Closing the Loop on Computational Cognitive Modeling:

Describing the Dynamics of Interactive Decision Making and Attention Allocation

Michael D. Byrne

Rice University

byrne@rice.edu

Alex Kirlik

University of Illinois at Urbana-Champaign

kirlik@uiuc.edu

Michael D. Fleetwood

Rice University

fleet@rice.edu

Draft chapter manuscript to appear in Foyle & Hooey (Eds), Title to be Determined, 2004.

INTRODUCTION

It has now been over 25 years since prominent human factors researchers such as Sheridan and Rasmussen called for a shift of focus in modeling human performance in technological systems (Sheridan, 1976, Rasmussen, 1976). Prior to this call, human manual control behavior, such as hand flying an aircraft or steering a ship occupied the majority of modelers' attention (see Jagacinski and Flach, 2002, for a recent overview and integration). But as researchers such as Sheridan and Rasmussen noted, many of not all of the tightest and fastest control loops in many human-machine systems are now available to be controlled by automation of one sort or another (e.g., autopilots). Additionally, a system operator's perceptual access to the controlled system is no longer direct but instead mediated by technological interfaces presenting processed and designed information rather than raw data. As such, the most pressing and important knowledge gap to be filled had changed. The central new questions had shifted to how a system operator monitors, manages, or otherwise interacts with highly automated systems, that is, "supervisory control" (Sheridan, 1976), and also how an operator uses the highly processed information locally available from a technological interface to diagnose and respond to system faults and environmental disturbances occurring in a remote location, that is, "behind" the interface, so to speak (Rasmussen, 1976).

Note the common nature of the supervisory control and fault diagnosis tasks spawned by increasing levels of automation. Both emerged due to automation and interface technology playing an increasingly *mediating* role, in standing between the operator and the controlled system and its environment. As such, an ever increasing need existed for the operator to go "beyond the information given" (Bruner, 1957) in the performance of these supervisory control and fault management tasks. As one of the fathers of the "cognitive revolution," Bruner pointed out that both the need and ability to effectively perform this “going beyond” is one of the hallmarks of a process that was once controversial in psychology, but is now taken as commonplace: *cognition*. Regardless of the exact terms in which Rasmussen and Sheridan originally framed their charge to the human performance modeling community, we can now safely say that their charge amounted to an appeal to extend modeling into the cognitive realm.

As a result, and due to this shift of emphasis, one can find many if not most authors, nearly all of them engineers, writing in the 1976 landmark volume “Monitoring Behavior and Supervisory Control (Sheridan and Johanssen, 1976) grappling with the issues of how to incorporate cognitive activities such as decision making, attention, memory, planning and problem solving into their quantitative human performance modeling approaches. And this trend, demonstrating a optimistic (some might say, “engineering”) outlook, that cognitive-level performance could largely be formulated in quantitative terms, held sway in the “man-machine systems” tradition throughout the late 1970s and early 80s, as revealed by the contents of the first volume in the well known “Advances in Man-Machine Systems Research” series, edited by the prominent human-machine systems engineer William Rouse (1984).

But history has demonstrated that modeling cognition in complex, operational task environments would not quickly yield to quantitative techniques such as Bayesian decision or estimation theories, optimal control theory, or even early AI-inspired models such as expert systems and “intelligent” planners or problem solvers (Kirlik & Bisantz, 1998). As such, to the extent that modeling has still proven useful and influential in human factors of the last 20 years or so, it has been aimed at a highly broad, abstract, and qualitative level, perhaps best illustrated by Rasmussen’s (1983) “Skills, Rules, and Knowledge” framework and Klein’s (1989) “Recognition-Primed Decision” model, or else aimed at a relatively micro-level, either with respect to one particular psychological "module" or one particular type of task. As noted by Kirlik (2003), one likely explanation for the slow progress in developing *both* more general *and* more formal cognitive performance models has been the difficulty in discovering or identifying sufficiently context-free abstractions of complex operational contexts; these abstractions being the hallmark of engineering modeling techniques.

But times are changing. As Byrne and Gray (2003) have observed, perhaps the time is ripe for formal, that is, computational or quantitative, approaches to once again provide a useful role in the engineering of complex human-machine systems. The present chapter, and indeed this volume itself, hopefully provides evidence for at least some merit to this claim. For during this past 20 years or so, another community of scientists, working within the traditions of cognitive science and human-computer interaction, have been developing, iteratively refining and validating, modeling techniques that we believe to hold promise in eventually becoming useful engineering tools for the analysis, design, and evaluation of human-machine systems.

Broadly speaking, these "cognitive architectures" (see Byrne, 2003, for an introduction and overview) represent an attempt to provide general frameworks for cognitive modeling in which the findings from scores of experimental studies and the latest psychological theory is implicitly embedded in the design of the architecture itself. As such, they can be viewed as computational artifacts that effectively embody much state-of-the-art theory and empirical knowledge in cognitive psychology, thus allowing the informed users of these artifacts to benefit by this knowledge to construct plausible cognitive models of particular cases of human-machine interaction. This "artifact as knowledge" view highlights another important feature, or way of looking at, these modeling techniques; i.e., the cognitive modeling architecture as the repository of cumulative scientific knowledge. To illustrate, neither of the current authors would call himself an expert on the retrieval of information from long-term memory. Yet, the models described below allow us to leverage exactly this expertise as it is in a very real sense “cached” directly into the design of the cognitive architecture we have chosen to use to model the taxiing and approach and landing scenarios.

Yet, as will be made clear, we believe that these architectures still have some way to go before they can be used as efficiently and usefully as, say, Fitts’s Law (Fitts, 1954) for the analysis, design, and evaluation of human-machine system performance. This point needs to be articulated further to illustrate the rationale underlying the presentation of much of the research to follow. In particular, we would like to pause to first explain why a considerable amount of time and effort had to be devoted to “applying” the particular modeling architecture we have chosen to work with to the NASA scenarios, and importantly, why we do *not* think that this time and effort bodes poorly for the future prospects of computational cognitive modeling becoming a useful engineering technique in human factors.

Model Versus Application

While an ideal modeling tool for the analysis and design of human-machine systems would require no special expertise of the part of the modeler and would not require the modeler to engage in a protracted and potentially effortful exercise in applying the model, we know of no engineering discipline in which models are used in this way. Were this the case, there would be no need to hire structural engineers to design bridges, no need to hire electrical engineers to design circuits, and no need to hire aeronautical engineers to design airplane wings, even though every one of these professionals is equipped with a rich toolbox of models. Especially as human-machine systems increase in technological complexity, we believe it is simply naive that the profession of cognitive engineering or human factors engineering will proceed any differently.

Next, we would like to point out that even in the simplest cases in human factors, the apparent simplicity of a model can belie the complexity of the tasks involved in its application. Take the aforementioned Fitts's Law, which is simple a log-linear mathematical model that is often said to be able to "predict" human discrete movement times for various tasks. But Fitts's Law, in and of itself, predicts nothing of the sort. Making such predictions requires performing work to parameterize the equation according to the details of the encounter between a particular human and a particular environment. The values of both the slope and intercept terms of Fitts's Law hardly have the status of physical constants, as they vary both across the degrees of freedom of the motor systems involved in the movement in question, and also across individuals. And we know of no truly reliable published tables for these values, thus requiring that to make any point "prediction" using Fitts's Law one will first have to study the performer of interest in order to estimate these human-centered parameters of the equation from behavioral data. As such, we do not find it defensible to suggest that the need for cognitive modelers use data associated with the prior behavior of performers in order to parameterize models in order to make analogous point predictions somehow reveals a deficiency of cognitive modeling methods.

