

Using Computational Cognitive Modeling to Diagnose Possible Sources of Aviation Error

Michael D. Byrne

*Department of Psychology
Rice University*

Alex Kirlik

*Aviation Human Factors Division
University of Illinois at Urbana–Champaign*

We present a computational model of a closed-loop, pilot–aircraft–visual scene–taxiway system created to shed light on possible sources of taxi error. The creation of the cognitive aspects of the model with ACT–R (Adaptive Control of Thought–Rational) required us to conduct studies with subject matter experts to identify the experiential adaptations pilots bring to taxiing. Five decision strategies were found, ranging from cognitively intensive but precise to fast and frugal but robust. We provide evidence for the model by comparing its behavior to a National Aeronautics and Space Administration Ames Research Center simulation of Chicago O’Hare surface operations. Decision horizons were highly variable; the model selected the most accurate strategy given the time available. We found a signature in the simulation data of the use of globally robust heuristics to cope with short decision horizons as revealed by the errors occurring most frequently at atypical taxiway geometries or clearance routes. These data provided empirical support for the model.

The purpose of models is not to fit the data but to sharpen the questions.
—Samuel Karlin, 1983

Aviation incident and accident investigators often find both cognitive and environmental contributing factors to these events. Environmental sources include such factors as flawed interface design (e.g., Degani, Shafto, & Kirlik, 1999), confusing automation (e.g., Olson & Sarter, 2000), and unexpected weather conditions (Wiegmann & Goh, 2001). Cognitive sources include such factors as poor situation awareness (SA; Endsley & Smolensky, 1998), procedural non-compliance (Bisantz & Pritchett, 2003), and poor crew coordination (Foushee & Helmreich, 1988).

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 highly trained and motivated crews, 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 (Connolly, 1999) so that most slips and errors have few serious consequences. Most significant incidents and accidents result when cognitive, environmental, and perhaps even other (e.g., organizational) factors collectively conspire to produce disaster (Reason, 1990).

For this reason, an increasing number of human factors and aviation psychology researchers have realized that the common terms *human error* and *pilot error* often paint a misleading picture of error etiology (e.g., Hollnagel, 1998; Woods, Johannesen, Cook, & Sarter, 1994). By their nature, these terms predicate error as a property of a human or pilot, in contrast to what has been learned about the systemic, multiply caused nature of many operational errors. These often misleading terms only contribute to the “train and blame” mindset still at work in many operational settings and perhaps contribute to the failure of such interventions to improve the safety landscape in settings from commercial aviation to military operations to medicine.

THE CHALLENGE POSED BY THE SYSTEMS VIEW OF ERROR

Although advances in theory may well present a more enlightened, systemic view of error, in our opinion, 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 an ultimate error and how these factors interact in typically dynamic, often complex, and usually probabilistic ways. To say that multiple-contributing factors conspire together to produce error is one thing. To provide techniques capable of representing these multiple factors and the precise manner in which they conspire is quite another. This problem is difficult enough

in the realm of accident investigation in which at least some evidence trail is available (Rasmussen, 1980; Wiegmann & Shappell, 1997). It is significantly more challenging, and arguably even more important, in the case of error prediction and mitigation (e.g., Hollnagel, 2000).

As a step toward addressing this problem, this article describes the results of a study in which dynamic and integrated, computational cognitive modeling, or more specifically, pilot–aircraft–scene–taxiway modeling, was performed to shed light on the possible sources of error in aviation surface operations, more specifically, taxi navigation. Modeling consisted of the integration of a pilot model developed within the ACT–R (Adaptive Control of Thought–Rational) cognitive architecture (Anderson et al., 2004; Anderson & Lebiere, 1998), a model of aircraft taxi dynamics, and models of both the visible and navigable airport surface, including signage and taxiways.

This modeling effort was motivated by experiments performed in a National Aeronautics and Space Administration (NASA) Ames' Advanced Concept Flight Simulator (for more detail, see Hooy & Foyle, 2001; Hooy, Foyle, & Andre, 2000). The purpose of the NASA experimentation was both to attempt to better understand the sources of error in aviation surface operations and to evaluate the potential of emerging display and communication technologies for lowering the incidence of error (Foyle et al., 1996).

The purpose of the cognitive system modeling research was to evaluate and extend the state-of-the-art in computational cognitive modeling as a resource for human performance and error prediction.

