component to the model (Carbonell, 1966; Kvalseth et al., 1977; Sheridan, 1970). Carbonell's Model (1966) adds a value component by adding a cost to missing a signal. Similarly, Kvalseth et al.'s Model (1977) explored adding a cost for errors (they also added a cost-based, sampling effort component). Sheridan (1970) added a positive reward-based value component for reducing error.

Carbonell (1966) suggested that different instruments involve various consequences of missing signals (e.g. missing fuel error more costly than speed). His model used a queuing theoretic approach, in which the central idea was that each instrument is competing for attention and the goal of the operator to minimize the cost of missed signals. Cost was calculated using Equation 1 and 2 (below), based probabilities and costs of errors.

$$C(t) = \sum_{i=1}^{M} \frac{C_i P_i(t)}{1 - P_i(t)}$$

Equation 1: Carbonell's model of probabilities and costs

The total cost of looking at an instrument at time t will be:

$$C_i^{'}(t) = C(t) - C_i P_i(t)$$

Equation 2: Carbonell's total cost equation

Where:

M = number of instruments
t = observation time
C(t) = cost associated with exceeding threshold (time independent)
P(t) = probability that the instrument will exceed threshold at time t

Kvalseth (1977) developed a similar model, but added a component based on the cost of sampling (an effort component). Kvalseth examined monitoring behavior in extremely low-bandwidth processes (printed numerical displays). The novel contribution from this model was the idea of combining sampling cost and cost of error into one model. However, a criticism of the experimental design Kvalseth used to validate this model is that it did not allow enough practice time to participants to obtain optimal sampling behavior (Moray, 1986).

Sheridan (1970) suggested that, instead of their being a cost associated with missing an event, there should be a reward for reducing error. The goal of the supervisor under this conception is to maximize return based on the value function and the bandwidth of the process. The optimal sampling interval, based on the net payoff, is calculated by taking the value of the last observation (a positive value) minus the cost per unit time of not sampling (a negative value).

6.1.4 Other Mediators and Models

Several other factors have been found to guide attention allocation in visual scanning, including effort, salience, habit, and contextual relevance (Wickens, 2004). For example, people often try to conserve effort, which suggests that highly effortful situations will negatively affect optimal scanning behavior (Wickens, 2004). Senders (1983) suggests that while frequency of fixations is due to the importance of the information, dwell durations do in fact reflect, in part, the difficulty of checking and interpreting the information. Kvalseth's (1977) model discussed in the previous section included an effort component as an attentional mediator. One way effort may influence scanning in piloting would be that AOI that are not within the 8 degrees of visual angle, where 90% of scanning takes place, might not be looked at as often as an optimal model might predict. During a dual task situation, the effort component might come into play even more because of already limited resources.

Another factor that affects scanning is salience, or the attention-capturing ability of an event (Wickens et al. 2001). Salient events are characterized by an increased ability to automatically capture attention. This said, salience may not always have predictable effects on visual attention allocation in human-machine systems. For example, Wickens et al. (2001) have suggested that operators do not necessarily look more frequently at places where salient events are likely to occur, and suggested that high salience may even reduce eye movements to an AOI.

Habit and context are two other factors that may influence scanning (Wickens, 2004). Influences on scanning behaviors due to habit include learned behavior such as scanning a display based on a learned, "hub and spoke" scanning pattern (Wickens, 2004). In driving, habit might influence scanning to all road signs, even if the driver expects little relevant information will be contained in the signs. Contextual influences are based on the expected consequences of events that have just happened (Wickens, 2004). For example, after seeing the driver in the vehicle ahead talking on her cell phone, one might expect that vehicle to behave more erratically, and thus spend more time visually attending to that vehicle.

6.2 The SEEV Model

Recently, Wickens and colleagues (see Wickens et al., 2002; Wickens et al., 2003) combined several of the components involved in monitoring strategies and in attention allocation into the

"SEEV" model of visual attention allocation. The SEEV model was designed to integrate the multiple mediators of attention allocation (discussed at length above) into a single model. This model suggests that attention is directed to an instrument or display based on four factors: Salience (S), Effort (EF), Expectancy (EX) (or bandwidth), and Value (V). Equation 3 reflects the functional form of the model, where P(A) is the probability of attending to a particular AOI, and s, e, ex, and v, are (estimated) parameters indicating the relative weighting or importance of the factors presumed to mediate attention allocation.

$$P(A) = sS - efEF + (exEX * vV)$$

Equation 3: The SEEV Model

Wickens et al. (2001) suggests that expectancy and value (expected value) are both top-down, knowledge driven components, guided by an internal mental model of the environment. These two factors are considered to be the "optimal" factors in guiding attention allocation; that is, these two components *should be* the only factors that drive scanning in goal-based task performance. With well-trained operators, these two factors should explain the majority of the variance in observed scanning behavior (Moray, 1986).

They are not, however, the only components that influence attention allocation. Salience and effort are identified as two bottom-up factors that inhibit the optimal scanning pattern. Effort is described as an inhibitor to eye movements, while salience draws extra eye movements to a location. Wickens (2001) suggests that in situations where operators are highly trained (as in Senders, 1964; Carbonell, 1966), these nuisance factors have a minimal effect on observed scanning patterns.

The SEEV model was designed to integrate previous models of information seeking (e.g. Senders, 1964; Carbonell et al. 1966; Sheridan, 1970), which combined bandwidth and value in a 1:1 mapping, with models of task management that incorporated a hierarchical goal structure into assigning values to the components. Weights are assigned to the components in the model by first identifying the areas and tasks of interest and putting them into a matrix. Each cell in the matrix is assigned a priority (e.g., aviate is higher priority than navigate which is a higher priority than communicate) based on the importance of directing attention to the location and event relative to the other cells in the matrix. Numerical values are assigned to the cells using the lowest integer assignment keeping ordinal values intact (e.g., 1, 2, 3, 4, etc.)

In an effort to validate the SEEV model in an operationally representative environment, Wickens et al. (2001, 2002, 2003) conducted a series of studies examining the expectancy (bandwidth) and value components of the model (effort and salience were not included) using well-trained pilots. Since the pilots were well trained, the effort and salience components were not expected to have a large influence on scanning strategies. Experiments 1 and 2 compared scanning behavior in free-flight vs. baseline conditions (ATC input) as well as conflict vs. non-conflict trials. It was found that the optimal expected-value model accounted for approximately 80% of the variance in scanning. Experiment 4 added a communications task (task-relevant communications) to the baseline condition from Experiments 1 and 2. The auditory input was modeled as a bandwidth parameter. The predicted model accounted for approximately 95% of the variance in actual scanning behavior.

Experiment 3 examined the effects of traffic density on scanning strategies. It was found that 95% of variance in scanning strategies was accounted by the EV model, suggesting that experienced pilots are able to adapt their scanning behavior well to changes in traffic levels. These findings are consistent with Sheridan (1976), which suggests that operators update their statistical model with each glance. That is, each time an eye movement is made to a control or an area of interest, the new expectancy is compared with the old model. The model is updated using the difference between the two models compared with the reliability that the new observation is correct.

However, it may be of interest to make an additional distinction between tasks that are visually demanding and those that are cognitively demanding when developing the predicted model. This distinction may inform solutions to the pilot distraction problem because limitations are most likely due to demands for divided attention. According to Senders (1983), areas that contain high bandwidth will be fixated on frequently, but areas that are difficult to obtain information from will be fixated on for long durations. This concept might be an additional consideration when applying priority and value components to the hierarchical model of control.

6.2.1 Summary

Moray (1990) has claimed that visual attention is primarily driven by unconscious automatic habits that are tightly coupled to the structure of the environment. He suggests that attention theory is one of the major places to look to reduce error in transportation systems and that predicting the causes of human error will come from using eye movements and other measures of operator behavior.

The previous sections of this review of the literature identified several factors that mediate visual scanning behavior, including bandwidth (or expectancy), value, effort, salience, habit, and context. The most influential factors, expectancy and value, have been modeled and validated in an aviation environment using the SEEV model (Wickens, 2003). It was suggested in the previous section that this model could be applied as a predictor of "optimal" scanning patterns in the piloting domain. Although it is recognized that SEEV is not an optimal model of scanning, it appears to be a reasonable predictor of visual attention allocation when the pilot is not distracted from monitoring by non-monitoring tasks.

6.3 The SEEV Model as a Rational Analysis of Visual Attention Allocation

The SEEV model is not a *process model* in the sense of ACT-R, Midas, or other computational models intended to mimic the *time course* of (attention allocation) behavior. However, we believe that the assumptions behind the SEEV model provide an important and valuable source of empirical constraint on the design of *any* process model. While the current NASA data set naturally provides an excellent source of empirical constraint relevant to the particular application of interest (SVS in aviation), the SEEV model captures a more general set of constraints about the adaptive nature of attention allocation in less task-specific terms.

In fact, we suggest that the SEEV model plays the role of providing a "Rational Analysis" (Anderson, 1990) of attention allocation behavior in that it provides 1) a description of features of the task that are relevant to shaping behavior (Salience, Expectancy, Effort, and Value); as well as 2) a candidate explanation of the functional (input-output) relations between these features and behavior. In this sense, like existing Rational Analyses of phenomena such as memory retrieval, categorization, and reasoning (Oaksford & Chater, 1998), the SEEV model. considered as a rational analysis, provides an empirically-motivated functional specification for the entire class of process models intending to mimic attention visual attention allocation behaviors.

As such. in parallel with the ACT-R modeling activities discussed earlier in this report, we undertook an effort to better understand the relationship between the SEEV model and our computational process modeling, in the hopes that, eventually, some of the more ad-hoc decisions we were required to make (regarding visual attention allocation processes) during ACT-R modeling could eventually be made in a fashion that was maximally informed by the best available current theory of visual attention allocation in human-machine systems (SEEV). Unfortunately, as of the date of this report, we have not been able to fully close the loop in this fashion, but (as discussed in Section 7), we intend to continue working in this direction as we continue to refine the ACT-R model. However, we do believe we have already taken at least a few important steps in bridging the process and product modeling approaches to modeling attention allocation, as will be discussed in greater detail in the following sections.

Specifically, during the process of ACT-R modeling and the analysis of eye movement data from the NASA SVS scenarios, we identified an issue had to be addressed head on in order to better understand the relationship between the SEEV model and our own computational modeling. This issue concerns how operators, such as pilots, allocate visual attention to information sources that are not perfectly reliable indicators of what environmental or system variables they were designed to represent. (Indeed, this was the exact motivation behind the design of some of the SVS scenario conditions). We discuss this issue, along with the experimental study to which it ultimately led, in the following section.

6.4 Ecological Validity as a Mediator of Attention Allocation

While performing our literature review and examining the SEEV model to provide guidance for the design of the attentional components of our ACT-R modeling, we came upon an issue that we believe has been insufficiently addressed by previous research. Specifically, both Senders' original model, and indeed all models that include bandwidth as a predictor, assume that bandwidth, in and of itself, is an important driver of attention. This assumption makes sense under the assumption that the operator's goal is to maintain an accurate internal model of the controlled system. As described above, some have gone so far as to contend that "optimal" attention allocation should be driven solely by bandwidth (and perhaps also value) factors. In his original research on this issue, Senders' claimed that his experimental participants (nearly) optimally adapted to bandwidth, with the implicit presumption that it was normative for a good operators in human-machine systems to do so.

However, we note that in many, if not most, cases in actual human machine systems, signal bandwidth is not an end in itself (as an object of adaptation), but instead merely a readily available, often salient (or "proximal") cue that is informative about some other task-relevant event (e.g., an indicator moving out of a safe range) which is the actual (or "distal") object of adaptation. This is certainly true in the current SVS context, as some experimental conditions were specifically designed to create a less than perfectly veridical relationship between the displayed information and the actual task environment. Senders' research did not, and in fact could not (as designed), address this issue, as his experiments perfectly confounded signal bandwidth (as measured in radians/sec) and alarm frequency (the rate with which each dial measurement went out of the safe range). As a result, it is unclear whether it is more correct to say Senders' participants had adapted to bandwidth, had adapted to alarm frequency, or had adapted to the *ecological validity* relations linking bandwidth, as a proximal (more readily available) cue, to alarm frequency, a distal (less readily available) criterion. As such, we decided to attempt to reduce this ambiguity by designing and performing an experiment resolving the confound in Senders' original study, but otherwise using a task modeled after his original task. This experimental study and findings are presented in the following sections.