Finally, consider the environmentally-centered variables in Fitts's Law, namely, target distance (or movement amplitude) and target width (or the need for accuracy). Estimates of the values of these variables must come from a detailed analysis of the task environment, and when those movements are of greater than one dimension, or if the target is irregularly shaped, making such measurements becomes notoriously tricky (MacKenzie, 2003). What is notable is that this task can even be non-trivial in the case of Fitts's Law, where the functionally-variables of the task environment (distance and target width) are known. This task is vastly more complicated in the case of applying a cognitive architecture of any kind since, in the vast majority of cases, they do not speak to the issue of specifying the environmental variables that play a functional role in cognition and behavior. This is not to say that the design of the architecture itself may not reflect a deep concern with environmental structure, and thus how the mechanisms comprising the architecture should be designed in an environmentally-adaptive fashion (see Anderson, 1990).

It is the case, however, that the relative lack of resources contained in these architectures to support environmental modeling is, admittedly, an arguable weakness of many such approaches (e.g., in comparison to a model such as Fitts's Law), and it owes to the origin of most cognitive architectures in the view that specifying functionally relevant environmental variables was outside the scope or purview of cognitive modeling *per se*. This issue is especially crucial for human factors applications, due to the longstanding recognition that:

> Human behavior, either cognitive or psychomotor, is too diverse to model unless it is sufficiently constrained by the situation or environment; however, when these environmental constraints exist, to model behavior adequately, one must include a a model for that environment. (Baron, 1984, p.6)

Contrasting views on what should be the appropriate scope of cognitive modeling do exist, of course, but a discussion of this issue is beyond the scope of the present chapter (e.g., see Hutchins, 1995; Hammond & Stewart, 2001; Kirlik, Miller & Jagacinski, 1993; Tolman & Brunswik, 1935). Suffice it to say, modeling a performer in a realistically complex task environment using a cognitive architecture not only requires a knowledge engineering component, but it can also require a parallel effort to discover the functionally relevant environmental variables of interest. Additionally, it may even require, as Baron suggests, actually including these variables in a model in its own right, or else "hooking up" the computational cognitive model to an existing simulation of the task environment, tasks that we discuss in some detail on in the following. In fact, we take one contribution of the research presented in this chapter to be demonstrating how, in at least the two scenarios of interest, the task of supplementing an architecturally-inspired cognitive model with models or simulations of either the controlled system (aircraft), and the task environment can at least be approached.

In a larger context, however, this was but one of the many challenges we faced in bringing what to this point had largely been a cognitive architecture applied and validated primarily in laboratory tasks to bear on realistically complex, operational scenarios (for another example, see Salvucci's example in the domain of automobile driving; Salvucci, 2001). In the following descriptions of the two modeling efforts we performed in taxi navigation and approach and landing with and without Synthetic Vision System (SVS) displays, we share what we consider to be not only our successes, but also our failures, compromises, and shortfalls, as we believe all of this information is potentially relevant to moving cognitive modeling research forward.

ACT-R MODELING: A GENERAL OVERVIEW

ACT-R (Anderson, et al., in press; see also Anderson & Lebiere, 1998) is a computational architecture designed to support modeling of human cognition and performance at a detailed temporal grain size. Figure 1 depicts the general system architecture. ACT-R allows for modeling of the human-in-the-loop as the output of the system is a time-stamped stream of behaviors at a very low level, such as individual shifts of visual attention, keystrokes, and primitive mental operations such as retrieval of a simple fact. In order to produce this, ACT-R must be provided two things: knowledge and a world or environment (usually simulated) in which to operate. The environment must dynamically respond to ACT-R's outputs and thus must also often be simulated at a high degree of fidelity. The knowledge that must be provided to

ACT-R to complete a model of a person in an environment is essentially of two types: declarative and procedural. Declarative knowledge, such as “George Washington was the first president of the United States,” or “‘IAH’ stands for Bush Intercontinental airport in Houston,” is represented in symbolic structures known as *chunks*.

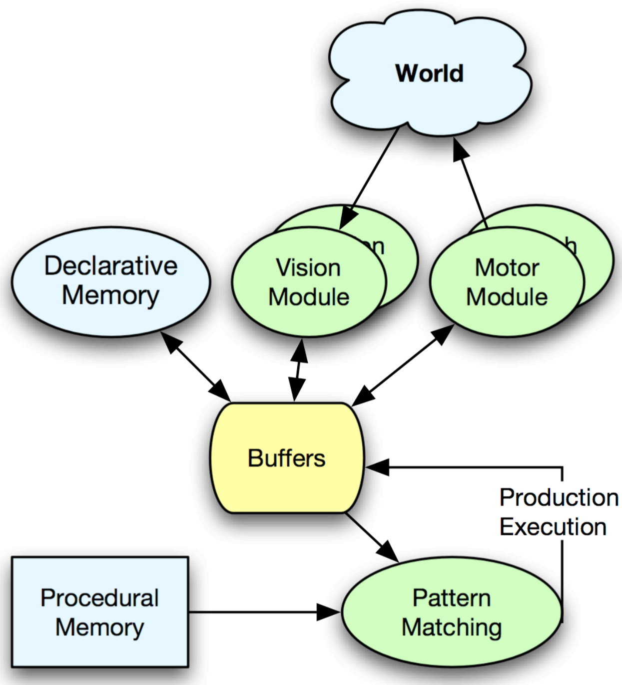

Figure 1. The ACT-R cognitive architecture

Procedural, sometimes referred to as “how-to,” knowledge, such as the knowledge of how to lower the landing gear in a 747, is stored in symbolic structures known as *production rules* or simply *productions*. These consist of IF-THEN pairs; IF a certain set of conditions hold, THEN perform one or more actions. In addition, both chunks and productions contain quantitative information that represents the statistical history of that particular piece of knowledge. For example, each chunk has associated with it a quantity called *activation* that is based on the frequency and recency of access to that particular chunk, as well as its relationship to the current context. Because the actual statistics are often not known, in many cases these values are left at system defaults or estimated by the modeler, though in principle ACT-R can learn them as well.

The basic operation of ACT-R is as follows. The state of the system is represented in a set of buffers. The IF sides of all productions are matched against the contents of those buffers. If multiple productions match, a procedure called *conflict resolution* is used to determine which production is allowed to fire, or apply its THEN actions. This generally changes the state of at least one buffer, and then this cycle is repeated every 50 ms of simulated time. In addition, a buffer can change without a production explicitly changing it. For example, there is a buffer that represents the visual object currently in the focus of visual attention. If that object changes or disappears, this buffer will change as a result. That is, the various perceptual and motor processes (and declarative memory as well) act in parallel with each other and with the central cognitive production cycle. These processes are modeled at varying levels of fidelity. For example, ACT-R does not contain any advanced machine vision component that would allow it to recognize objects from analog light input. Rather, ACT-R needs to be given an explicit description of the visual object to which it is attending.

A COMPUTATIONAL COGNITIVE MODEL OF TAXI NAVIGATION

One of the first decisions that had to be made in modeling the NASA taxi navigation simulation and scenarios (Hooey & Foyle, this volume) was a decision about scope. To limit the scope of the project, we chose to model only the captain in ACT-R, and treated both the ground controller and the first officer (FO) as items in the environment. We felt this decision balanced tractability and relevance, since the captain made the final decisions and the captain also controlled the aircraft.

A second important aspect of scoping model coverage was to select the focal psychological activities to model. Our research team was one of many teams also creating cognitive models of the same T-NASA2 data (as described elsewhere in this volume). In this light, we considered both the strengths and weaknesses of our ACT-R approach with the alternative approaches taken by other research teams, with the goal of providing a unique contribution to the overall research effort. For example, we ruled out focusing on multi-tasking and situation awareness (SA) issues (losing track of one's location on the airport surface), as these were the focal points for other approaches. All things considered, including our own previous experience in human performance modeling (e.g., Kirlik, 1998; Kirlik, Miller & Jagacinski, 1993), we decided to focus on the interactive, dynamic decision making aspects of the task in its closed-loop context. As a result, we focused on those contributions to error that may result from the interaction of the structure of a task environment and the need to make often-rapid decisions on the basis of imperfect information, resulting from decay of clearance information from memory, low visibility, and sluggish aircraft dynamics. Our focus on decision making, which assumed pilots had accurate knowledge of their current location, was complemented by another modeling team's focus on situation awareness (SA) errors associated with losing track of one's location on the airport surface (McCarley et al., this volume).