THE PROBLEM: TAXI ERRORS AND RUNWAY INCURSIONS

Errors made during navigation on an airport surface have potentially serious consequences, but this is not always the case. Many such errors are detected and remedied by flight crews themselves, others are detected and remedied by controllers, and many uncorrected errors still fail to result in serious negative consequences due to the sometime-forgiving nature of the overall multiagent space that constitutes the modern taxi surface. However, some errors in taxi navigation can result in drastic consequences.

A particularly pernicious type of error is the runway incursion, which is any occurrence involving an aircraft or other object creating a collision hazard with another aircraft taking off or landing or intending to take off or land. Since 1972, runway incursion accidents have claimed 719 lives and resulted in the destruction of 20 aircraft (Jones, 2000). The problem of runway incursion accidents continues to only get worse, despite acknowledgment of the importance of the problem by both the Federal Aviation Administration and the National Transportation Safety

Board and plans to remedy the problem with technologies such as the Airport Movement Area Safety System ("Runway Incursions," 2003). For example, the number of U.S. runway incursions in 1996, 1997, and 1998, totaled 287, 315, and 325, respectively. In 1999, a Korean Airlines airliner with 362 passengers swerved during takeoff at Chicago O'Hare International Airport (ORD) to avoid hitting a jet that entered the runway, and an Iceland Air passenger jet at John Fitzgerald Kennedy Airport (JFK) came within 65 m of a cargo jet that mistakenly entered the runway (Jones, 2000).

These problems show no immediate sign of going away. There were a total of 337 U.S. runway incursions in 2002, more than 1.5 times the number reported a decade earlier. "Runway Incursions," (2003) noted that "Despite FAA programs to reduce incursions, there were 23 reported in January, 2003, compared with only 14 in January 2002" (p. 15). Due in part to the inability to deal with incursion problems to date, NASA established an Aviation System-Wide Safety Program to address this and other challenges to aviation safety. The NASA simulation and technology evaluation study described in the following section represents one attempt to use aviation psychology and human factors research techniques to address critical challenges to aviation safety.

SIMULATION, EXPERIMENTATION, AND DATA COLLECTION

Called T-NASA2 (for more detail, see Hooy & Foyle, 2001; Hooy, Foyle & Andre, 2000) throughout this article, the experimental scenario required 18 flight crews, consisting of active pilots from six commercial airlines, to approach, land, and taxi to a gate at ORD. The flight crews had varying levels of experience with the ORD surface configuration. Experimentation used baseline conditions (for this study, chart technology only) as well as conditions in which pilots were provided with various new display and communication technologies, including a moving map and head-up displays with virtual signage (e.g., a superimposed STOP sign at a hold point). The modeling performed in this research focused solely on performance in the baseline (current technology) conditions.

T-NASA2 Data Set

Nine different taxiway routes were used in the baseline trials of the T-NASA2 simulation. Each of the 18 crews were tested over a balanced subset of three different routes for a total of 54 trials. Each trial began approximately 12 nm out on a level approach into ORD. Pilots performed an autoland, and the first officer (FO) notified the captain of their location with respect to the runway exit on the

basis of clearance information obtained during the final stages of flight and the paper airport diagram. As the aircraft cleared the runway, the crew tuned the radio to ground controller frequency, and the controller provided a taxi clearance (a set of intersections and directions) from the current location to the destination gate. Crews were then required to taxi to the gate in simulated, visually impoverished conditions (RVR¹ 1000'). Further details can be found in Hooey and Foyle (2001). It should be noted that the simulation represented neither all standard operating procedures (after-landing checklists, log, and company paperwork) nor all communication activities (with the cabin crew, dispatch, and gate). As a result, the level of crew workload was somewhat lower than a crew might experience in operational conditions, lending support to the idea that the experimental situation with respect to error was closer to best-case rather than worst-case conditions (other than low visibility).