6.5 An Empirical Investigation of Ecological Validity as a Mediator of Attention Allocation

We performed an experiment modeled as closely as possible after Senders' original scanning task except for the fact that we orthogonally decoupled signal bandwidth (rad/sec) and alarm frequency (the number of times each gauge went into the alarm region per unit time). The latter value (alarm frequency) also mimicked Senders' study and was a constant across all conditions of the experiment and was set to the value of 0.25 (i.e., one of the four gauges on the display exceeded its alarm threshold, on average, every 4 seconds). The four bandwidth values we used (across the four gauges on the display) were identical to those used in Senders' original study (0.5, 1.0, 2.0, and 4.0 radians per second).

In order to also make the study relevant to models that take the *value* of an information source in addition to bandwidth into account, we varied the value (points earned for a correct detection, points deducted for a miss or false alarm). These values were 1, 2, 4, and 8 points, again, varied across the 4 gauges (see Figure 27 for a depiction of the experimental display, including horizontal bars representing the different values of the different gauges). Finally, we introduced three ecological validity (EV) conditions in the experiment, thus decoupling bandwidth and alarm frequency, but reflecting the correlation between the two. The three values of EV were 1.0, 0.75, and 0.25. Ecological validity did not vary by gauge, but rather, is a measure of the overall correlation between the bandwidths of the four gauges and the frequency with which the four gauges transitioned into the alarm region over a period of time, such as a 5-minute session.

As such, one can think of EV as the predictiveness or reliability of the more readily available (proximal) cue of bandwidth, and the actual, but less readily available (distal) task criterion of alarm frequency, which, combined with value, was the true object of successful adaptation in this experiment. One purpose of the experiment was to determine how EV (the reliability of a proximally displayed cue) mediate how bandwidth and value influence attention allocation, as a factor to be potentially introduced into future models. A second purpose of the experiment was to obtain a perhaps clearer picture of how value and bandwidth combine to influence attention allocation, regardless of EV (to date, the SEEV model has been validated

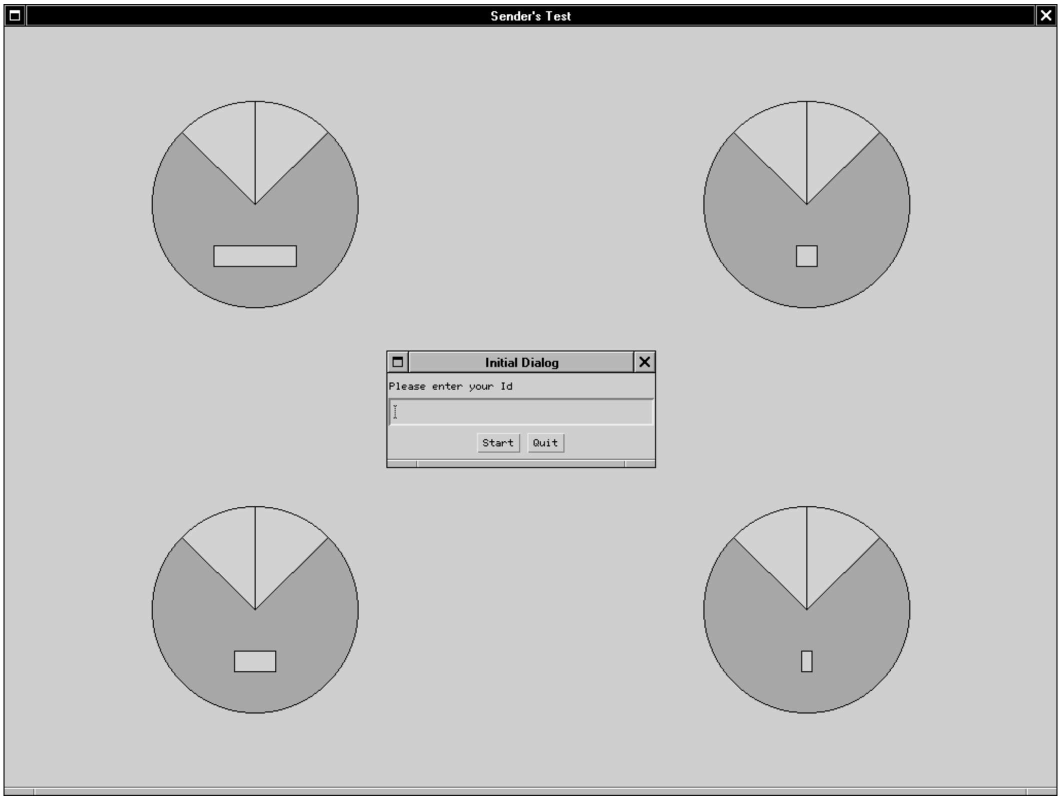

Figure 27. Experimental Display at Session Initiation

only in complex task environments in which value and bandwidth could be measured only on an ordinal scale; the present experiment provides ratio scale measures of bandwidth, value, task performance, and even eye fixation duration percentages per gauge – see below for details).

6.5.1 Method

Fifteen students with normal or corrected-to-normal vision, recruited from the University of Illinois participated in the experiment for one hour each day for five consecutive days (Monday-Friday). The original experimental design incorporates 30 participants, the second group of 15 to be run after the date of this report, under conditions identical to those described below. Each participant was remunerated $40 for his or her participation in the study and was informed that he or she would receive a $50 bonus for being the high scorer in his or her experimental group (EV condition). The fifteen participants were randomly assigned to one of the three EV conditions described above (1.0, 0.75, 0.25), resulting in 5 participants per EV condition. At this low value of *N* (per condition), we believe that any statistically significant results based on the data obtained so far are likely to represent practically relevant (large) differences, and we expect that some of the marginal effects observed here may reach statistical significance after running the final set of 15 additional participants.

In Table 16 below, The G values across the top of the table refer to the four gauges on the experimental display. G0 is the upper left gauge, with G1, G2, and G3 being the upper right, lower left, and lower right gauges, respectively. Subjects 1-5 comprised the EV= 1 condition, 6-

10 the EV = .75 condition, and 11-15 the EV = .25 condition. The entries in the table indicate how the Value and bandwidth (BW) independent variables were mapped onto each of the four gauges across the entire set of 15 participants, indicating (by rows) how participants were yoked across EV conditions.

Participants performed eight, 5-minute trials of the task per day, requiring them to press one of four response keys in response to a gauge moving out of range (into its alarm region). The four keys were spatially mapped to be compatible with the gauges to which they were associated. They received points equal to the values in the table below for "hits," i.e., pressing the correct response key within 1 second of a gauge going out of range, with a similar number of points being deducted for misses and false alarms associated with each gauge.

Table 16. Experimental Design (see text for details*)*

	G0		G1		G2		G3	
Subject Number	Value	BW	Value	BW	Value	BW	Value	BW
1,6,11	8	0.5	2	1	1	2	4	4
2,7,12	4	1	8	2	2	4	1	0.5
3,8,13	2	2	4	0.5	8	1	1	4
4,9,14	8	4	1	1	2	0.5	4	2
5,10,15	1	0.5	8	4	4	2	2	1

The scoring algorithm ensured that that participants could score only one hit per alarm. Data on total score, hits, misses, false alarms, and eye movements (subsequently coded as the percentage of dwell time per gauge) were recorded. Pilot testing revealed that it could be possible that participants might be able to detect some periodicity in the "random" (sum of 3 sine waves of various frequencies, amplitudes and phase shifts) signals presented at a given bandwidth and ecological validity combination on a given gauge. As such, we created four such signal streams for each bandwidth/EV combination, and randomized their presentation over the course of the experiment. We have no evidence (in the data, nor in verbal reports) that any participants noticed periodicity.

Participants sat in a position such that the mean distance between their eyes and the display was 35 cm. This distance facilitated coding eye movement data from a set of Apple ISight cameras. The horizontal separation between left and right gauges on the display was 17.5 cm, or 28.1 degrees of visual angle. The vertical separation between upper and lower gauges was 13 cm, or 21 degrees of visual angle. The diameter of each gauge was 5.25 cm. The safe range for the gauge pointer was plus and minus 45 degrees from vertical, resulting in an overall 90 degree safe zone, vertically centered. The alarm region was any pointer position outside this zone.

Finally, the experiment was designed to include a transfer condition consisting of the final four 5-minute experimental trials on the last day of the experiment. Participants were not informed of the transfer manipulation. In the transfer condition, EV=1 participants performed the task at an EV of .75, the EV=.75 participants performed the task at an EV=.25, and the EV=.25 group performed the task at an EV = 1. To summarize, each member of the three EV groups performed

36, 5-minute trials at a specified level of EV over the first 4 and one-half days of the experiment, and completed the study by performing 4, 5-minute transfer trials at a different level of EV on the final day. In the presentation of results below, where the nature of the trials is not mentioned, we refer to the analysis of data from the first 36 “normal” trials only. We will specifically mention when we are discussing findings based on data from the transfer condition.

6.5.2 The Displayed Signals and Their Properties

Although to this point we have written as if there were simply three EV conditions (1.0, 0.75, and 0.25) and a constant alarm rate of .25 per second across the entire experiment, this was not strictly true. To construct the signal streams used to drive each of the four gauges in each 5-minute trial, we had to search an enormous space to identify a set of four signal streams, each being the sum of three sine waves of varying frequency, amplitude, and phase shift in order to constrain alarm rate to be approximately .25/sec, while satisfying the bandwidth (.5, 1, 2, and 4 rad/sec) and EV (1.0, .75. and .25) values required by the experimental design. We also had to ensure that the displayed set of four signals every 5 minute trial did not result in any gauge having zero alarms, thus precluding participants from offloading the task of monitoring any particular gauge. Somewhere on the order of 100 hours of 1-Gigahertz *CPU* time were necessary to identify the entire set of signal streams used for the experiments. Even with this amount of computation (accomplished by a genetic algorithm we developed specifically for this purpose), we found that we had to tolerate some deviation from an exact alarm rate of 0.25/sec per trial in order to obtain acceptable solutions. In addition, we had to similarly tolerate some deviation from the exact ecological validity levels desired.

Table 17 therefore presents information on the exact levels of bandwidth and alarm frequency that participants in each EV condition observed. Recall that four different signal streams per EV condition were generated and presented in random order to preclude participants from detecting periodicity. This explains why four signal sets are provided per EV condition in Table 17. In the following, “Signal Set” refers to the set of *all four* signal streams used to drive the four gauges on the experimental display in a single, 5-minute, experimental trial.

Table 17. Actual EV and Alarm Rate Values Used Per Experimental Condition

EV = 1.0 Condition

	Actual EV	Actual Alarm Rate
Signal Set 1	0.99	0.28
Signal Set 2	0.96	0.25
Signal Set 3	0.98	0.29
Signal Set 4	1.00	0.31
Average Signal	**0.98**	**0.28**

Table 17 (Cont'd). *Actual EV and Alarm Rate Values Used Per Experimental Condition*

EV = .75 Condition

	Actual EV	Actual Alarm Rate
Signal Set 5	0.76	0.22
Signal Set 6	0.77	0.28
Signal Set 7	0.70	0.23
Signal Set 8	0.79	0.25
Average Signal	**0.75**	**0.25**

EV = .25 Condition

	Actual EV	Actual Alarm Rate
Signal Set 9	0.22	0.25
Signal Set 10	0.30	0.25
Signal Set 11	0.25	0.25
Signal Set 12	0.30	0.24
Average Signal	**0.27**	**0.25**

6.5.3 Instructions to Participants

All participants were read the following instructions, followed by an exchange it which the experimenter answered any (permissible) questions that the participants had about the task.

For this study we are trying to understand the factors that affect where you decide to direct your visual attention. We are hoping to use this information to guide display design in airplane cockpits, so it is very important that you try and do your best throughout the study. We will be giving the top performer in each group (of five people) a $50 bonus. I will tell you more about that when I explain what you will be doing.

It is important that you attend your session each day. If you cannot make your scheduled session please let me know as soon as possible, and we can try and re-schedule another time in the day. For this study we are asking you to complete 8 five-minute trials per day for the entire week. You will have a short break following each session and a longer break in the middle. We will make a video recording of each session to look at your eye movements.

To obtain accurate eye-movement measures, it is necessary that you are a predetermined distance from the computer monitor. We have set up the computer monitor so that it is the correct distance away from your eyes when you are lined up with this string. … Please try and keep this distance from the monitor during the entire set of trials. We have also set it up that the correct distance is when your eyes are lined up with the edge of the desk. If you are not the correct distance from the monitor during one of your trials, we will let you know.