The Model's Environment

We created an ACT-R model of one human pilot, but this pilot model still had to be situated in an accurate environment. Thus, three external entities were modeled to describe the environment: the simulated aircraft controlled by the pilot model; the simulated visual information available to the pilot model; and the simulated runway and taxiway environment through which the aircraft traveled. Each of these three environmental entities was modeled computationally and then integrated with the cognitive components of the pilot model to create an overall representation of the interactive human-aircraft-environment system.

Code for the vehicle dynamics that was used to drive the actual NASA flight simulator in which behavioral data was collected was unfortunately unavailable. We therefore had to create a simplified vehicle model with which the pilot model could interact. Although we were not interested in control issues per se, the dynamics of the aircraft played an important role in determining decision time horizons, a key factor in the cognitive representation of the pilot's activities. Given vehicle size, mass, and dynamics, however, we still did require a somewhat reasonable approximation to the actual aircraft dynamics used in the experiments in order to be able to get a handle on timing issues. The aircraft model we constructed assumed that the pilot controlled the vehicle in three ways: applying engine power, braking, and steering. For the purposes of modeling an aircraft during taxiing, these three forms of control are sufficient. Based on Cheng, Sharma, and Foyle's (2001) analysis of the NASA simulated aircraft dynamics, we proceeded with a model in which it was reasonable to assume that throttle and braking inputs generated applied forces that were linearly related with aircraft speed.

Steering, however, was another matter. After consideration of the functional role that steering inputs played in the T-NASA2 scenario, we decided that we could finesse the problem of steering dynamics by assuming that the manual control aspects of the steering problem did not play a significant role in the navigation errors that were observed. That is, we assumed that making an appropriate turn was purely a decision-making problem, and that no turn errors resulted from correct turn decisions that were erroneously executed. Note that this assumption does not completely decouple the manual and cognitive aspects of the modeling, however. It was still the case that manual control of the acceleration and braking aspects of the model did play a role in determining the aircraft's position relative to an impending turn, and importantly, placed a hard constraint on the aircraft's maximum speed of approach to each turn.

The maximum aircraft speeds for the various types of turns required in the NASA simulation were calculated under the constraint that lateral acceleration be limited to 0.25 g for passenger comfort (Cheng et al., 2001) and also the field data reported in Cassell, Smith and Hicok (1999). These maximum speeds partially determined the time available to make a turn decision, and as will be seen, as this time is reduced there was a greater probability of an incorrect decision. Our simplification regarding steering merely boiled down to the fact that once the model had made its decision about which turn to take, that turn was executed without error.

To implement this aspect of the model, we decided to model the ORD airport taxiway as a set of interconnected "rails" upon which travel of the simulated aircraft was constrained. Taxiway decision making in this scheme, then, boiled down to the selection of the appropriate rail to take at each taxiway intersection. In this manner, we did not have to model the dynamics of the aircraft while turning: we simply moved the aircraft along each turn rail at the specified, turn-radius-specific speed.

The model used to represent the visual information available to our ACT-R pilot model was obtained from the actual NASA flight simulator in the form of a polygon database. This database consisted of location-coded objects (e.g., taxiways, signage) present on the ORD surface, or at least those objects presented to flight crews during NASA experimentation. Distant objects became "visible" to the pilot model at similar distances to which these same objects became visible to human pilots in T-NASA2 experimentation.

Task Analysis and Knowledge Engineering

The task-specific information required to construct the model was obtained by studying various task analyses of taxiing (e.g., Cassell, Smith & Hicok, 1999) and through extensive consultation with two subject matter experts (SMEs) who were experienced airline pilots. We first discovered that, in many cases, pilots have multiple tasks in which to engage while taxiing. Based on this finding, our ACT-R model only concerned itself with navigation decision making when such a decision was pending. In the interim, the model iterated through four tasks deemed central to the safety of the aircraft.

These four tasks included monitoring the visual scene for incursions, particularly objects like ground vehicles which are difficult to detect in poor visibility, maintaining the speed of the aircraft, since the dynamics of a commercial jetliner require relatively frequent adjustments of throttle and/or brake to maintain a constant speed, listening for hold instructions from the ground-based controller, and maintaining an updated representation of the current position of the aircraft on the taxi surface and the location of the destination. While these tasks often have little direct impact on navigation, they do take time to execute, and time is the key limited resource in making navigation decisions in our integrated pilot-aircraft-environment system model.

With respect to navigation decisions, we found that decision-making is highly local. That is, the planning horizon is very short; flight crews are quite busy in the time after landing and thus, in situations like ORD in poor visibility, report they do not have the time to "plan ahead" and consider turns or intersections other than the immediately pending one. Second, the decision process tends to be hierarchical: pilots first decide if the next intersection requires a turn, and if it does, then decide which turn to make. We found that in the error corpus available to us (NASA, 2001), errors in the first decision (whether to turn or not) were rare (which was also consistent with our SME reports), and so we concentrated our efforts on understanding how pilots made the second decision.

The first issue to be addressed was: What kinds of knowledge and strategies are actually brought to bear by real pilots in the kinds of conditions experienced by the pilots in the NASA study? Largely though interviews with SMEs, we discovered a number of key strategies employed by pilots, and also discovered that some of these strategies would *not* have been available to our model (which we thus excluded from consideration). At the end of both our task analyses and SME interviews, we had identified five primary decision strategies available for making turn decisions:

1. Remember the correct clearance: While fast, this strategy is increasingly inaccurate as time lapses between obtaining the list of turns described in the clearance and the time at which turn execution is actually required.

2. Make turns toward the gate: While somewhat slower than the first strategy, this strategy has a reasonable level of accuracy at many airports.

3. Turn in the direction that reduces the larger of the X or Y (cockpit-oriented) distance between the aircraft and the gate. We deemed this strategy to be moderately fast, like strategy 2, but with a potentially higher accuracy than strategy 2, since more information is taken into account.

4. Derive from map/spatial knowledge. This is the slowest strategy available, with high accuracy possible only from a highly experienced (at a given airport) flight crew.

5. Guess randomly. This is a very fast strategy, although it is unlikely to be very accurate, especially at multi-turn intersections. However, we did include it as a possible heuristic in the

model for two reasons: a) it may be the only strategy available given the decision time available in some cases; and b) it provides insights into chance performance levels.

The next modeling issue to be dealt with was how to choose between strategies when faced with a time constrained decision horizon. This type of meta-decision is well modeled by the conflict resolution mechanism ACT-R uses to arbitrate between multiple productions matching the current situation. The accuracy of strategies 1 (recall the clearance) and 4 (derive from map knowledge) is primarily a function of the accuracy of the primitive cognitive operations required of these tasks, moderated by factors such as ACT-R's memory decay and effectively constrained working memory. However, the accuracy of strategies 2, 3, and 5 is less cognitively constrained and instead critically dependent on the geometry of actual clearances and taxiways. As such, we thus employed an SME as a participant in a study to provide data for an analysis of the heuristic decision strategies 2 and 3 (the accuracy of strategy 5, random guessing, was determined by the taxiway geometry itself). We would have perhaps never thought of performing this study had the ACT-R model not required us to provide it with high level (i.e., airport neutral) strategies pilots might use in deciding what turns to make during taxi operations, along with their associated costs (times required) and benefits (accuracy).

Identifying Taxi Decision Heuristics

To obtain this information, which was required to inform modeling, we provided our SME (a working B-767 pilot for a major U.S. carrier) with Jeppesen charts for most major U.S., airports, and then asked him to select charts for those airports for which he had significant experience of typical taxi routes and clearances. He selected 9 airports (DFW, LAX, SFO, ATL, JFK, DEN, SEA, MIA, ORD). The SME was asked to draw, using a highlighter on the charts themselves, the likely or expected taxi routes at each airport from touchdown to his company's gate area. A total of 284 routes were generated in this way.