Across the 54 baseline T-NASA2 trials, a total of 12 off-route navigation major errors were committed. Major errors were defined as deviation of 50 ft or more from the center line of the cleared taxi route. These major errors were used for the modeling effort because they were objectively determined with simulation data and did not require subjective interpretation for classification. On each, crews proceeded down an incorrect route without any evidence of immediate awareness or else required correction by ground control. The T-NASA2 research team designated these 12 to be major errors. Additionally, 14 other deviations were observed but were detected and corrected by the crews. These latter 14 deviations were thus classified as minor errors by the NASA team, and we were instructed that the modeling effort should focus solely on the major errors. NASA provided our modeling team with descriptions of each major error, in terms of intersection complexity, turn severity, and their own classification of each in terms of planning, decision making, or execution (Goodman; 2001; Hooey & Foyle, 2001).

Two aspects of the T-NASA2 data set provided the primary motivation for this modeling effort. First, it was believed that modeling might shed light on the underlying causes of the errors observed in the experimental simulations. A second motivation was the fact that the suite of SA and navigation aids used in the new technology conditions of the T-NASA2 experiments were observed to eliminate navigation errors almost entirely (Hooey & Foyle, 2001). The goal of our research, therefore, was to provide a systemic explanation for the errors that were observed in a fashion that was consistent with the finding that no errors were observed when the quality of information available to support navigating was improved.

¹Runway visual range (RVR) is the range over which the pilot of an aircraft on the centerline of a runway can see the runway surface markings or the lights delineating the runway or identifying its centerline.

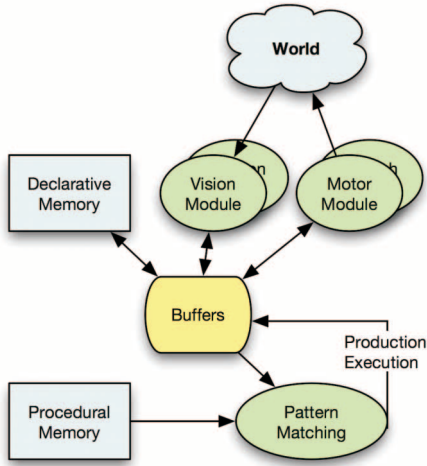


FIGURE 1 ACT-R (Adaptive Control of Thought-Rational) cognitive architecture.

ACT-R MODELING: A GENERAL OVERVIEW

ACT-R (Anderson & Lebiere, 1998; see also Anderson et al., 2004) is a computational architecture designed to support the modeling of human cognition and performance at a detailed temporal grain size. Figure 1 depicts the general system architecture. ACT-R allows for the 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 the retrieval of a simple fact. 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 the outputs of ACT-R 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*.

Procedural knowledge, 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 relation to this context. Because the actual statistics are often not known, in many

cases, these values are left at system defaults or are estimated by the modeler, although 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 msec 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 object to which it is attending.

ACT-R was originally designed to model the results of cognitive psychology laboratory experiments and is often considered a bottom-up or first-principles approach to the problem of modeling human cognition and performance. Whether ACT-R scales up to more complex domains is an empirical question, but so far, it has done well in dynamic domains such as driving (Salvucci, 2001), and we believe it is now mature enough to be tested in aviation.

CONSTRUCTING AN ACT-R MODEL OF TAXI PERFORMANCE

Taxiing a commercial jetliner is obviously a complex task, and the construction of an ACT-R model of a pilot performing this task was similarly complex along multiple dimensions.

Model Scope

One of the first decisions that had to be made was a decision about scope. In one sense, there are clearly multiple humans in the taxi loop, even in the somewhat simplified NASA simulation. These include the captain, who is actually head up, looking out the window, and controlling the aircraft, and the FO, who looks primarily head down and assists both the captain and the ground-based controller. 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 FO as items in the environment. We thought this decision was a good balance between tractability and relevance because the captain made the final decisions and also controlled the aircraft.

A second, important aspect of scoping model coverage was to select the psychological activities on which we would focus our efforts. Our research team was one of many teams also creating cognitive models of the same T-NASA2 data (e.g., see Deutsch & Pew, 2002; Gore & Corker, 2002; Lebiere et al., 2002; McCarley, Wickens, Goh, & Horrey, 2002). 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 multitasking, as ACT-R is less mature in this area than some other models, and we ruled out focusing on SA issues (losing track of one's location on the airport surface), as our model was less mature in this area than some other models. 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 a 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 the focus of another modeling team on SA errors associated with losing track of one's location on the airport surface (McCarley et al., 2002).