At the beginning of each trial you will see a box that asks you to enter your ID. You don't need to put anything in this box. Pressing start will begin your session. Please make sure an experimenter is in the room and tells you to begin before you press start.

As you can see there are four gauges on the screen. Each gauge has a pointer that is initially pointing upwards. Once you press start, all of these dials will begin to move back and forth. There is a safe range for all the values as indicated by the light gray areas, and an alarm range as indicated by the dark gray areas. Your goal is to detect when each of the gauges goes into the alarm range by pressing a button on the keyboard corresponding to the gauge that is currently out of bounds. You have one second to press the button once the gauge has gone out of bounds. If the upper left gauge is out of range, press the "a" key. The lower left gauge corresponds to the "z" key. If the upper right is out of bounds press the " ' " key. Lower right is the "/" key. You only need to press the button once each time the pointer goes out of range. If you press it more than once each time it goes out of bounds, it will be considered a false alarm (which we will go over later). So, your overall goal of this experiment is to monitor all four gauges continuously and detect when any one of them goes out of bounds.

Also, in the alarm range of each gauge, there are bars of different lengths. These bars are used to indicate the number of points each gauge is worth. The longest bar is worth 8 points, the bigger medium-sized bar is worth 4 points, the smaller bar is worth 2, and the smallest bar is worth 1 point. Each time you correctly detect that a gauge has gone out of bounds you will receive the number of points indicated by the length of the bar. There are two ways to loose points. If you miss that an alarm has gone out of range you will loose the number of points based on the value of the gauge. You will also loose the same number of points, based on the value of the gauge, when you press the key indicating that the gauge has gone out of bounds when it is still in the safe zone (light gray). If you press a key more than once when the alarm is out of bounds, you will not receive more points. This is considered to be a false alarm, so you will loose points.

Your overall score will be the positive points you earn for each time you correctly detect that a gauge has gone out of bounds minus the points you loose for both misses and false alarms. Since there are many ways to loose points, it is possible (or even likely) that you will have negative points – especially at the beginning.

Your total score will be given to you at the end of the each trial.

To calculate your score for the bonus at the end, we will take the average of your top five scores. over the whole week. The top performer in each group (1 in 5 chance) will receive an extra $50.

6.6 Results

Recall that a participant's score for a 5 minute trial was calculated as the number of hits for each gauge times the value (1, 2, 4, or 8) of each gauge, less the number of misses plus false alarms associated with each gauge times this same value. Prior to data analysis, there was a need to normalize this score (across all trials) due to the fact that some trials had a slightly higher alarm rate than .25 per second, and some a slightly lower rate (thus differing in maximally achievable score), as discussed in Section 6.5.2 and Table 17 above. As such performance scores were normalized to reflect the number of hits actually possible per trial, based on the information in Table 17. Since this normalization was required, we decided to simultaneously normalize the participants' scores prior to analysis so that a score of 100 always meant the highest score achievable per trial, with lesser scores reflecting the percentage deviation from peak performance. Recall, however, that especially in early trials, that participants total scores may be negative (due to a high number of misses and false alarms relative to the number of hits).

Since the number of hits possible varied slightly as a function of actual alarm rate (as shown in Table 17), we decided to calculate a normalized performance score as follows. The potentially achievable score (PAS) was naturally the sum, over the 5-minute trial, of the number of hits per gauge times the value associated with each gauge. This PAS value for a given 5-minute trial was used in the denominator of an equation to obtain a score reflecting a percentage of the PAS actually achieved. The numerator of that equation was the number of actual hits actually achieved per gauge times the value associated with each gauge, less the number of false alarms plus misses per gauge, also weighted by the value associated with each gauge. This normalized performance measure, where a value of 100 indicated maximally obtainable performance, is referred to simply as *score* in the following.

In addition to results based on the score measure, we also present results on hits, false alarms, and dwell time percentages (per gauge) calculated from eye movement data. Data on misses are excluded because misses are simply the complement of hits. Again, unless specifically mentioned otherwise, the data and results discussed below are based solely on the first 36 trials of the experiment, prior to four transfer trials in which the EV value was changed without informing participants of this intervention. Finally, we again note that at the time of this report we are only halfway though data collection, so in addition to the following analyses reflecting a relative small N (5 per EV condition), we have yet to perform a variety of additional analyses that we expect to shed even greater light on the collective influence of bandwidth, value, and ecological validity on visual attention allocation.

6.6.1 Overall Score

Figure 28 provides a concise snapshot of the results of the experiment in terms the influence of ecological validity as a moderator of performance. Analysis of variance of the unpooled data (i.e., not averaged over participants nor the 8 trials in each of days 1-4, nor the four pre-transfer trials on day 5), indicated significant effects practice, $F(33, 509) = 16.09$, $p < .01$ and EV, (F(2, 35) = 13.08, $p < .01$, but not their interaction ($p = .356$). Pairwise (by EV) analyses revealed that the EV = 1 group outperformed the EV = .75 group ($p < .01$) and also the EV = .25 group ($p <$ 01), and *suggested* ($p = .12$) that the EV = .25 group actually may have outperformed the EV =.75 group.

We are naturally curious to determine if the additional power gained by running the 2nd half of the experimental participants will provide additional evidence regard to this latter finding. Should that be the case, such a result, while perhaps initially surprising, *may* be a reflection of the fact that it could be harder to make use of highly salient but imperfect (EV = .75) proximal (i.e., not averaged over participants nor the 8 trials in each of days 1-4, nor the four pre-transfer

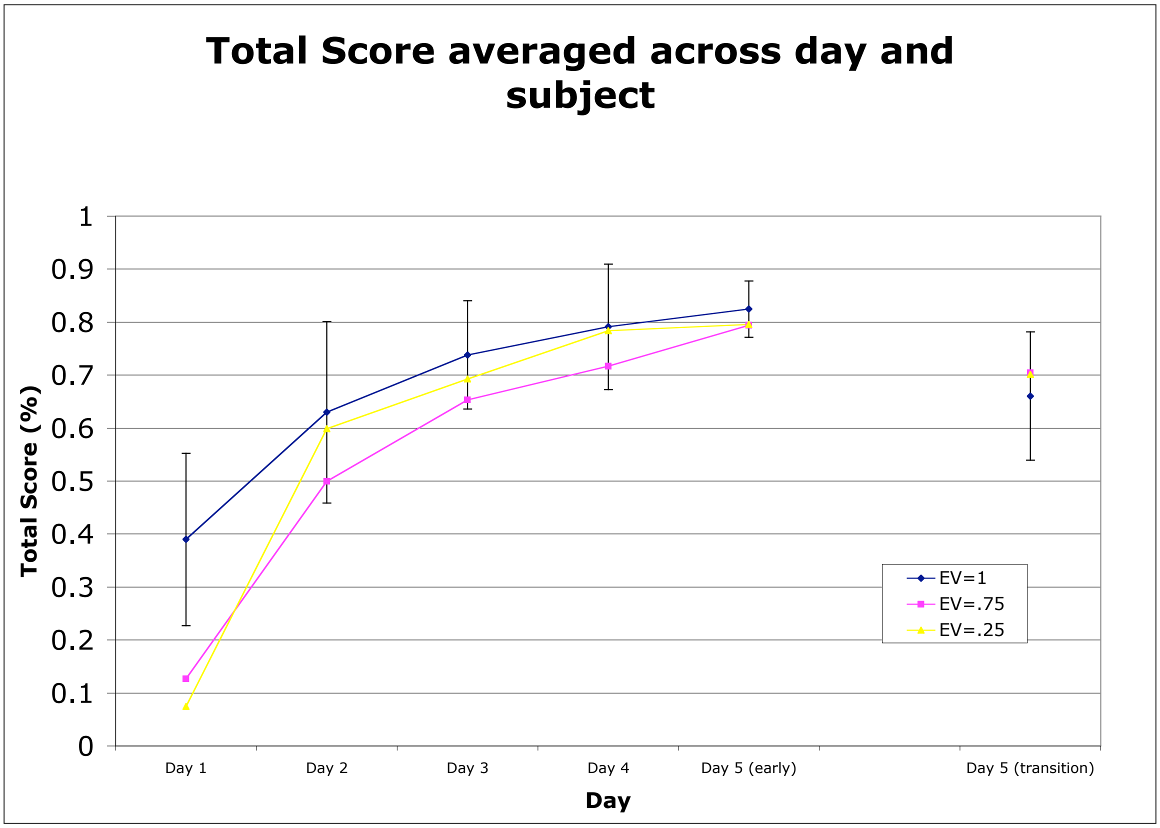

Figure 28. Total Score by Ecological Validity Condition Averaged Over Group and Day

trials on day 5), indicated significant effects practice, $F(33, 509) = 16.09$, $p < .01$ and EV, ($F(2, 35) = 13.08$, $p < .01$, but not their interaction ($p = .356$). Pairwise (by EV) analyses revealed that the EV = 1 group outperformed the EV = .75 group ($p < .01$) and also the EV = .25 group ($p <$ 01), and *suggested* ($p = .12$) that the EV = .25 group actually may have outperformed the EV =.75 group.

We are naturally curious to determine if the additional power gained by running the 2nd half of the experimental participants will provide additional evidence regard to this latter finding. Should that be the case, such a result, while perhaps initially surprising, *may* be a reflection of the fact that it could be harder to make use of highly salient but imperfect (EV = .75) proximal information in adapting to a criterion, than it is to adapt to the perhaps less readily available, but true task criterion itself (alarm frequency). Under this interpretation, the notion is that perhaps the EV=.25 group may have realized early on that bandwidth was a very poor predictor of alarm frequency, and thus did not attempt to make use of bandwidth, despite its perceptual salience, in adapting to the demands of the task. While the statistical support for this hypothesis remains somewhat weak to this point, in the following section we will present the results of regression modeling of dwell time per gauge, using bandwidth and value as predictors, that provides converging evidence in this regard. Finally, note that Figure 28 includes, at the far right, the mean score for each EV group across the final four transfer trials. While the graph gives the appearance of a general decline across all 3 EV groups, we will examine this issue at a finer grain of detail in a following section.

Further analysis revealed that the improvement in performance shown in Figure 28 across all three EV groups owed to an non-EV specific decrease in the number of false alarms generated, as shown in Figure 29, and an EV-specific increase in the number of hits, as shown in Figure 30. That is, in the analysis of false alarm data, EV was not a significant predictor ($p = .77$), but it was a significant predictor in the analysis of hit data $F(2, 35) = 17.48, p < .01$. These findings are consistent with our initial expectations. Increasing one's number of hits is a measure of adaptation to the statistical and value structure of the task environment, a structure that was intentionally manipulated across EV groups. On the other hand, false alarms are essentially self-generated errors of commission, and there was no *a priori* reason to believe that a manipulation of EV would influence this form of learning in this task.