Our goal at this point was to identify whether any of the heuristic strategies identified during task analysis and knowledge engineering would be likely to yield acceptable levels of decision accuracy. We obtained an estimate of the accuracy of heuristic strategies 2 (turn toward the company's gates), and 3 (turn in the direction that minimizes the largest of the X or Y distance between the current location and the gates) by comparing the predictions these heuristics would make with the data provided by the SME for the 9 airports studied. We recognize that these accuracy estimates may be specific to the (major) carrier for whom the SME flew, since other carriers' gates may be located in areas at these 9 airports such that their pilots are provided more or less complex, or geometrically intuitive, clearances than those providing the basis of our SME's experience. However, we do believe that this study resulted in enlightening results regarding the surprisingly high level of accuracy of simple, "fast and frugal" decision heuristics (Gigerenzer & Goldstein, 1996) in this complex, operational environment.

Figure 2 presents the results of an analysis of the effectiveness of these two heuristic strategies. Note that the XY heuristic is quite good across the board, and the even simpler "toward terminal" heuristic is reasonably accurate at many major U.S. airports. As such, we created the turn decision making components of the pilot model to make decisions according to the set of 5 strategies described previously, including the two surprisingly frugal and robust "Toward Terminal" and "XY" heuristics portrayed in Figure 2. One can think of these 5 strategies as being hierarchically organized in terms of their costs (time requirements) and benefits (accuracies). The decision components of the cognitive model worked by choosing the strategy that achieved the highest accuracy given the decision time available.

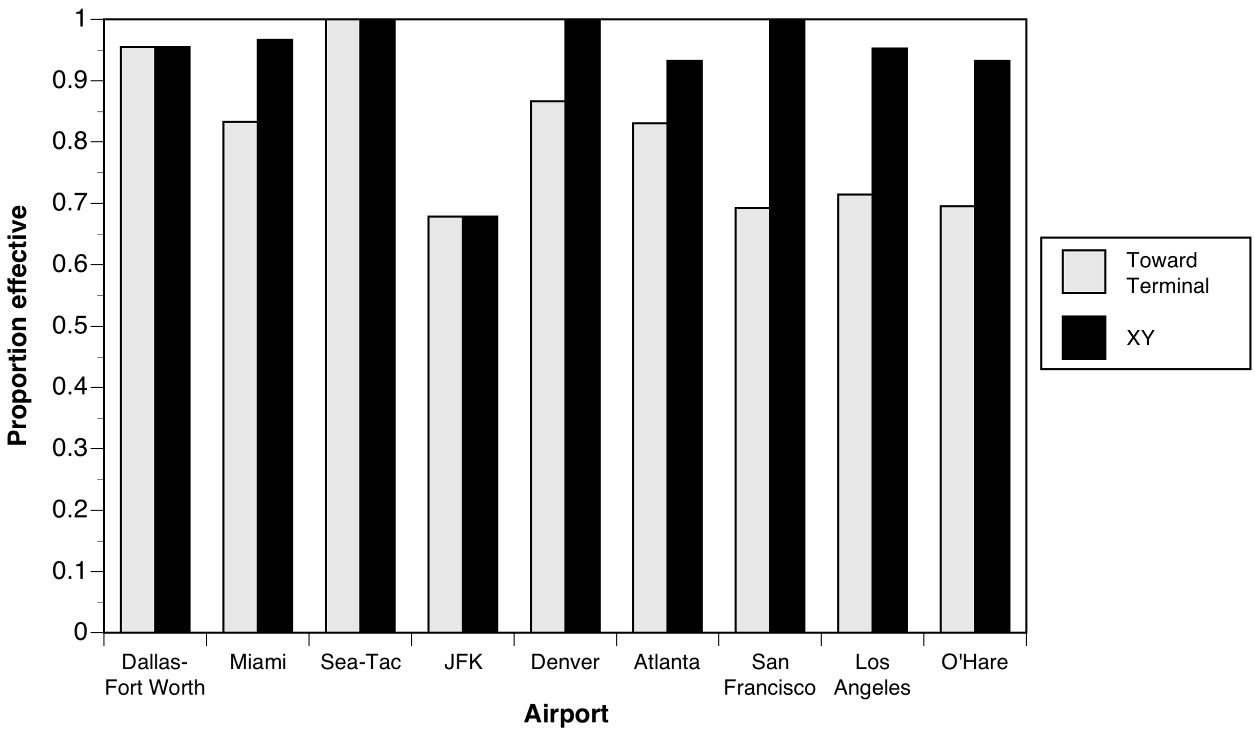

Figure 2. Accuracy of “Toward Terminal” and “Minimize Greater of XY Distance” heuristics

Detailed Description of Dynamic Decision Modeling

From a time-horizon (cost) perspective, the selection of decision strategies was informed by a procedure for estimating the time remaining before a decision had to be made. Time remaining was based on the aircraft’s distance to an intersection and the amount of slowing necessary to make whatever turns were available, which was thus dependent on aircraft dynamics. Recall that we had an algorithm available to calculate the maximum speed with which a turn of a given type could be negotiated. Thus, the computation of time remaining assumed a worst case scenario for each specific intersection. That is, the time horizon for decision making was determined by the intersection distance combined with knowledge of aircraft dynamics, used to determine whether breaking could slow the aircraft sufficiently to negotiate an intersection’s sharpest turn.

Each turn-related decision strategy was one production rule, which was allowed to enter conflict resolution only if the average time it would take the model to execute the procedure was less than 0.5 seconds less than the decision horizon. This somewhat conservative approach was used to compensate for the fact that both the time estimation and strategy execution times were noisy. Those productions meeting this criteria competed in a modified version of ACT-R's conflict resolution procedure. In the default ACT-R procedure, the utility of each production is estimated by the quantity PG-C, where P is the probability of success if that production is selected, G is a time constant (20 seconds is the default), and C is the time taken until an outcome is reached if that production fires. Because time cost was irrelevant in this application as long as the cost was less than the time remaining, this term was removed, though there was a 1 sec penalty applied to productions whose time cost was within 0.5 seconds of the remaining time, again a conservative move to ensure that a decision strategy likely to be completed will be selected (one of our SMEs indicated a conservative bias in this direction). The utility of each production is also assumed in ACT-R to be a noisy quantity, so the system was not always guaranteed to select the strategy with the highest utility as computed by the PG-C measure. (Amount of noise in this computation is a free parameter in ACT-R and a value of 1 was used as the *s* parameter in the logistic noise distribution. This yields a standard deviation of about 1.8, which was not varied to fit the data.) Thus, there were two sources of noise in this situation: estimation of time remaining, and the utilities of the strategies themselves.

In the pilot model, the P for each production was estimated according to the actual probability of success of each of the decision strategies. Thus, P for the production initiating the "turn toward the gate" production was 80.7% since that was the success rate for that strategy as determined by the SME study. P values for the other two decision heuristics (3 and 5 above) were calculated in an analogous fashion, and P values for strategies 1 (recall the actual clearance) and 4 (derive from the map) were determined by the boundedly-rational cognitive mechanisms inherent in the ACT-R cognitive architecture. With the entire model in place, we then ran a Monte Carlo simulation (300 repetitions at each of 50 time horizons) to determine the probability of selection for each strategy as a function of decision time available. These simulation results are presented in Figure 3.