Model Environment

Thus, we created an ACT-R model of one human pilot, but this pilot model still had to be situated in an accurate environment. In this research, 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 simulated aircraft traveled. Each of these three environmental entities was computationally modeled and 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. Given vehicle size, mass, and dynamics, however, we still did require a somewhat reasonable approximation to the actual aircraft dynamics used in the experiments to be able to get a handle on timing issues. Although we were not interested in control issues per se, the dynamics of the aircraft played an important role in the determination of decision-time horizons, a key factor in the cognitive

representation of the pilot's activities. The aircraft model we constructed assumed that the pilot controlled the vehicle in three ways: by applying engine power, braking, and steering. For the purposes of modeling an aircraft during taxiing, these three forms of control are sufficient. On the basis of 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 the manual control of the acceleration and braking aspects of the model did play a role in the determination of the aircraft position relative to an impending turn and, importantly, placed a hard constraint on the maximum speed of approach of the aircraft 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). For our model, these speeds were found to be 20 knots for a soft (veer) turn, 16 knots for a right turn, and 14 knots for a U-turn and were based on actual turn-radius measurements from the ORD taxiway layout (all turns made in these scenarios could be classified according to this scheme). Although due to airport layout constraints, taxiing would not always occur at the maximum possible speed, these maximum speeds partially determined the time available to make a turn decision, and in our model, as this time was reduced, there was a greater probability of an incorrect turn 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 then executed without error.

To implement this aspect of the model, we decided to model the ORD airport taxiway as a set of interconnected rails on 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 software 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.

Modeling Pilot Background Knowledge

Obviously, the environment and its dynamic properties are critically important in understanding pilot performance in this domain, but they do not, of course, completely determine pilot behavior; thus, the use of a knowledge-based performance model such as ACT-R is necessary. As mentioned earlier, the ACT-R model must be supplied with the knowledge of how to do this task. This part of the model-building process is often referred to as *knowledge engineering* because the demands of gathering and structuring the knowledge necessary to perform the tasks in such domains are significant. We focused our efforts on the identification of procedures and problem-solving strategies used by pilots in this domain as well as the cost-benefit structure of those procedures and strategies.

Task Analysis and Knowledge Engineering

The task-specific information required to construct the model was obtained by the study of various task analyses of taxiing (e.g., Cassell et al., 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. On the basis of 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 such as ground vehicles that are difficult to detect in poor visibility; maintaining the speed of the aircraft because the dynamics of a commercial jetliner require relatively frequent adjustments of throttle, brake, or both 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. Although these tasks often have little direct impact on navigation, they do take time to execute, and time is the key limited resource in the making of navigation decisions in our integrated pilot-aircraft-environment system model.

With respect to navigation decisions, we found that decision making was highly local. That is, the planning horizon is very short; flight crews are quite busy in the time after landing and, thus, in situations such as 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, decide which turn to make. We found that in the error corpus available to us, 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 actual pilots in the kinds of conditions experienced by the pilots in the NASA study? Largely through 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. Many of these strategies involved open communications between ground-based controllers and other aircraft. For example, if Qantas Flight 1132 has just been given a clearance that overlaps with the clearance given to United Flight 302, one viable strategy for the United pilot is to simply follow the Qantas aircraft for the overlapping portion of the clearance.

Similarly, pilots can use dissimilar clearances to rule out certain decision alternatives. For example, when faced with an intersection that forces the pilot to choose between taxiways A10 and D, if the pilot has just heard another flight given a clearance, which involves A10, D is the more likely choice because the ground controller is unlikely to assign two aircraft to be on the same taxiway approached from different directions. It is unclear the extent to which these strategies were available to the pilots in the T-NASA2 study because details of what clearances were given to the (simulated) other aircraft and when such clearances were given were not available to us. Thus, we had no choice but to exclude these strategies from the model.

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: Although fast, this strategy is increasingly inaccurate as time lapses between the time at which the list of turns described in the clearance is obtained and the time at which turn execution is actually required.
2. Make turns toward the gate: Although 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 because more information is taken into account.
4. Derive from map or 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 multiturn intersections. However, we did include it as a pos-

sible 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 constrained working memory. However, the accuracy of Strategies 2, 3, and 5 is less cognitively constrained and instead is critically dependent on the geometry of actual clearances and taxiways. As such, we used an SME as a participant in the 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).