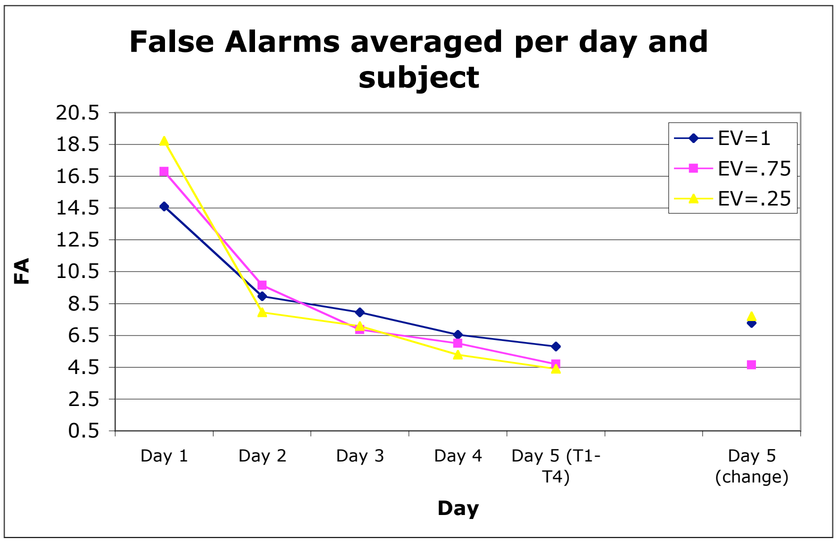

Figure 29. False Alarms by Ecological Validity Condition Averaged Over Group and Day

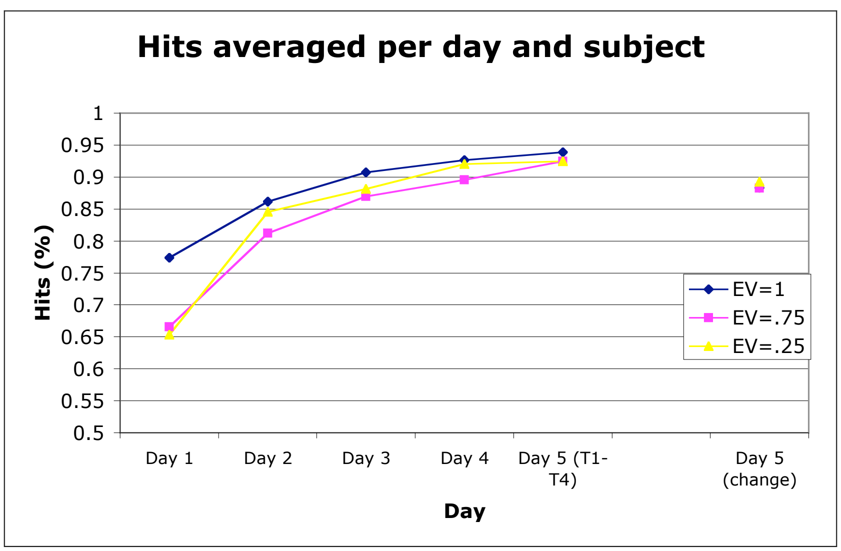

Figure 30. Hits by Ecological Validity Condition Averaged Over Group and Day

Even at the level of data aggregation represented in Figures 28-30, it is clear that all three EV groups managed to learn to perform the task to (near) asymptotic levels, as represented by the flattening of the graphs over days. This level represented approximately 80% of the maximal achievable points that could be scored in an experimental trial. As additional evidence for participants reaching asymptote, it should be noted that the trial on which the overall (across all EV groups) highest mean score was achieved in the experiment occurred on Trial 34, the second, rather than the fourth (last), trial on the final day of the experiment prior to transfer. It is also clear that ecological validity plays a significant role in mediating adaptation, and as will be seen below, visual attention allocation in this task as well.

6.6.2 Regression Modeling of Eye Movement Data

In order to shed light on the collective influence of EV, bandwidth, and value on visual attention allocation, we coded the eye movement data for both an early trial (Trial 2) and a late trial (Trial 34, the abovementioned trial at which overall task performance across EV groups peaked). In particular, we coded these data in terms of the percentage of time the participant spent fixating on each of the four gauges during the 5-minute experimental trial. Then, for each EV group separately (1.0, .75, and .25), we created regression models attempting to predict these percent dwell times from the particular bandwidth and values of the gauges those participants viewed. Thus, these models were functionally in the spirit of the SEEV models discussed previously.

We executed two different model formulations for each data set (there were 6 data sets: 3 EV conditions x the early vs. late trials). For each set, we created regression models that attempted to predict percent dwell times (per gauge, over a 5-minute trial) by both additive (bandwidth + value) and multiplicative (bandwidth x value) predictors (the latter being consistent with the latest formulation of the SEEV model). As such, in total we created 12 regression models (the 6 data sets x the additive vs. multiplicative models. Notably, in every one of the 12 cases, we achieved higher both R^2 and adjusted R^2 values for the *additive* rather than the *multiplicative* regression models. Additionally, 5 of the 6 additive models explained a significant ($p < .05$) portion of the variance in dwell time percentage, whereas only 3 of the 6 multiplicative models

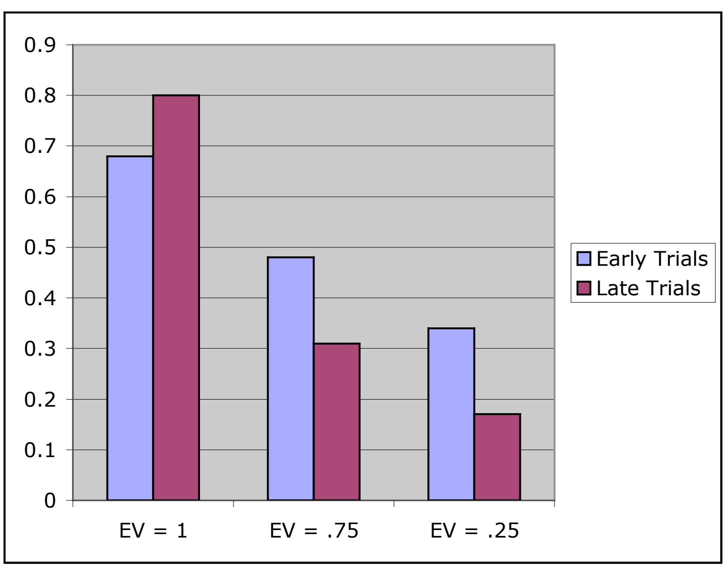

Figure 31. R^2 Values for Linear Regressions Predicting Dwell Time from Bandwidth and Value

did so. Note that this finding contrasts with the current formulation of the SEEV model, which assumes bandwidth and value combine multiplicatively in influencing attention allocation. However, due to the fits obtained we will restrict our presentation below to the additive models.

Figure 31 (above) depicts the fit (R^2) of the 6 linear-additive models broken down by both EV condition and whether the eye movements were measured on early (Day 1, Trial 2) versus late (Final Day, Highest Overall Performance) trials. The only one of these 6 models which failed to explain a significant portion of the variance was the model of dwell time per gauge in the EV = .25 condition during late trials (when these participants had maximized their performance). Consistent with the speculation made earlier in Section 6.6.1 that perhaps these low EV condition participants may have recognized that bandwidth was not a reliable cue, it may not be a coincidence that a model using bandwidth as a predictor fails to explain the performance of these participants while they were at their level of peak performance. To get additional insight into this issue, we now consider the patterns of cue usage (i.e., beta weights) revealed by modeling these data.

6.6.2.1 Bandwidth and Cue Usage Patterns by Ecological Validity Condition

To enable the beta weights produced by regression modeling to be meaningfully interpreted, we normalized the predictor variables (bandwidth and value) to a common [0, 1] range prior to modeling. Figures 32, 33, and 34 depict how the beta weights or “cue usage” patterns are influenced by EV and practice. For these graphs, when the regression analysis revealed that a particular beta weight did not significantly differ from zero ($p > .05$), it is graphed as a zero value (i.e., it does not appear).

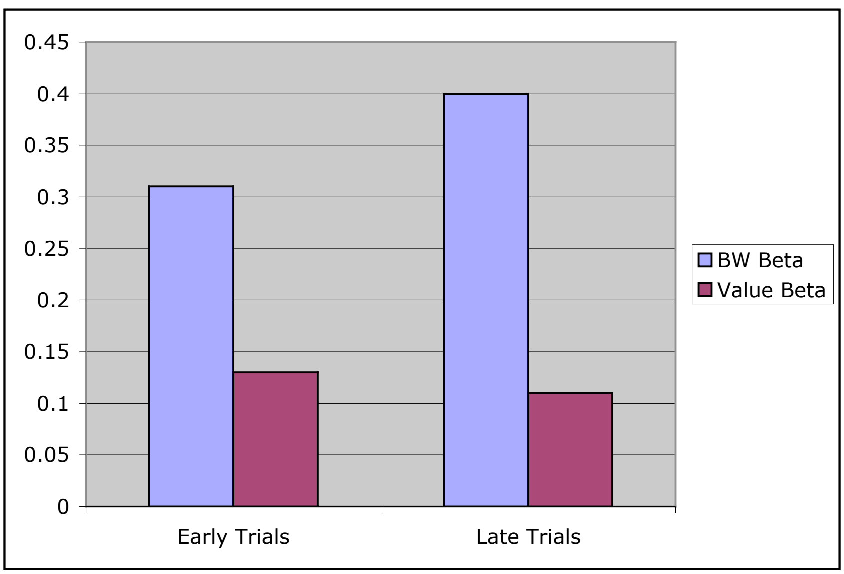

Figure 32. Beta Values for BW and Value in the **EV = 1 Condition**

Figure 32 is consistent with what one might expect from an analysis of an alarm detection situation where a performer is provided with a cue (bandwidth) that is both highly salient and highly predictive of alarm frequency, but must also tailor the use of this cue to an additional concern, the value achieved by detecting the various alarms (some weighting of value). Since EV=1 in predicting alarm frequency (though independent of the points earned for detecting an alarm), it is not surprising to find that participants in this condition appeared to become increasingly sensitive to bandwidth as a cue over experience (although statistical testing of this trend awaits). What is notable about Figure 32 is that the degree to which the *value* of a gauge influenced the visual dwell time percentages: EV=1 participants did not change over experience. While it is possible that this group already displayed the optimal adaptation to value by Trial 2, it is unlikely. Shown as static bars against a background of often wildly changing gauge needles, one would think that adaptation to value would have come more slowly than adaptation to bandwidth, at least in shaping eye movements and dwell durations. The data from Figure 32 may indeed reveal that all the learning displayed by the EV=1 group was associated solely with adaptation to bandwidth or alarm frequency (as they were perfectly correlated), rather than value.

Figure 33, in contrast, tells quite a different story. These participants performed the task in the EV=.75 condition, where bandwidth was an imperfect predictor of alarm frequency. While this would naturally make this a more challenging task, the less than perfect validity of the bandwidth cue appeared to *also* have the side effect of rendering their instrument scanning behavior completely uninfluenced by the differing values associated with the various gauges.

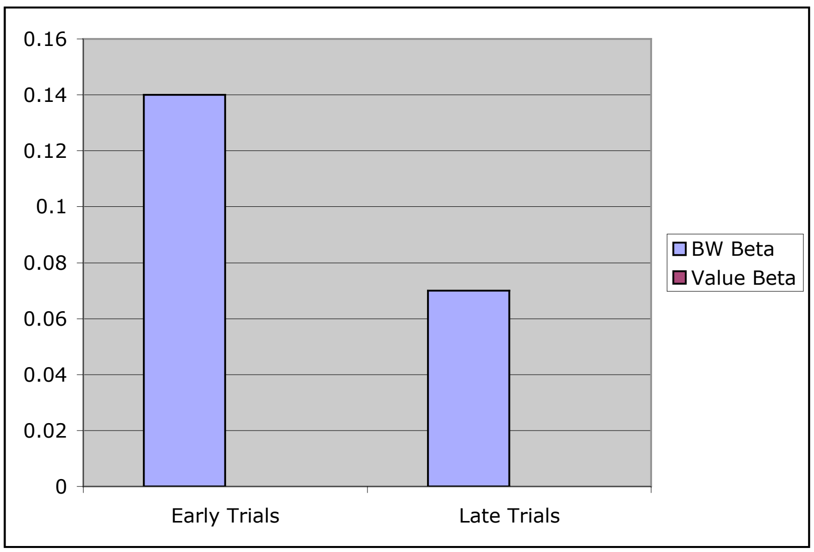

Figure 33. Beta Values for BW and Value in the **EV = .75 Condition**

Note that Figure 33 indicates an complete lack of sensitivity to value information for these participants (the beta weights on value in neither the early nor late trial regression models in this EV condition were not significant from zero). Since this was the only EV condition in which a complete lack of adaptation to value information was evidenced in gauge dwell times (see below for a discussion of the EV = .25 condition), one might *speculate* that this result may be due to the problem of negotiating an environment which provides neither a highly salient and perfectly reliable cue (EV = 1), nor an environment in which one can readily recognize that the most salient cue (bandwidth) is of little value, and can then focus attention on adapting to other task-relevant features of the environment, such as the criterion itself (EV=.75). This speculation clearly requires additional empirical scrutiny.