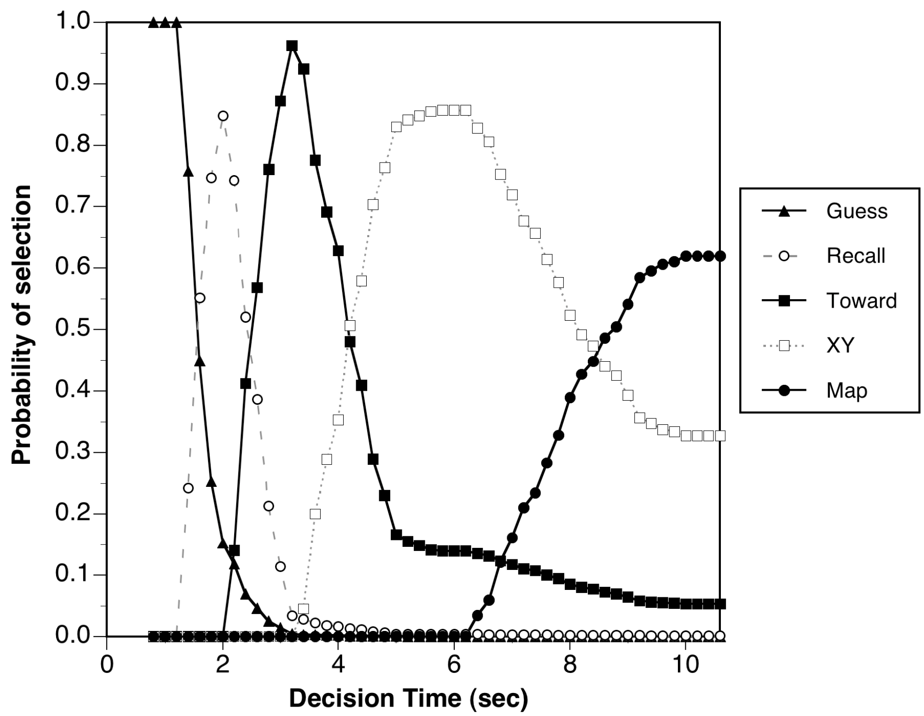

Figure 3. Selection probability for each decision strategy by decision time horizon

As is clear from Figure 3, as the decision horizon decreases, so does the likelihood that the pilot model will select a less accurate strategy. In fact, in the time window from about 2.5 to about 8 seconds, the environmentally-derived heuristics dominate alternative strategies, and based on our viewing of the NASA-supplied scenario videos (NASA, 2002), this time horizon was not unrepresentative of many actual decision horizons observed. However, this can be viewed as adaptive since a fast and frugal strategy that can run to completion can frequently outperform an analytically superior decision strategy that must be truncated due to time constraints (Gigerenzer & Goldstein, 1996).

Empirical Adequacy

Appropriate techniques for the verification and validation of human performance models based on computational cognitive modeling is an issue of great current interest (see, e.g., Leiden, Laughery & Corker, 2001). In the following, we present two sources of empirical evidence in support of our dynamic, integrated, computational model of this pilot-aircraft-visual scene-taxiway system. The first source of support is a global analysis of the frequency of taxi navigation errors as a function of intersection type. The second is a more finely grained analysis at an error-by-error level.

Global Evidence for Decision Heuristic Reliance

Nine different taxiway routes were used in the T-NASA2 baseline scenarios, covering a total of 97 separate intersection crossings. Since each route was run 6 times, a total of 582 intersection crossings occurred in the baseline trials. In only 12 instances were crews observed to make significant departures from the cleared route, resulting in an error rate (per intersection, rather than per trial) of approximately 2% (Goodman, 2001).

As Goodman (2001) reported, of the 582 intersection crossed, the clearance indicated that crews should proceed in a direction toward the destination gate in 534 cases (91.8%), while the clearance directed crews in directions away from the gate in only 48 cases (or 8.2%). Upon examining this information with respect to the predictions of both the “Toward Terminal” and “XY” heuristics embodied in our model, we discovered that at *every* one of the 97 intersection crossings in T-NASA2 scenarios at which the cleared route conflicted with *both* these two heuristics, at least one taxi error was made. These accounted for 7 of the 12 taxi errors observed.

In addition, and as will be discussed in the following section, 4 of the 12 taxi errors were attributed not to decision making, but rather to a loss of situation awareness or SA (i.e., losing track of one's position on the airport surface, see Goodman, 2001 and Hooey and Foyle, 2001), a cognitive phenomenon beyond the scope of the present modeling. Our modeling approach assumed that the primary source of errors was time-stressed decision making combined with what might be called "counter-intuitive" intersection and clearance pairs; i.e., those at which both the "Toward Terminal" and "XY" heuristics failed due to either atypical geometry or clearances.

Local Evidence of Decision Heuristic Reliance

The Goodman (2001) report provided a detailed analysis of each of the 12 taxi errors observed in the baseline conditions of T-NASA2 experimentation. In the Byrne & Kirlik (in press), we analyzed each error individually. In the present context, we will merely summarize these results by saying that 4 of the 12 errors were unambiguously classified by Goodman as owing to a loss of situation awareness (e.g., crews indicating their uncertainty about their location on the airport surface). Of the 8 remaining errors, which we consider to be decision making errors, every one of the 8 involved either an incorrect or premature turn toward the destination gate.

Summary

The crux of the interpretation of taxi errors in T-NASA2 is that pilots had multiple methods for handling individual turn decisions, and used the most accurate strategy possible given the time available (cf. Payne and Bettman, 2001). When time was short, as a function of poor visibility, workload, and aircraft dynamics, the model assumed that the pilot tended to rely on computationally cheaper, but less specific information gained from experience with the wider class of situations of which the current decision is an instance. In the case of the T-NASA2 scenario, this more general information pertained to the typical taxi routes and clearances that would be expected from touchdown to gate at major U.S. airports. This interpretation is also consistent with the fact that the suite of display aids used in the high-technology conditions of T-NASA2 experimentation, by providing improved information to support local decision making, effectively eliminated taxi errors. We believe that the ACT-R model we constructed not only received a reasonable degree of validation in terms of agreement with the empirical data, but perhaps even more importantly, prompted us to ask questions (e.g., about pilots' decision heuristics) that would likely would have never asked had we not attempted to accomplish the task of constructing a running computational cognitive model of this behavioral situation.

A COMPUTATIONAL COGNITIVE MODEL OF THE IMPACT OF SVS DISPLAYS

Hooey and Foyle (this volume) provide a detailed account of the experiments performed by NASA to examine the effects of introducing Synthetic Vision Display (SVS) technology into a cockpit during an simulated approach and landing scenario at Santa Barbara airport. Due to the relatively low number of participants (3) and the limited amount of overt behavioral data available due to flying a largely automated approach, we decided early on to focus our modeling efforts on trying to assess the influence of the SVS on pilot's visual attention allocation, for which much data was available through eye tracking. As such, our first step in this modeling exercise was to conduct a fairly encompassing analysis of visual attention allocation, based largely on an examination of fixation frequencies at the various areas-of-interest (AOIs) within the cockpit (including the out-the-window or OTW scene). A detailed analysis of these data is presented in Byrne and Kirlik (2004). The manner in which these analyses guided model construction and validation will be described in following sections.

Additionally, very early on in this exercise, and strongly influenced by our experience in modeling the taxi scenarios, we realized the importance of having a fairly veridical simulation of both the cockpit, aircraft, and task environment that we could couple with our ACT-R model to enable truly closed-loop simulation. To accomplish this, we needed a dynamic and accurate model of the pilot's environment with which ACT-R could interact. For the aircraft/flight simulation, we selected X-Plane for this purpose (note that X-Plane has been certified by the FAA for training, see http://www.x-plane.com/FTD.html) and linked ACT-R to X-Plane via a UDP network interface that we constructed 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 visual scene. This is a problem since the ACT-R model needs something to "see." Most of this problem was solved by mocking up the primary displays (navigation or NAV, primary flight display or PFD, mode control panel or MCP, etc.) in the language of ACT-R so that the cognitive model could directly "view" those pieces of the display. These displays were updated based on data sent from X-Plane over the network connection.
In addition, we had to supply X-Plane with the aircraft specifications (a 757) and the appropriate approach/navigation and FMC programming (e.g., approach waypoints) 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 4 depicts the overall interactive system.

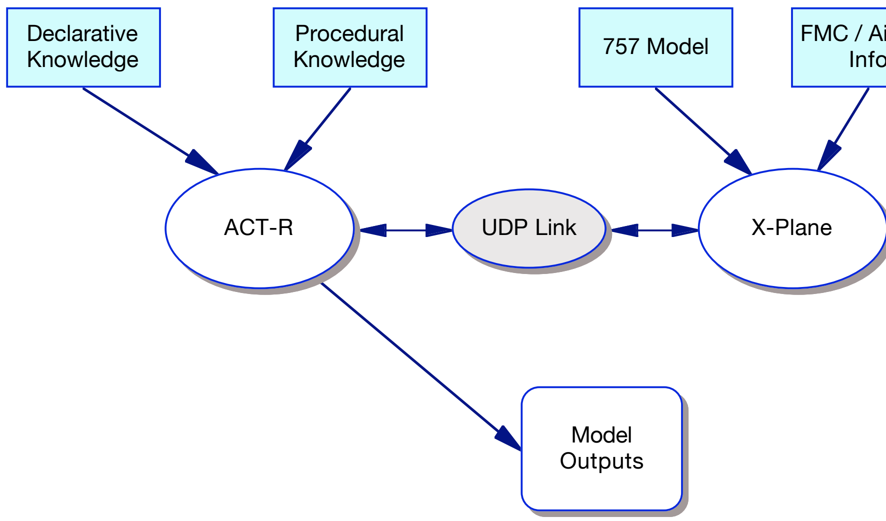

Figure 4. A Model of the closed-loop, interactive system

Modeling and Task Analysis

After detailed eye movement data analysis and the creation of an interactive, closed-loop simulation capability, our first order of business was to try to understand the task at a detailed level. We relied on three primary sources of information: the task analysis information collected and supplied by NASA Ames (Keller, Leiden, & Small, 2003); related work in the human factors of aviation; and conversations with our subject matter expert (SME). We synthesized these into the ACT-R formalism. An example of some of the resulting control structure appears in Figure 5.

The first insight from the knowledge engineering process is that the bulk of the task, particularly for the first two phases of flight, consists primarily of monitoring the state of the aircraft and maintaining an up-to-date representation of that state. 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 the bulk of the experimental trials, there is little activity in the form of over control actions while in fact the pilots are in fact highly engaged. 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.

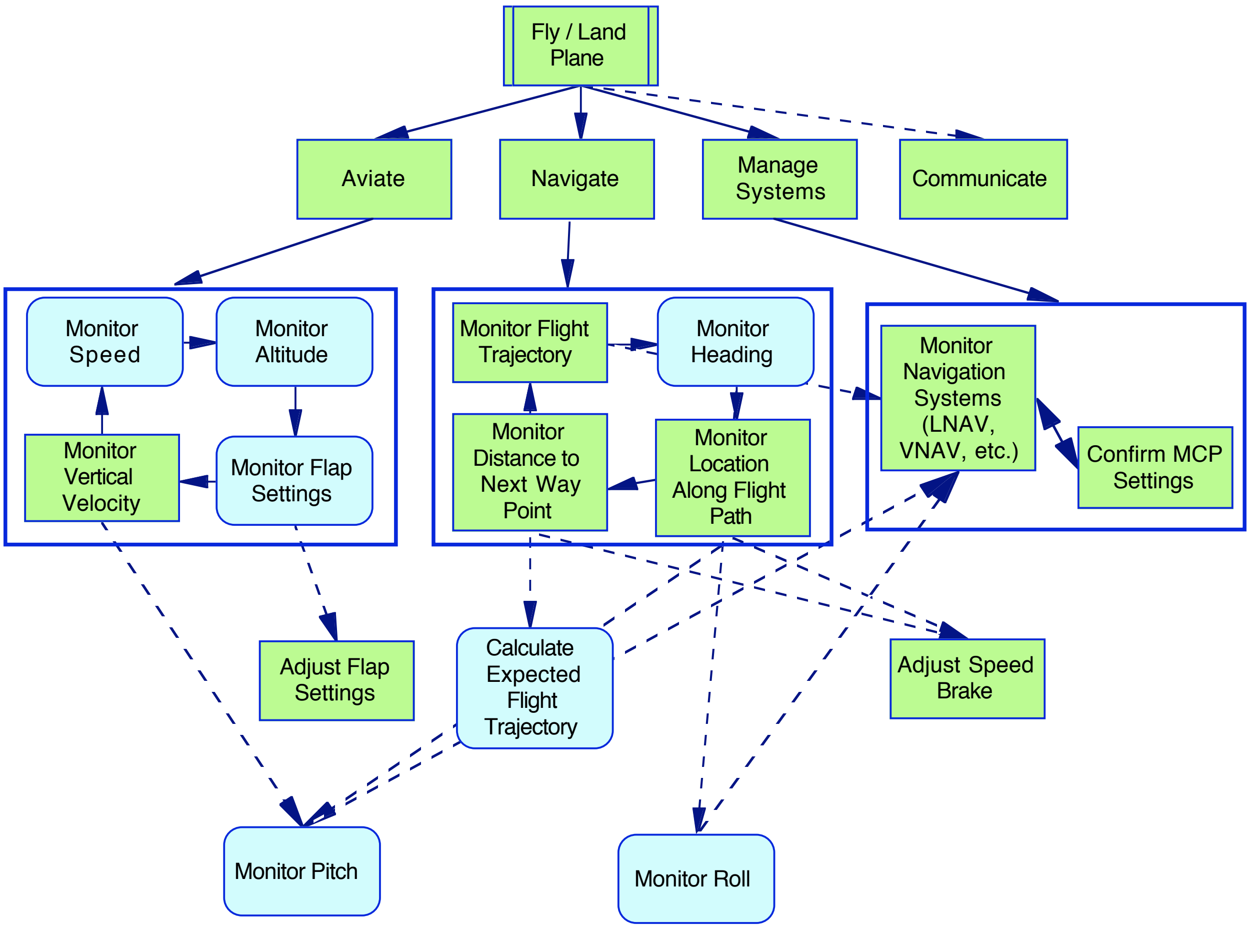

Figure 5. Flow of control resulting from task analyses. Dashed lines represent conditional subgoals that may not be executed every time; blue rounded boxes indicate that some information required by that subgoal may be found on the SVS, if present.

The ACT-R architecture provided a great deal of constraint as well. Working within the parameters of the architecture set certain boundaries and delimits scope. In particular, it meant that we were 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 in order to be maintained accurately.

Because the eye-movement data were the primary focus of the modeling effort, we 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; one of our ultimate aims is 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.

The final ACT-R model is driven primarily top-down by the goal structure, and seeks information from the environment in order to fulfill the requirements of keeping its representation of flight state up-to-date. Many times the model will determine that it needs to know the current altitude. When this information is available in more than one location (e.g., both on the PFD and the SVS), the current location of gaze is a large determiner of where the model will look. Thus, the model was 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 (that is, the symbology on the SVS is redundant with information on the PFD and NAV displays) and low-level parameters of the human visual system, such as saccade latency and accuracy.

The Main Empirical Findings

As described in detail in Kirlik & Byrne (2004), our analyses of the NASA supplied eye movement data over selected phases of flight and experimental conditions. Here we will focus on a few particularly central findings. 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. The SVS should thus be particularly relevant in later phases of approach and landing as the pilot visually acquires the airport and runway.

When we collapsed eye movement data over flight phase, the following results emerged. Pilots clearly spent a fair proportion (21.8% of total fixations, when present) of their gaze on the SVS. An interesting question we wanted to address is from where they "stole" these gazes. Namely, they must have reduced some portion of fixations associated with other areas of interest. A natural suspect for the location from which fixations would be stolen was the OTW display, due to the redundancy between the perceptual information gained from the SVS and OTW displays.

Surprisingly, however, the fixation dwell time and frequencies on different areas of interest under SVS versus baseline (No SVS) conditions revealed no obvious difference in the amount of gaze directed out the window. Instead, it appeared the SVS was 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 found this result both counterintuitive and quite interesting. Instead, our data analyses indicated 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.