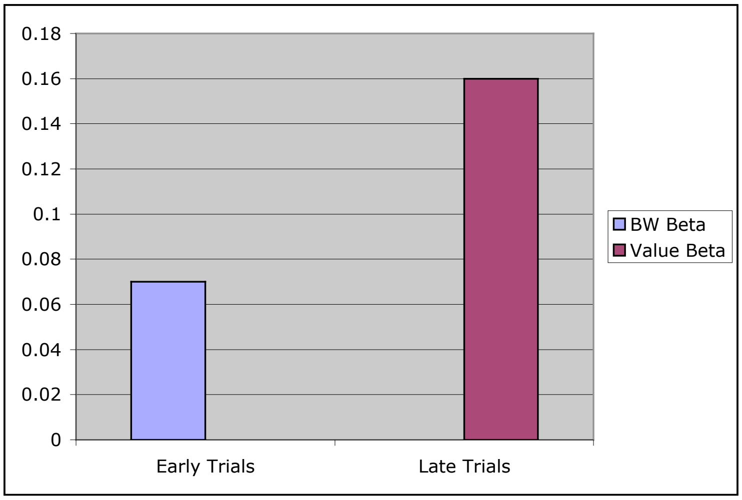

Figure 34. Beta Values for BW and Value in the **EV = .25 Condition**

Figure 34, however, does lend at least some additional support to the above speculation. This figure depicts the change in beta weights from early to late trials for the EV=.25 condition participants. Note the initial minimal, yet significant weighting on the highly salient bandwidth cue, with no initial weighting on value. In contrast, the EV=1 participants (see Figure 32) displayed early, presumably adaptive (note the extremely large performance difference between the EV = 1 group and the other two groups on Day 1 from Figure 28) reliance on *both* bandwidth and value cues in the initial sessions. The "early" component in Figure 35, like Figure 33 as a whole (EV = .75), may possibly reflect the difficulty of trying to effectively use two cues when the more salient one is only partially informative.

The striking difference between these latter two figures, however, could be explained in the following way. By the end of the experiment, it is quite possible, and indeed largely adaptive, for the EV=.25 participants to have largely given up on the bandwidth cue (perhaps quite early on). As a result, (unlike the EV = .75 group who may have been still struggling to adaptively use an only partially informative cue), this group would then be free to focus attention on adapting to a remaining task-relevant cues, in this case, value and alarm frequency itself. Note that of all three groups, the EV=.25 group displayed the highest beta weight on value in the late trials, although this conclusion awaits statistical testing.

Clearly, however, Figure 34 indicates that unlike both the EV = 1 and EV = ,75 groups, the EV = .25 group displayed no variance in (the measured aspects of) visual scanning associated with bandwidth by the end of the experiment. In short, all evidence points to the conclusion that they had given up on this cue, however salient it may have been. The results from this group thus provide additional support for the assumption of Wickens and his colleagues that salience will not be a good predictor of the attention allocation of well adapted operators, as discussed previously in our literature review. Our results lead us to temper their observation slightly, however. Our results from the EV=1 condition suggest that when a highly salient cue is also highly predictive, it will indeed be used, and continue to be used, throughout the course of adaptation. One general lesson learned from this experiment is that statements about the influence of various predictors of visual attention allocation must be conditioned on an examination of the uncertainty (in our case, the varying levels of EV), of the task environment in which adaptation takes place.

6.6.2.2. Analysis of the Transfer Condition

Recall that in the final four trials of the experiment, the EV=1 group performed the EV=.75 condition, the EV=.75 group performed the EV=.25 condition, and the EV=.25 group performed the EV=1 condition. What was the effect of this manipulation? Empirically, the answer appears to be quite simple, as demonstrated in Figure 35 below. This figure indicates that the EV = .25 group was able to successfully maintain its level of performance post-transfer, while the other two groups were hurt by the transfer manipulation. However, the statistical analysis is only partially supportive of this interpretation, as ANOVA indicated a p value of only .16 for the main effect of EV, and an even higher value for the apparent interaction effect. We should note that this is one result that may clearly be due to a combination of a small sample size and the existence of one outlier in the EV = .25 condition. Specifically, for the two post-transfer trials entered into the ANOVA, the five EV = .25 participant scores are shown in Table 18.

Table 18. EV=.25 Participant Scores in the Final Two Post Transfer Trials

Participant	Score on Trial 39	Score on Trial 40
1	0.793	0.980
2	0.838	0.869
3	**0.283**	**0.546**
4	0.910	0.882
5	0.967	0.978

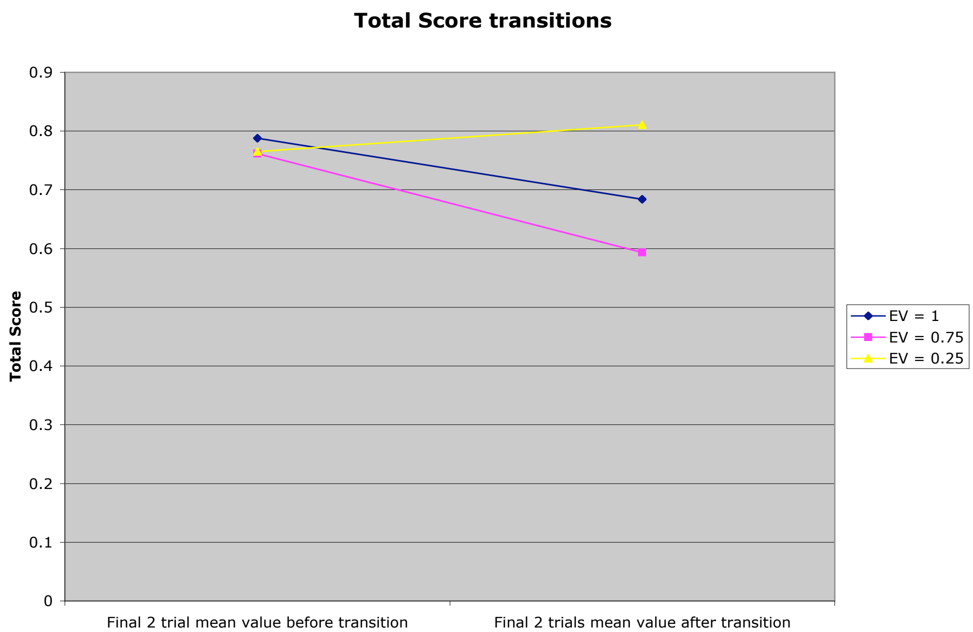

Figure 35. Pre- and Post-Transfer Performance Scores by EV Condition

Due to the low N and the highlighted outlier in Table 18, we will indeed be examining the transfer effects in more detail once the second half of the experiment is completed. Note that should the trend indicated in Figure 35 be found reliable, at least two interpretations of this effect are possible. In the first case, it may simply be due to the fact that two of the groups have been transferred to a less informative environment while the third group (EV = .25) has been transferred to a more informative environment. While perhaps the most elegant, and clearly the simplest explanation, if correct it should be noted that this would represent remarkably fast adaptation to the bandwidth cue on the part of the EV=.25 participants, especially after 36. 5-minute trials in which that cue had very low ecological validity. Recall that Figure 33 indicated that by the late (pre-transfer) trials, the beta weight for bandwidth for this group could not be said to be different from zero, suggesting that at the time the transfer trials were administered, this group had largely learned to ignore even an extremely salient visual cue during visual scanning. To conclude that they were the only group not "hurt" by the transfer manipulation because they were the only group transferred into a more informative environment is actually a fairly extreme claim about the rate with which one can acquire reliance on a cue one has learned over 36, 5-monute trials, to largely ignore.

A second possible interpretation of these data would be that all previous evidence points to the fact that the EV = .25 group had dispensed with reliance on the bandwidth cue entirely, and had learned to scan displays in a manner influenced by value and perhaps other cues, including the criterion itself (alarm frequency). As such, one might expect that changing the ecological validity of the bandwidth cue would leave their scanning, and thus their performance, relatively unaffected, as compared to the other two groups, who were still relying on the highly salient bandwidth cue at the time of transfer. We expect that a detailed analysis of the eye movement data, not yet complete to date, may enable us to adjudicate between these competing explanations, and also shed additional light about the many questions that remain unanswered in our experimentation and analysis to date.

7. Future Work

We believe that this research domain is very rich and there are many things that could be done to extend the contributions reported here. First, there is near-term work on finalizing the process modeling to be done in the time remaining on the project, including, perhaps most notably, the integration of the Land/Go Around decision logic that we have developed into the ACT-R model. Second, we intend to complete the experimentation and analysis reported in Section 6, aimed at bridging, and mutually informing, process and product models of visual attention allocation in human machine systems. Third, we intend to refine the visual attention allocation components of the ACT-R model in light of any relevant findings obtained by experimentation. Fourth, there are longer-term possibilities for this work and other work which could be derived from it.

7.1 Finalizing the Current Project

There are a number of things to be finished in the months remaining. First and most obviously, the go/no-go decision model needs to be integrated with the ACT-R model. This will allow us to do the same evaluation and validation of the model for Phase 3 as was done for the first two phases. Second, the exact nature of the relationship between the current ACT-R model and the Wickens SEEV model should be refined, using the empirical study presented in Section 6 as as an initial point of contact, one that we believed could be mined much more deeply.

We would also like to undertake another form of validation. One of the things the ACT-R model is sensitive to is exactly what information is displayed where. The model qualitatively predicts that the addition or removal of instruments from the SVS overlay will affect both how and when the pilots would look at the SVS, and this may impact their downstream situation awareness as indexed by the model's representation of the aircraft state. In essence, we should be able to play out "what if" scenarios, such as "what if the altitude was not overlaid on the SVS?" or "what if the state of the flaps was overlaid on the SVS?" While not strictly validation, we think this is a an avenue with great potential for modeling, since running a new field trial with a tweaked SVS is both difficult and expensive, but running the model with a tweaked SVS is not—this could potentially show off the value-added of modeling.

7.2 Future Extensions

There are numerous ways in which this research may be followed up or extended. One possibility is a very direct extension of the present work. While we have not had the time to extend this to other airports, aircraft, approach plates, and scenarios, such extensions should be possible. There may be some real value in such an approach to evaluating the SVS, since by necessity the experiment already run and field trials are limited in the level of risk they may present to human pilots. However, this is not an issue for a simulated pilot. This may be useful in evaluating how the SVS may (or may not) be of assistance in extreme circumstances.

We are also interested in extensions to the theoretical work which have been suggested by this line of research. One such avenue in this vein is the aforementioned attempt to relate ACT-R's mechanisms with the higher-level SEEV model. It would be interesting to consider the relationship between the current research and other related frameworks for considering problems of attention allocation and visual sampling, akin to our experimentation motivated by the original Senders (1964) research. We would like to explore the relationship between other models which also touch on ideas involving allocation of cognitive resources, such as information foraging theory (Pirolli & Card, 1999).

In addition, we believe that we may have some leverage on some other high-level and abstract human factors constructs, such as "situation awareness." There is no box or section of the ACT-R architecture that one could point to as being situation awareness. Rather, we have observed that the model has to keep a number of pieces of information available at various times (some things, like altitude, all the time); the accessibility of the set of needed information about the aircraft's state might be termed the model's situation awareness, but it is not a unitary thing. It is both distributed, in that it lives in multiple declarative memory elements, and dynamic, in that different pieces are needed and "refreshed" by checking the environment at different rates. We hope ultimately that this work will lead toward more formal definitions, at least in an ACT-R context, of a number of terms from the human factors literature (e.g., situation awareness, workload) that are currently somewhat vague.

8. Implications and Lessons Learned

Computational cognitive modeling is about more than just generating model fits; perhaps more important than those are the insights generated by the work. One of the larger goals of this project is to help understand the impact of the SVS, about which we believe we have important insights. However, there are also wider lessons which were learned along the way.

8.1 Synthetic Visual Systems in Commercial Cockpits

It seems reasonably likely that SVS technology will indeed make its way into commercial cockpits sometime in the indeterminate future. What does our modeling work have to say about such a deployment? We believe there are numerous ramifications.

First, as has been mentioned previously, while the intent of an SVS may be to serve simply as a proxy for looking out the window in circumstances when looking out the window is not terribly useful, it is clear that pilots can and almost certainly will use the SVS not simply in this role, particularly if other information relevant to pilots' goals is displayed there. This leads to what we believe is one of our most important take-home messages for SVS design: the overlaid symbology matters, and may matter a great deal. Ultimately, someone will have to decide for SVS systems deployed in commercial aircraft what to overlay. We believe it would be egregiously bad for the content and visual properties of the overlay to be lightly considered, as it can have a tremendous impact on pilots' attention allocation. Note that the overlay used in this research is nearly identical to the symbology available on at least one commercial HUD system. The question our research raises here is whether or not this is the right thing to do, or was this done just because it resembles another display? The SVS used in Prinzel, et al. (2002) had substantially different symbology, including the moving vertical strips for altitude and airspeed found on the typical PFD. We believe it is important for the system designers to very seriously consider what information should and should not included on the overlay, and how what is included is displayed.

Furthermore, we believe that the overlay should be dynamic in the sense that it should display different pieces of information at different times. Pilots have different information needs in various circumstances (e.g., near waypoints) and we believe this can and should be addressed.

Additionally, we see an opportunity. While few commercial carriers in the U.S. are currently making extensive use of RNAV approaches, it is likely that these will increase in frequency in the future. Since any SVS-equipped aircraft will have high-precision GPS integrated with other flight systems, RNAV approaches could even become common. However, in many current commercial cockpits, even those with RNAV-capable instrumentation, display support for RNAV is not optimal. In many cases, one of the pieces of information useful for RNAV landings, the vertical deviation from the RNAV path, is only displayed on the head-down FMC display, and only after that display is toggled into the appropriate mode via the head-down keypad.

However, such additions much be considered carefully. By altering the symbology, it should be possible to greatly influence how much attention pilots allocate to the SVS. By putting more flight-relevant information on the SVS, it is likely that more attention would be allocated to it, even in phases of flight where use of the SVS is not a particular goal. However, this needs to be balanced against need for other information which is not displayed on the SVS, particularly information like warnings. This problem is already starting to become apparent among pilots who fly with HUDs, who report that sometimes they catch themselves "flying the HUD" rather than flying the aircraft. There are tradeoffs here between having enough appropriate information to support pilots and having too much, causing them to be drawn in. While the current empirical results and modeling work are not quite in a position to parameterize this kind of tradeoff, being able to do so in the future is not an unreasonable goal.

8.2 Human Performance Modeling

We learned, or had strongly reinforced, several other lessons which apply more generally than to just the SVS. In general, the lessons learned in the first project on taxiing applied here as well. For example, we believe it would have been extremely difficult, if not impossible, to model the data down to the level of individual saccades without a “closed-loop” model which includes both the human in the loop as well as the environment. In some sense it is not surprising that the SVS had such a large impact on attention allocation, as when one considers the wider human-machine system, the visual environment faced by the human pilot was substantially changed so a change in the interaction pattern is likely. However, while this is a good general platitude, it is too vague. It is only with modeling that we came to understand why pilots made such extensive use of the SVS when it was present and only through modeling that we had any leverage at all on quantitative prediction of transition frequency.

Beyond the quantitative benefits of modeling, we also feel that modeling in this style forces the analyst(s)—in this case, us—to consider multiple interacting factors which jointly constrain interactive behavior. One of the most significant barriers to the development of human factors interventions based on the systems view is the lack of techniques and models capable of simultaneously representing the many potential factors contributing to human performance, and how these factors interact in typically dynamic, often complex, and usually probabilistic ways. To say that multiple contributing factors “conspire together” to produce behavior is one thing. To provide techniques capable of representing these multiple factors, and the precise manner in which they “conspire” is quite another.

As a step toward addressing this problem, we took up a broader variant of a perspective first sketched out elsewhere (Byrne, 2001; Gray, 2000; Gray & Altmann, 1999) and presented in Figure 36 as the interactive behavior triad (IBT). That is, we consider interactive behavior, such as a pilot interacting with a cockpit, as jointly constrained by three factors: embodied cognition, the environment, and the task. We take “embodied cognition” to mean the capabilities and limitations of the integrated human perceptual-cognitive-motor system. We take the environment to include the constraints and affordances provided to the operator (in this case, the pilot) by his or her environment. This includes the various artifacts and displays in that environment, the properties of the system being controlled directly or indirectly by the pilot (e.g., the aircraft), and the broader environment in which that behavior is embedded (e.g., cockpits in general). Finally, we take the task to mean the goal, or more often, the set of goals the operator is trying to accomplish and the knowledge necessary to fulfill those goals, per the description in Card, Moran, and Newell (1983).

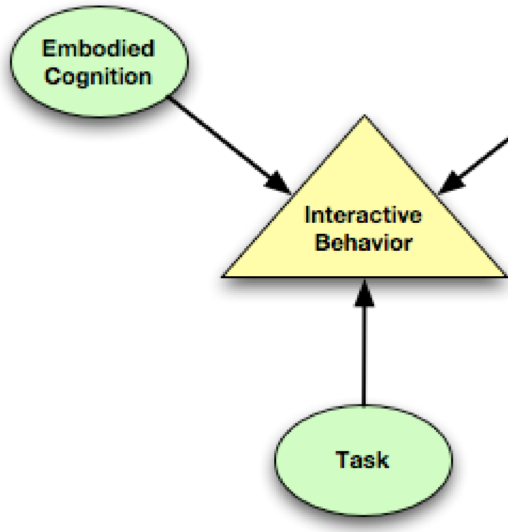

Figure 36. Interactive behavior triad

As Gray and Altmann noted, traditional disciplines have generally considered these three entities in a pairwise fashion, rather than as a triad. The experimental psychology community has typically considered the operator, but often with either artificial tasks or in context that minimize or eliminate the role of the environment. Ethnographic or ecological analysis typically considers the environmental context and the tasks, but often overlooks issues rooted in the capabilities and limitations of the human element. Engineers have traditionally considered the design of artifacts (an aspect of the environment) to support particular tasks, but often ignored constraints imposed by the capabilities and limitations of the user. And indeed, there are likely many situations in which all three need not be considered. For example, there are situations where the environment provides sufficient constraints that an analysis of the environment alone, or with a few relatively uncomplicated assumptions about task and environment, can generate useful predictions, even in the aviation domain (e.g., Kirlik, 1993, Casner, 1994). And many task analysis techniques have been very successful with only the crudest assumptions about human capabilities and environmental limitations (e.g., Kirwan & Ainsworth, 1992). However, we believe that to understand and predict human performance in many real-world situations, all three sources of constraint must be jointly considered and, ideally, incorporated into quantitative modeling formalisms.

The particular ACT-R approach we took essentially requires joint consideration of all three factors because the model will not run without a detailed specification of all the interacting pieces. We believe being forced to consider all three factors is good discipline even if it is not always possible to produce a perfect fit to the data.

However, we also had some of the drawbacks of this approach thrown into sharp relief during the project. The technical integration issues can be very resource-demanding and can often bottleneck the more substantively fruitful work. While it seems unlikely that this problem will ever go away entirely, there are numerous people working on aspects of this problem and over time, these problems can become more tractable. We seriously doubt that producing an ACT-R model of this complexity, linked to a real-time flight simulator and producing quantitatively respectable behavior, would have been remotely feasible with the ACT-R and associated software technology of ten years ago. But the technical details in fine-grained computational cognitive modeling can indeed be tiresome at times.

Overall we believe efforts like the present one will help facilitate future work of this type. The experience we have gained, as well as some of the code, could most certainly be re-used in similar projects in the future. We are also hopeful that work like this will become more common in the future as well; as costs of using highly-trained experts as subjects continues to rise and the modeling technology and methodology becomes more mature, modeling becomes more viable. We see no reason for that trend not to continue.

9. References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Quin, Y. (in press). An integrated theory of the mind. To appear in *Psychological Review*.

Bowers, C. A. and Jentsch, F. (2001). Use of commercial, off-the-shelf, simulations for team research. In E. Salas (Ed.), *Advances in Human Performance and Cognitive Engineering Research, Vol. 1*. Mahwah, NJ: Erlbuam. (pp. 293-318).

Byrne, M. D. (2001). ACT-R/PM and menu selection: Applying a cognitive architecture to HCI. *International Journal of Human-Computer Studies, 55(1)*, 41–84.

Byrne, M. D., & Gray, W. D. (2003). Returning human factors to an engineering discipline: Expanding the science base through a new generation of quantitative methods—preface to the special section. *Human Factors, 45*, 1-4.

Byrne, M. D., & Kirlik, A. (in press). Using computational cognitive modeling to diagnose possible sources of aviation error. To appear in *International Journal of Aviation Psychology*.

Carbonell, J. R. (1966). "Queueing Model of Many-Instrument Visual Sampling." *Transactions on Human Factors in Electronics HFE-7(4)*: 157-164.

Carbonell, J. R., J. L. Ward, et al. (1968). "A Queueing Model of Visual Sampling: Experimental Validation." IEEE Transactions on Man-Machine Systems MMS-9: 82-87.

Casner, S. M. (1994). Understanding the determinants of problem-solving behavior in a complex environment. *Human Factors, 36*, 580–596.

Craik, K.J.W. (1947). Theory of the human operator in control systems. In: The operator as an engineering system. *British Journal of Psychology, 38*, 56-61.

Goodman, A., Hooey, B. L., Foyle, D. C., & Wilson, J. R. (2003). Characterizing visual performance during approach and landing with and without a synthetic vision display: A part task study. In D. C. Foyle, A. Goodman & B. L. Hooey (Eds.), *Proceedings of the 2003 NASA Aviation Safety Program Conference on Human Performance Modeling of Approach and Landing with Augmented Displays*: NASA Conference Proceedings NASA/CP-2003-212267.

Gray, W. D., & Altman, E. M. (1999). Cognitive modeling and human-computer interaction. In W. Karwowski (Ed.), *International encyclopedia of ergonomics and human factors.* New York: Taylor & Francis, Ltd.

Keller, J., Leiden, K., & R., S. (2003). Cognitive task analysis of commercial jet aircraft pilots during instrument approaches for baseline and synthetic vision displays. In D. C. Foyle, A. Goodman & B. L. Hooey (Eds.), *Proceedings of the 2003 NASA Aviation Safety Program Conference on Human Performance Modeling of Approach and Landing with Augmented Displays*: NASA Conference Proceedings NASA/CP-2003-212267.

Kirlik, A. (1993). Modeling strategic behavior in human-automation interaction: Why an "aid" can (and should) go unused. *Human Factors, 34*, 221-242.

Kirwan, B., & Ainsworth, L. K. (Eds.). (1992). *A guide to task analysis*. Taylor & Francis.

Moray, N. (1986). Monitoring behavior and supervisory control. In, K. Boff, I. Kaufmann and J. Beatty (Eds.), *Handbook of Perception and Human Performance*. New York, Wiley. H.

Moray, N. (1990). Designing for transportation safety in the light of perception, attention, and mental models. *Ergonomics 33*(10-11), pp. 1201-1213.

Moray, N. (1993). Designing for Attention. Attention: Selection, Awareness, and Control. In, A. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, Awareness, and Control*. Oxford: Clarendon Press.

Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review, 106*, 643-675.

Prinzel, L. J., Kramer, L. J., Comstock, J. R., Bailey, R. E., Hughes, M. F., Parrish, R. V. (2002). NASA Synthetic Vision EGE Flight Test. In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting* (pp. 135-139). Santa Monica, CA: Human Factors and Ergonomics Society.

Senders, J. (1964). The human operator as a monitor and controller of multidegree of freedom systems. *IEEE Transactions on Human Factors in Electronics, HFE-5*, 2-6.

Senders, J.W., Kristofferson, A. B., Levinson, W.H., Dietrich, C.W., & Ward, J.L. (1967). The attentional demand of automobile driving. *Highway Research Record*. 195, 15-33.

Senders, J.W. (1983). *Visual Scanning Processes*. Netherlands: University of Tilburg Press.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Science*. (pp 1-19).

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & R. Davies (Eds.) *Varieties of Attention* (pp.29- 61). New York: Academic Press.

Wickens, C.D., Tsang, P., Pierce, B. (1985). The dynamics of resource allocation. In W.B. Rouse, (Ed.), *Advances in Man-Machine Systems Research, Vol. 2.* Greenwich, CT: JAI Press.

Wickens, C. D., Gordon, S.E., and Liu, Y. (1998). *An Introduction to Human Factors Engineering. New York*: Longman. (pp. 392-395)

Wickens, C. D. and J. G. Hollands (2000). *Engineering Psychology and Human Performance*. Upper Saddle River, NJ: Prentice Hall.

Wickens, C. D., J. Helleberg, et al. (2001). Pilot task management: testing an attentional expected value model of visual scanning. Savoy, IL, UIUC Institute of Aviation Technical Report.

Wickens, C. D., J. Goh, et al. (2003). Attentional models of multitask pilot performance using advanced display technology. *Human Factors 45*(3): 360-380.

Wickens, C. D. (2004). Lesson 8 Class Notes. Psychology 404: Theories of Attention. Departmen of Psychology, University of Illinois at Urbana-Champaign.