Subsequent analysis broken down by phase of flight (approach toward landing) indicated that this tendency only increased with flight phase, and that pilots were not using the SVS as merely a proxy for OTW information, but mainly using it as a proxy for the PFD and NAV displays, due to redundantly displayed information on both.

Evaluating and Validating the ACT-R Model

One evaluation and validation approach that could have been taken would be to perform split-half validation based on the eye movement data provided. 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, we decided to capitalize on the strengths of the supplied data. While there are few subjects, there was 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 (Wickens, 2002); 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 the more fine-grained level of transitions. That is, we asked how well the model, with parameters selected to fit the more global data, fits the more detailed data in the transition matrices.

Flight 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?

There were a few parameters that could be tweaked in this model which affect how well it captures the data. These generally involve strategy selection (conflict resolution) 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 could 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, as the values could be estimated by considering the distance to be traveled by the saccade, which affects how long it takes on average to successfully complete. Note that these parameters 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 display 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. These two parameters were then kept constant through all simulations. Values were selected to produce a good fit to the 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. At the most qualitative level of fit assessment, 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.

Since this was the first criterion, the model's best fit is to the total fixations in each region for Phase 2 with SVS, as depicted in Figure 6. 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. 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 manipulation, and is shown in Figure 7. 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.

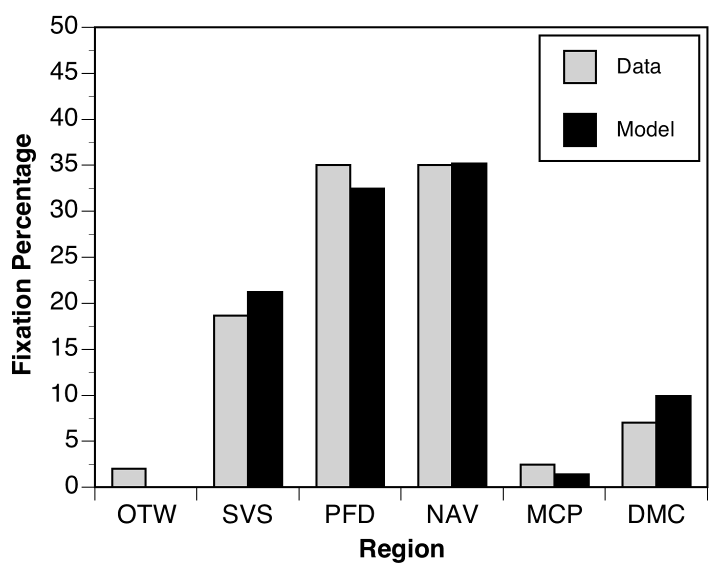

Figure 6. Model vs. data overall attention allocation for Phase 2 with SVS

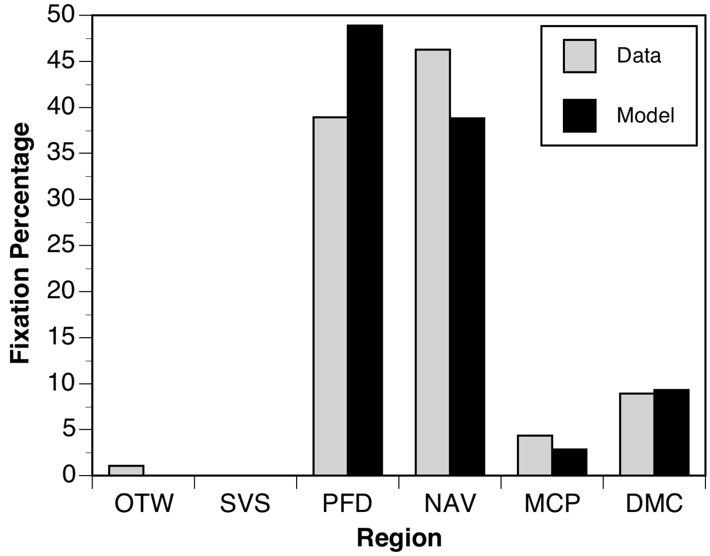

Figure 7. Model vs. data overall attention allocation for Phase 2 without SVS

The next question is how the model did 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 a useful test to see if the model can capture the differences between the two phases. Figure 8 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).

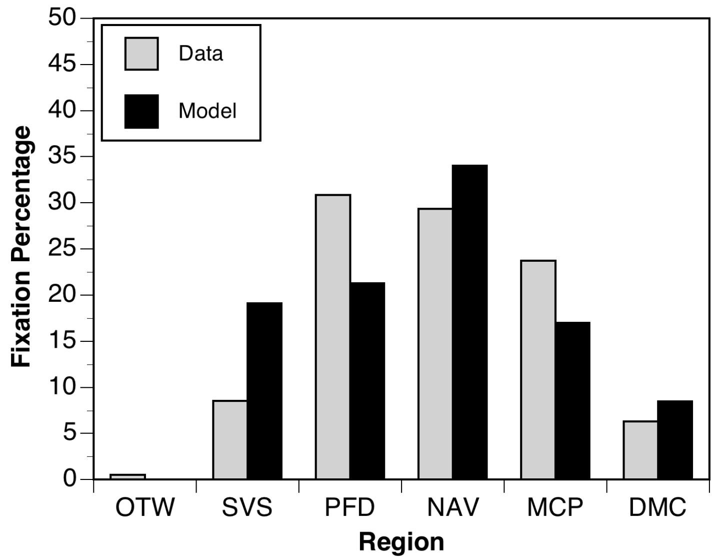

Figure 8. Model vs. data overall attention allocation for Phase 1 with SVS

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 9.

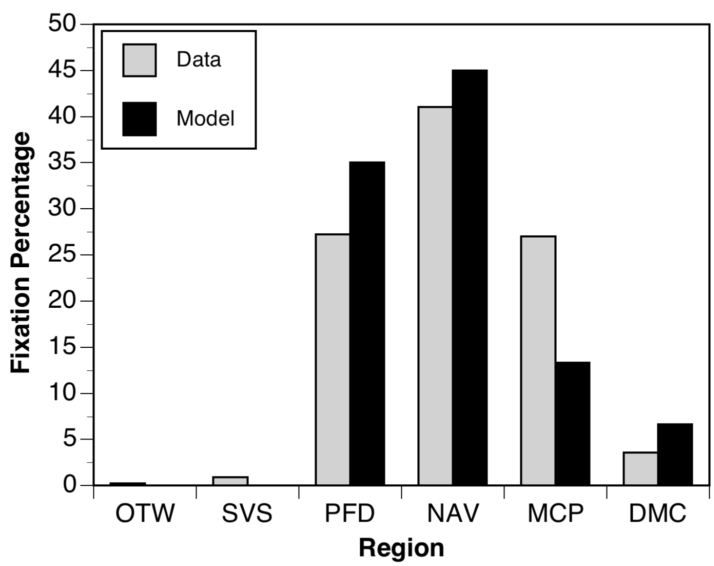

Figure 9. Model vs. data overall attention allocation for Phase 1 without SVS

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, as mentioned before, there exists a more fine-grained level at which the model may be validated: 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.

Presentation of all the data here would be somewhat laborious so we will illustrate with a particular selection of it (for a more complete presentation, see Byrne & Kirlik 2004). 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. Table 1 presents the human fixation transition matrix for phase 2 without SVS, while Table 2 presents the model's matrix under the same conditions. The model explains nearly 80% of the variance in the data with an r-squared of 0.772.

Table 1. Human 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 2. 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 does a not quite as good a job overall. The NAV and PFD displays were well-fit, but was 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. For the no-SVS Phase 1, the match to empirical data 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. Just as with the overall attention allocation, the model does not do quite as well in the SVS condition with Phase 1. 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. Overall, thought, 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.

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-fitting, 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, at least in the early phases of flight, entirely as a result of the symbology overlaid on the SVS. 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.