Appendix 1

Land/Go Around Decison Logic Code

```
/**
 * gngLogic.java
 *
 * Version of go/no-go landing logic
 *
 * Customizability and compartmentalization emphasized
 * over Java features ... for clarity
 * ... and reproduction in other languages
 *
 * @author  Piotr Adamczyk
 * @version 0.1 - stable release
 */

import java.util.*;
import java.util.Arrays;

public class gngLogic {

        /**
         * Empty constructor method
         *
         */
        public gngLogic() {
    }


        /**
         * We've got three Opinion types Boolean, Fuzzy, and Bayesian
         * they can be used in the following decision styles...
         * (just set the variables as outlined)
         *
         * decideMajority - Boolean
         *  fuzzy = FALSE
         *  opnThreshold = [0...1]
         *
         * deicdeMajority - Fuzzy
         *           fuzzy = TRUE
         *  opnThreshold = [0...1]
         *
         * deicdeMajority - Bayesian
         *  opnThreshold = [0...1]
         *
```

```
     * decideDisjunction - Boolean
     *  fuzzy = FALSE
     *  opnThreshold = [0...1]
     *
     * decideDisjunction - Fuzzy
     *  fuzzy = TRUE
     *  opnThreshold = [0...1]
     *
     * decideDisjunction - Bayesian
     *  opnThreshold = [0...1]
     *
     * decideThreshold - Boolean
     *  fuzzy = FALSE
     *  opnThreshold = [0...1]
     *
     * decideThreshold - Fuzzy
     *  fuzzy = TRUE
     *  opnThreshold = [0...1]
     *
     * decideThreshold - Bayesian
     *  opnThreshold = [0...1]
     *
     * decideBayesian - Bayesian - Momentary
     *  updateBayes = FALSE
     *  opnThreshold = [0...1]
     *
     * decideBayesian - Bayesian - Temporal Difference
     *  updateBayes = TRUE
     *  opnThreshold = [0...1]
     */

/** Maim method - contains the main thread of decision logic
 *
 * @param args String array of arguments - does nothing
 */
public static void main(String[] args) {
  /** Variables under consideration
   *  + Altitude and Pitch/Roll for testing purposes
   *  - Offset, Alignment, and Approach for full version
   */
   double Altitude = 11000.0;
   double Pitch = 32.0;
   double Offset = 0.0;
   double Alignment = 0.0;
   double Approach = 0.0;
```

```
/** number of variables under consideration
 *  + 2 for testing, 5 for the full version
 */
int numVar = 2;
//int numVar = 5;

/** Storage for objective quality measures
 *  one for each variable
 */
double objAltitude = 0.0;
double objPitch = 0.0;
double objOffset = 0.0;
double objAlignment = 0.0;
double objApproach = 0.0;

/** Storage for subjective opinions
 *  one for each variable
 */
double opnAltitude = 0.0;
double opnPitch = 0.0;
double opnOffset = 0.0;
double opnAlignment = 0.0;
double opnApproach = 0.0;

/** Array for storing the opinions on vars
 *  makes passing the collection a bit simpler
 */
double[] opnArray = new double[numVar];

/** Array for storing the weights on vars
 *  makes passing the collection a bit simpler
 */
double[] weightArray = new double[numVar];

/** Storage for weights on subjective opinions
 * The weights assigned to a particular opinion in decision making
 * - a scaling factor from 0 to 1
 */
 double opnWeightAltitude = 1.0;
 weightArray[0] = opnWeightAltitude;
 double opnWeightPitch = 0.75;
 weightArray[1] = opnWeightPitch;
 double opnWeightOffset = 0.0;
 double opnWeightAlignment = 0.0;
 double opnWeightApproach = 0.0;
```

```
/** Aversion - a measure of confidence in the gauges and the data
 *  A real exponent used in the opinion scaling functions
 *  Aversion > 1 : objective values are discounted
 *  Aversion = 1 : objective values remain constant
 *  Aversion < 1 : objective values are promoted
 *
 *  So those who are more averse don't quite trust the gauges...
 */
 double aversion = 0.95;

/** Opinion Type - Fuzzy or Boolean
 *  A boolean flag as to whether opinions should be hard or fuzzy
 *  fuzzy = TRUE = fuzzy opinion, scaled [0...1]
 *  fuzzy = FALSE = boolean opinion, [0, 1]
 *
 *  --- Note: Both still go through the opinion calculation the same way ---
 *
 *  --- Note: if fuzzy is false, be sure to check the opnThreshold ---
 */
 boolean fuzzy = true;
 //boolean fuzzy = false;

/** Opinion Threshold
 *  A real value above which opinions become a GO, 1
 *    and below which become a NO-GO, 0
 *
 *  The higher the threshold, the more ideal a situation needs to seem for
 *   a pilot to be satisfied with it
 *
 *  opnThreshold = 1 = Always NO-GO
 *  ..
 *  ... everything in between
 *  ..
 *  opnThreshold =  0 = Always GO
 */
 double opnThreshold = 0.5;

 /** Prior probability for Bayesian reasoning
  * pLand - storage for the pilot's predisposition for landing
  *
  * pLand = 1 = Always GO
  * ..
  * ... somewhere in between
  * ..
  * pLand = 0 = Always NO-GO
  *
```

```
 */
 double pLand = 0.8;

 /** pLandArray - storage for the updates to the pilot's predisposition
  *  all the same for now, but might want different priors for each variable
  *
  *  pLand = 1 = Always GO
  *  ..
  *  ... somewhere in between
  *  ..
  *  pLand = 0 = Always No-Go
  */
 double[] pLandArray = new double[numVar];
 Arrays.fill(pLandArray, pLand);

 /**
  * updateBayes - Boolean flag, whether or not to engage the
  * temporal difference learning component
  */
 boolean updateBayes = true;
 //boolean updateBayes = false;

 /**
  * alpha - a learning rate for the temporal difference update
  */
 double alpha = 0.1;

 /** Go/No-Go - storage for the ultimate decision
  *  goNoGo = FALSE = No-go
  *  goNoGo = TRUE = Go
  */
 boolean goNoGo = false;

System.out.println("Testing the Logic");
System.out.println("-----------------");
System.out.println("Initial Altitude = " + Altitude);
System.out.println("Initial Pitch = " + Pitch);
for (int timestamp = 1; timestamp <= 15; timestamp++) {
   System.out.println("-- Time " + timestamp + " -------");
   Altitude = updateAltitude(Altitude);
   System.out.print("Altitude = " + Altitude + " : ");
   objAltitude = convertAltitude(Altitude, timestamp);
   System.out.print("Fitness = " + objAltitude + " : ");
   opnAltitude = opineAltitude(objAltitude, aversion, fuzzy, opnThreshold);
   opnArray[0] = opnAltitude;
   System.out.println("Opinion = " + opnAltitude);
```

```
    if (updateBayes) {
            pLandArray[0] = pLandArray[0] + alpha * (bayes(pLandArray[0], opnAltitude,
(1.0 - opnAltitude)) - pLandArray[0]);
    } else {
            pLandArray[0] = bayes(pLand, opnAltitude, (1.0 - opnAltitude));
    }

    Pitch = updatePitch(Pitch);
    System.out.print("   Pitch = " + Pitch + " : ");
    objPitch = convertPitch(Pitch);
    System.out.print("Fitness = " + objPitch + " : ");
    opnPitch = opinePitch(objPitch, aversion, fuzzy, opnThreshold);
    opnArray[1] = opnPitch;
    System.out.println("Opinion = " + opnPitch);
    if (updateBayes) {
            pLandArray[1] = pLandArray[1] + alpha * (bayes(pLandArray[1], opnPitch, (1.0
- opnPitch)) - pLandArray[1]);
    } else {
            pLandArray[1] = bayes(pLand, opnPitch, (1.0 - opnPitch));
    }

} /* END For */

System.out.println("\n-- Time to make a decision -------");

/**
 * decideDisjunction (array of opinions, array of weights, fuzzy?, big threshold)
 * look through the opinion array and if Fuzzy, if all of weighted opinions is above the
threshold then GO
 * if Boolean, if all of the weighted opinions is above the threshold then GO
 *
 * in other words, if any of the opinions are negative, either Boolean or Fuzzy below
threshold -> NO-GO
 */
 // goNoGo = decideDisjunction(opnArray, weightArray, fuzzy, opnThreshold);
 // goNoGo = decideDisjunction(pLandArray, opnThreshold);

/**
 * decideMajority (array of opinions, fuzzy?, number of variables, opinion threshold)
 * look through the opinion array and if Fuzzy, if a pure majority is above threshold then
GO
 * if Boolean, if a pure majority say GO, then Go
 */
 // goNoGo = decideMajority(opnArray, fuzzy, numVar, opnThreshold);
 // goNoGo = decideMajority(pLandArray, numVar, opnThreshold);
```

```
        /**
         * decideThreshold (array of opinions, array of weights, fuzzy?, big threshold)
         * look through the opinion array and if Fuzzy, if the sum of weighted opinions is above the
        threshold then GO
         * if Boolean, if the sum of the weighted opinions is above the threshold then GO
         */
         // goNoGo = decideThreshold(opnArray, weightArray, fuzzy, (opnThreshold*numVar));
         // goNoGo = decideThreshold(opnArray, weightArray, (opnThreshold*numVar));

        /**
         * decideBayesian (array of Bayesian posteriors, threshold)
         * calculate the product of all of the Bayesian posteriors - same effect as a long conjunction
         * if above the threshold, then GO
         */
         goNoGo = decideBayesian(pLandArray, opnThreshold, updateBayes);

        if (goNoGo)
           System.out.println("SAFE to Land");
        else System.out.println("--- NOT --- Safe To Land");

    } /* END Main */

  /**
   * Altitude - Descent rate
   * @param oldAltitude
   * @return
   */
    public static double updateAltitude(double oldAltitude) {
        double newAltitude;
        Random rand = new Random();
        double randDblValue;
        int randIntValue;

        /* make a safe (within the Fitness measure) drop 95% of the time */
        randDblValue = rand.nextDouble();
        if (randDblValue > 0.05) {
           /* So 95% of the time make a move within five percent of the altitude */
           newAltitude = oldAltitude - (0.05 * rand.nextDouble() * oldAltitude);
        } else
           /* While 5% of the time make a more dramatic 10% move */
           newAltitude = oldAltitude - (0.1 * rand.nextDouble() * oldAltitude);

        return newAltitude;
    }

    /**
```

```
 * Pitch/Roll (Wobble) rate
 * @param oldPitch
 * @return
 */
public static double updatePitch(double oldPitch) {
   double newPitch;
   Random rand = new Random();
   double randDblValue;
   int randIntValue;

   /* establish a cutoff between fine and coarse action
    * in this case, the angle from level flight
    */
   double cutoffPitch = 20;
   if (java.lang.Math.abs(oldPitch) > cutoffPitch) {
      /* If over 20 degrees from level, sharp move to level [0...30% of current] */
      randIntValue = rand.nextInt(10);
      newPitch = java.lang.Math.abs(oldPitch) - (0.3 * randIntValue *
   java.lang.Math.abs(oldPitch));
   } else {
      /* If under 20 degrees just a smaller move [0...3 degrees] */
      randIntValue = rand.nextInt(3);
      newPitch = java.lang.Math.abs(oldPitch) - randIntValue;
   }

   if (oldPitch > 0) {
      return newPitch;
   }
   else return -newPitch;
}


/**
 * Offset angle updates - cleared for brevity
 * @param oldOffset The offset angle from the previous timestamp
 * @return newOffset A double representing the new offset angle
 */
public static double updateOffset(double oldOffset) {
   double newOffset = 0.0;
   return newOffset;
}

/**
 * Alignement angle update - cleared for brevity
 * @param oldAlignment The Alignment angle from the previous timestamp
 * @return newAlignment A double representing the new alignment angle
```

```
 */
public static double updateAlignment(double oldAlignment) {
   double newAlignment = 0.0;
   return newAlignment;
}

/**
* Approach angle update - cleared for brevity
* @param oldApproach The Approach angle from the previous timestamp
* @return newApproach A double representing the new approach angle
*/
public static double updateApproach(double oldApproach) {
   double newApproach = 0.0;
   return newApproach;
}

/** Methods to convert from raw gauge/data values to "Fitness" values
 * Apply a function to the values from the gauges...
 * - effectively converts the various data metrics to a single scale
 *   that can be compared and reasoned with effectively
 *
 * each of the five variables has a different translation function...
 */

/** Altitude Conversion
 *
 * [0...10] with 10 at best elevation at timestamp and 0 at +/-(20%)
 * the timestamp is required to determine the ideal altitude
 *
 * @param rawAltitude actual altitude
 *                timestamp required to calculate the ideal altitude
 * @return goodAltitude quality of the altitude when compared to ideal approach
 */
public static double convertAltitude(double rawAltitude, int timeStamp) {
   double goodAltitude = 0.0;

   /* linear rate of descent from 10,000 feet */
   double initBestAlt = 10000;
   double nowBestAlt = initBestAlt - (initBestAlt*0.01*timeStamp);

   double upperAlt = 1.2 * nowBestAlt;
   double lowerAlt = 0.8 * nowBestAlt;

   //System.out.println("Acceptable Range [" + lowerAlt + ", " + upperAlt + "]");