More generally, we believe we have developed a model which is dynamically sensitive to both top-down (i.e, task structure and information-seeking goals) and bottom-up (i.e., parameters of visual attention and structure of the information environment) constraints. This is an important insight in and of itself; one cannot predict performance by looking just at a system interface or by considering just the task structure. Integrated, closed-loop computational cognitive modeling seems necessary to reach such goals.

CONCLUSIONS

Along with our colleagues also reporting their modeling research in this volume, we hope we have demonstrated that computational cognitive modeling, suitably extended and amended to handle the complexities of operational environments, has a promising future as an engineering tool for the analysis, design, and evaluation of human-machine systems. Just as importantly, our chapter should also indicate how far we have to go before this sort of modeling can reliably be put in the hands of professional practitioners. Compromises had to be made, extensions had to be improvised, and some *ad hoc* assumptions made simply to close the loop on a computational modeling approach that has, until very recently, been applied primarily to static and isolable tasks characteristic of the experimental psychology laboratory. As these modeling techniques mature, we certainly should expect even more rigor in model validation than could be provided in this effort, motivated largely to push the boundaries and test the limits of this style of modeling.

That said, no science has yet sustained itself without spawning a socially-relevant and valuable branch of engineering or technology, and we expect that cognitive science will be no different in this regard. As such, we invite those cognitive scientists interested in furthering the development of cognitive architectures or modeling techniques to join us in this venture of taking, head on, the complexities of modeling human cognition and behavior in contexts that are at once both scientifically challenging and also of enormous practical consequence.

ACKNOWLEDGMENTS

The authors would like to acknowledge the following people who played an essential role in conducting the research reported here. They include David Huss and Chris Fick of Rice University, and Sarah Miller, Alex Kosorukoff, Ruei-Sung Lin, Sam Zheng 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, computer 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. The authors would also like to thank the editors of this volume for their generous support and assistance throughout this research.

REFERENCES

Anderson, J. R. (1990). *The Adaptive Character of Thought.* Hillsdale, NJ: Lawrence Erlbaum Associates.

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*.

Anderson, J. R., & Lebiere, C. (1998) *The atomic components of thought.* Mahwah, NJ:Lawrence Erlbaum.

Baron, S. (1984). A control theoretic approach to modeling human supervisory control of dynamic systems. In W. B. Rouse (Ed.), *Advances in Man-Machine Systems Research, Vol. 2.* Greenwich, CT: JAI Press.

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).

Bruner, J. S. (1957). Going beyond the information given. In J. S. Bruner, E. Brunswik, L. Festinger, F. Heider, K. F. Muenzinger, C. E. Osgood, & D. Rapaport, (Eds.), *Contemporary Approaches to Cognition* (pp. 41-69). Cambridge, MA: Harvard University Press. Reprinted in Bruner, J. S. (1973). *Beyond the Information Given* (pp. 218-238). New York: Norton.

Byrne, M. D. (2003). Cognitive architecture. In J. A. Jacko & A. Sears (Eds.), *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications* (pp. 97–117). Mahwah, NJ: Lawrence Erlbaum.

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

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.

Cassell, R., Smith, A., and Hicok, D. (1999). Development of airport surface required navigation performance (RNP). NASA/CR-1999-209109.

Cheng, V. H. L., Sharma, V., and Foyle, D. C. (2001). Study of aircraft taxi performance for enhancing airport surface traffic control. Manuscript accepted for IEEE-ITS publication.

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

Foyle, D. C., Andre, A. D., McCann, R. S., Wenzel, E., Begault, D. & Battiste, V. (1996). Taxiway navigation and situation awareness (T-NASA) system. Problem, design, philosophy, and description of an integrated display suite for low visibility airport surface operations. *SAE Transactions: Journal of Aerospace, 105*, 1511-1418.

Gigerenzer, G. & Goldstein, D. G. (1996). Reasoning the Fast and Frugal Way: Models of Bounded Rationality," *Psychological Review*, *103*, pp. 650-669.

Goodman, A. (2001). Enhanced descriptions of off-route navigation errors in T-NASA2. NASA Ames Research Center, Moffet Field, CA, USA.

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.

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.

Hammond, K.R. & Stewart, T.R. (2001). *The Essential Brunswik: Beginnings, Explications, Applications*. New York: Oxford University Press.

Hooey, B. L. & Foyle, D. C. (2001). A post-hoc analysis of navigation errors during surface operations. Identification of contributing factors and mitigating strategies. *Proceeedings of the 11th Symposium on Aviation Psychology*. Ohio State University, Columbus, OH.

Hooey, B. L., Foyle, D. C. & Andre, A. D. (2000). Integration of cockpit displays for surface operations: The final stage of a human-centered design approach. *SAE Transactions: Journal of Aerospace, 109*, SAE, Warrendale, VA, USA, 1053-1065.

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge: MIT Press.

Jagacinski, R. & Flach, J. (2002) *Control Theory for Humans*. Mawhah, NJ: Erlbaum.

Kirlik, A. (2003). Human factors distributes its workload. Review of E. Salas (Ed.), "Advances in Human Performance and Cognitive Engineering Research, Vol. 1." *Contemporary Psychology, Vol. 48,* No. 6.

Keller, J., Leiden, K., & R., S. (2003). Cognitive task analysis of commercial jet aircraft pilots during instrument approaches for baseline and synthetic vision displays.

Kirlik, A., Miller, R.A., and Jagacinski, R.J., (1993). Supervisory control in a dynamic uncertain environment: A process model of skilled human-environment interaction. *IEEE Transactions on Systems, Man, and Cybernetics*, *Vol 23*(4).

Kirlik, A. and Bisantz (1998). Cognition in human-machine systems: Experiential and environmental components of adaptation. In P. Hancock (Ed). Human Performance and Ergonomics. New York: Academic Press.

Klein, G., (1989). Recognition-primed decisions. In W. B. Rouse (Ed.), *Advances in Man-Machine Systems Research, Vol. 5* (pp. 47-92). Greenwich, CT: JAI Press.

MacKenzie, I. S. (2003). Motor behaviour models for human-computer interaction . In J. M. Carroll (Ed.) *Toward a Multidisciplinary Science of Human-Computer Interaction,* (pp. 27-54). San Francisco, CA: Morgan Kaufmann.

National Aeronautics and Space Administration (2001). Information and Data Package for Human Performance Modelers: Airport Taxi Operations with and without T-NASA at Chicago O'Hare. Unpublished documents, NASA Ames Research Center, Human Factors Research and Technology Divsion.

Payne, J. W. and Bettman, J. (2001). Preferential choice and adaptive strategy use. In G. Gigerenzer & R. Selten (Eds.). (pp. 123-146). *Bounded Rationality: The Adaptive Toolbox*. Cambridge, MA: MIT Press.

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.

Rasmussen, J. (1976). Outlines of a hybrid model of the process plant operator. In T. B. Sheridan & G. Johannsen, (Eds.) *Monitoring Behavior and Supervisory Control*, (pp. 371-383). New York: Plenum Press.

Rasmussen, J. (1983). Skills, rules, and knowledge: signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-13*, 257-266.

Rouse, W. B. (1984), *Advances in Man-Machine Systems Research, Vol. 2.* Greenwich, CT: JAI Press.

Salvucci, D. D. (2001). Predicting the effects of in-car interface use on driver performance: An integrated model approach. International Journal of Human-Computer Studies, 55, 85-107.

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.

Sheridan, T.B. (1976). Toward a general model of supervisory control. In T.B. Sheridan and G. Johansen, (Eds.), *Monitoring Behavior and Supervisory Control*. New York: Plenum Press.

Sheridan, T. B. & Johannsen, G. (1976). *Monitoring Behavior and Supervisory Control*. New York: Plenum Press.

Tolman, E.C. & Brunswik, E. (1935). The organism and the causal texture of the environment. *Psychological Review, 42*, 43-77.

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