   /* kept it this way in case we don't want to keep these sections linear
    * and it lets us tweak each bit individually...
```

```
    */
   if (rawAltitude > upperAlt) {
      /* If over 20% from best, it's bad */
      goodAltitude = 0;
   }
   if (rawAltitude < lowerAlt) {
      /* If under under 20% from best, it's bad */
      goodAltitude = 0;
   }
   if (rawAltitude < upperAlt && rawAltitude >= nowBestAlt) {
      double run = upperAlt-nowBestAlt;
      goodAltitude = (-10.0/run) * rawAltitude + 60.0;
   }
   if (rawAltitude > lowerAlt && rawAltitude < nowBestAlt) {
      double run = nowBestAlt-lowerAlt;
      goodAltitude = 10.0 + ((10.0/run) * rawAltitude - 50.0);
   }

   return goodAltitude;
}

/** Pitch/Roll (Wobble) Conversion
 *
 * [0...10] with 10 at level flight and 0 at 90 degrees from level
 * @param rawPitch
 * @return goodPitch A double of the objective Fitness of the Pitch
 */
public static double convertPitch(double rawPitch) {
   double goodPitch = 0.0;
   goodPitch = -0.111111111111 * java.lang.Math.abs(rawPitch) + 10;
   return goodPitch;
}

/**
 * Offset angle conversion - cleared for brevity
 * @param rawOffset
 * @return goodOffset
 */
public static double convertOffset(double rawOffset) {
   double goodOffset = 0.0;
   return goodOffset;
}

/**
 * Alignement angle conversion - cleared for brevity
 * @param rawAlignment
```

```
 * @return goodAlignment
 */
public static double convertAlignment(double rawAlignment) {
   double goodAlignment = 0.0;
   return goodAlignment;
}

/**
 * Approach angle conversion - cleared for brevity
 * @param rawApproach
 * @return goodApproach
 */
public static double convertApproach(double rawApproach) {
   double goodApproach = 0.0;
   return goodApproach;
}

/**
 * Opinion translators
 * Now that we have the objective fitness measures for each of the values we can
 * modify the behavior of the pilot by looking at how their opinions might be formed.
 *
 * these functions can create two values - one for hard opinion and one for fuzzy decision
   making
 * the sensitivity to individual gauges can be tweaked through the parameters for each gauge
 *
 * opinion formula :
 *  (objective value ^ aversion) / 10^(aversion - 1)
 *
 * keeping them separate for now in case we want to do something different with each
 */

/**
 * Altitude opinion function
 * @param objAltitude
 * @param aversion
 * @param fuzzy
 * @param opnThreshold
 * @return opnAltitude
 */
public static double opineAltitude(double objAltitude, double aversion, boolean fuzzy, double
   opnThreshold) {
   double opnAltitude = 0.0;
   opnAltitude = Math.pow(objAltitude, aversion) / Math.pow(10, aversion - 1.0);
   opnAltitude = opnAltitude / 10;
```

```
    if (fuzzy == false)
        if (opnAltitude > opnThreshold)
            opnAltitude = 1.0;
        else opnAltitude = 0.0;

    return opnAltitude;
}

/**
 * Pitch Opinion function
 * @param objPitch
 * @param aversion
 * @param fuzzy
 * @param opnThreshold
 * @return opnPitch
 */
public static double opinePitch(double objPitch, double aversion, boolean fuzzy, double
    opnThreshold) {
    double opnPitch = 0.0;
    opnPitch = Math.pow(objPitch, aversion) / Math.pow(10, aversion - 1.0);
    opnPitch = opnPitch / 10;

    if (fuzzy == false)
        if (opnPitch > opnThreshold)
            opnPitch = 1.0;
        else opnPitch = 0.0;

    return opnPitch;
}

/**
 * Offset angle opinion - cleared for brevity
 * @param objOffset
 * @return opnOffset
 */
public static double opineOffset(double objOffset) {
    double opnOffset = 0.0;
    return opnOffset;
}

/**
 * Alignement angle opinion - cleared for brevit
 * @param objAlignment
 * @return opnAlignment
 */
public static double opineAlignment(double objAlignment) {
```

```
    double opnAlignment = 0.0;
    return opnAlignment;
}

/**
 * Approach angle opinion - cleared for brevit
 * @param objApproach
 * @return opnApproach
 */
public static double opineApproach(double objApproach) {
    double opnApproach = 0.0;
    return opnApproach;
}

/**
 * Decsision Making section
 * We now have the opinion ratings for our simulated pilot - and now we need to create
 * a set of decisions based on the opinions held about the data available...
 */

/**
 * decideMajority
 * @param opnArray
 * @param fuzzy
 * @param numVar
 * @param threshold
 * @return
 */
public static boolean decideMajority(double[] opnArray, boolean fuzzy, int numVar, double
    threshold){
    boolean decision = false;
    double sum = 0.0;

    if (!fuzzy) {
        System.out.println("Majority Decision with Boolean Opinions");
        // pure majority, so no-go if evenly divided
        if (sumArray(opnArray) > (numVar /2))
            decision = true;
    }
    else {
        System.out.println("Majority Decision with Fuzzy Opinions");
        System.out.print("Threshold value = " + threshold + " : ");
        for (int i=0; i<opnArray.length; i++) {
            if (opnArray[i] > threshold)
                sum = sum + 1.0;
        }
```

```
      // pure majority, so no-go if evenly divided
      if (sum > (numVar /2))
         decision = true;
   }

   System.out.println("Number of GO votes = " + sum);

   return decision;
}

/**
 * decideMajority
 * @param opnArray
 * @param numVar
 * @param threshold
 * @return
 */
public static boolean decideMajority(double[] opnArray, int numVar, double threshold){
   boolean decision = false;
   double sum = 0.0;

   System.out.println("Majority Decision with Bayesian Opinions");
   System.out.print("Threshold value = " + threshold + " : ");
   for (int i=0; i<opnArray.length; i++) {
      if (opnArray[i] > threshold)
            sum = sum + 1.0;
      }
   // pure majority, so no-go if evenly divided
   if (sum > (numVar /2))
      decision = true;

   System.out.println("Number of GO votes = " + sum);

   return decision;
}

/**
 * decideDisjunction - if any one of the opinions are negative - NO-GO
 *  - applies to both boolean and fuzzy decisions
 * @param opnArray
 * @param weightArray
 * @param fuzzy
 * @param threshold
 * @return
 */
```

```
public static boolean decideDisjunction(double[] opnArray, double[] weightArray, boolean
   fuzzy, double threshold){
   boolean decision = false;
   double opnValue = 0.0;
   System.out.println("Opinion Threshold: " + threshold);
   if (fuzzy) {
      System.out.println("Disjunction decision with Fuzzy opinions");
      for (int i=0; i<opnArray.length; i++) {
         opnValue = opnArray[i] * weightArray[i];
         if (opnValue < threshold) {
               return decision;
         }
      }
   } else {
      System.out.println("Disjunction decision with Boolean opinions");
      for (int i=0; i<opnArray.length; i++) {
         if (opnArray[i] == 0.0) {
               return decision;
         }
      }
   }

   // everything points to an OK landing
   decision = true;
   return decision;
}
/**
 * decideDisjunction - if any one of the opinions are negative - NO-GO
 *  - applies to both boolean and fuzzy decisions
 * @param opnArray
 * @param threshold
 * @return
 */
public static boolean decideDisjunction(double[] opnArray, double threshold) {
   boolean decision = false;
   double opnValue = 0.0;
   System.out.println("Opinion Threshold: " + threshold);
   System.out.println("Disjunction decision with Bayesian opinions");
   System.out.println("Threshold for p(land) = " + threshold);
   for (int i=0; i<opnArray.length; i++) {
      opnValue = opnArray[i];
      if (opnArray[i] < threshold) {
               return decision;
      }
   }
   // everything points to an OK landing
```

```
    decision = true;
    return decision;
}

/**
 * DecideThreshold
 * @param opnArray
 * @param weightArray
 * @param fuzzy
 * @param threshold
 * @return
 */
public static boolean decideThreshold(double[] opnArray, double[] weightArray, boolean
    fuzzy, double threshold){
    boolean decision = false;
    double opnValue = sumArrays(opnArray, weightArray);

    if (fuzzy)
        System.out.println("Threshold decision with Fuzzy opinions");
    else System.out.println("Threshold decision with Boolean opinions");

    System.out.print("Threshold value = " + threshold + " : ");
    System.out.println("opnValue = " + opnValue);

    // if the sum of the opinions is over the threshold run with it
    if (opnValue > threshold)
        decision = true;
    // this catches opnValue <= threshold
    else decision =  false;

    return decision;
}

/**
 * DecideThreshold
 * @param opnArray
 * @param weightArray
 * @param threshold
 * @return
 */
public static boolean decideThreshold(double[] opnArray, double[] weightArray, double
    threshold){
    boolean decision = false;
    double opnValue = sumArrays(opnArray, weightArray);

    System.out.println("Threshold decision with Bayesian opinions");
```

```
    System.out.print("Threshold value = " + threshold + " : ");
    System.out.println("opnValue = " + opnValue);

    // if the sum of the opinions is over the threshold run with it
    if (opnValue > threshold)
        decision = true;
    // this catches opnValue <= threshold
    else decision =  false;

    return decision;
}

/**
 * decideBayesian
 * @param bayesArray
 * @param threshold
 * @return
 */
public static boolean decideBayesian(double[] bayesArray, double threshold, boolean
    updateBayes){
    boolean decision = false;
    double product = 0.0;

    if (updateBayes) {
        System.out.println("Bayesian decision - belief revision (temporal difference) model");
    } else System.out.println("Bayesian decision - momentary decision");

    product = productArray(bayesArray);
    System.out.print("Threshold = " + threshold + " : ");
    System.out.println("Product of all updated p(land) = " + product);

    if (product > threshold) {
        decision = true;
    }

    return decision;
}

/**
 * sumArrays - helper method that returns the sum of the weighted opinions
 * @param array1
 * @param array2
 * @return sum
 */
public static double sumArrays(double[] array1, double[] array2){
```

```
    double sum = 0.0;
    if (array1.length != array2.length) {
        System.out.println("hey - the arrays aren't of the same length - sumArrays");
        return 0.0;
    }
    for (int i=0; i<array1.length; i++) {
        sum += array1[i] * array2[i];
    }
    return sum;
}

/**
 * sumArray - helper method that returns the sum of the elements of the array
 * @param dblArray
 * @return sum
 */
public static double sumArray(double[] dblArray) {
    double sum = 0.0;
    for (int i=0; i<dblArray.length; i++) {
        sum += dblArray[i];
    }
    return sum;
}

/**
 * productArray - helper method that returns the product of the elements of the array
 * @param dblArray
 * @return product
 */
public static double productArray(double[] dblArray) {
    if (dblArray.length <= 0) {
        System.out.println("Hey - give me something to work with - productArray");
        return 0.0;
    }

    double product = dblArray[0];
    for (int i=1; i<dblArray.length; i++) {
        product = product * dblArray[i];
    }
    return product;
}

/**
 * Bayes - normalized update function
 *
 *                          p(opinion|land)*p(land)
```

```
     * p(land|opinion) = ---------------------------------------------------
     *                   p(opinion|land)*p(land) + p(opinion|~land)*p(~land)
     *
     * in words:
     *
     * the probability of landing given the pilot's opinion
     *  is equal to
     *    (the probability of the opinion given the choice to land) times (the probability of the
       choice to land)
     *   divided by
     *   [ (the probability of the opinion given the choice to land) times (the probability of the
       choice to land)
     *     plus
     *    (the probability of the opinion given the choice NOT to land) times (the probability of the
       choice NOT to land)
     *   ]
     *
     * @param land Prior probability of landing
     * @param opnLand Conditional probability of opinion given land - the subjective opinion
     * @param opnNotLand Conditional probability of opinion given not land
     * @return posterior probability od
     */
    public static double bayes(double land, double opnLand, double opnNotLand){
       double posterior = 0.0;
       double notLand = 1.0 - land;

       posterior = (opnLand*land) / ((opnLand*land) + (opnNotLand*notLand));

       System.out.print("p(land) = " + land + " : ");
       System.out.print("p(opinion|land) = " + opnLand + " : ");
       System.out.println("p(land|opinion) = " + posterior);

       return posterior;
    }

} /* END gnglogic */
```