For this study, Jeppesen charts for all major U.S. airports were made available to the SME, a working B-767 pilot for a major U.S. carrier. He was asked to select charts for those airports for which he had significant experience of typical taxi routes, and he was asked to draw, with a highlighter on the charts themselves, the likely or expected actual taxi routes at each airport from touchdown to the gate area of his company. 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).

MODELING TAXI DECISION HUERISTICS

To obtain this information, which was required to inform modeling, we provided our SME Jeppesen charts for all 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 nine airports (Dallas-Fort Worth, Los Angeles, San Francisco, Atlanta, JFK [Kennedy Airport, New York], Denver, Sea-Tac [Seattle-Tacoma], Miami, and O'Hare). The SME was asked to draw, with a highlighter on the charts themselves, the likely or expected taxi routes at each airport from touchdown to the gate area of his company. A total of 284 routes was 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 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 nine airports studied. We recognize that these accuracy estimates may be specific to the (major) carrier for whom the SME flew because other the gates for other carriers may be located in areas at these nine 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 was quite good across the board, and the even simpler toward-terminal heuristic was 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 the five strategies described previously, including the two surprisingly frugal and robust toward-terminal and XY heuristics portrayed in Figure 2. One can think of these five 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.

Detailed Description of Dynamic Decision Modeling

From a time–horizon (cost) perspective, the selection of decision strategies was informed by a procedure for the estimation of the time remaining before a decision had to be made. Time remaining was based on the distance of the aircraft to

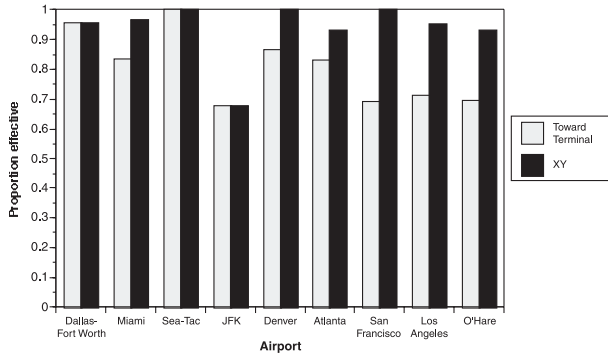


FIGURE 2 Accuracy of the toward-terminal and minimize the greater of the XY distance heuristics.

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, which was used to determine whether breaking could slow the aircraft sufficiently to negotiate the sharpest turn of an intersection.

This time remaining calculation was not implemented in ACT-R (i.e., we did not create a cognitive model of how the pilot estimated this time) but rather was made by a call from ACT-R to an external algorithm so that the model could determine which of the five decision strategies were available in any particular instance. Because we believed the pilots' abilities to estimate these times were imperfect, noise was added to the result of the computations on the basis of the aircraft model, such that the result returned was anywhere from 80 to 120% of the true time remaining.

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 sec 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 slightly modified version of the standard conflict resolution procedure of ACT-R. 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 sec 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, although there was a 1-sec penalty applied to productions whose time cost was within 0.5 sec of the remaining time, again, a conservative move to ensure that a decision strategy likely to be completed would 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. (The 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 yielded 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: the estimation of time remaining and the utilities of the strategies themselves.

In the pilot model, 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% because 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) 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 the decision time available. These simulation results are presented in Figure 3.

As is clear from Figure 3, as the decision horizon decreased, so did the likelihood that the pilot model would select a less accurate strategy. In fact, in the time window from about 2.5 to about 8 sec, the environmentally derived heuristics dominated alternative strategies. However, this could be viewed as adaptive because a fast and frugal strategy that could run to completion could frequently outperform an analytically superior decision strategy that had to be truncated due to time constraints (Gigerenzer & Goldstein, 1996). As such, these results are not necessarily surprising but do suggest that error-reduction efforts requiring new decision strategies will have to be evaluated in light of the availability of potentially more frugal heuristics that may yield relatively robust performance yet fail in situations where the environmental regularities embodied in these heuristics are not satisfied (Reason, 1990). For example, modeling indicated that the turn toward gate heuristic took approximately 2.5 sec to compute with 80% accuracy. A rational pilot would not favor a new strategy or technology over this heuristic unless the increased benefit–cost ratio of a novel decision strategy was significantly superior to this quick and dirty method.

EMPIRICAL ADEQUACY

Appropriate techniques for the verification and validation of human performance models based on computational, cognitive modeling is an issue of great

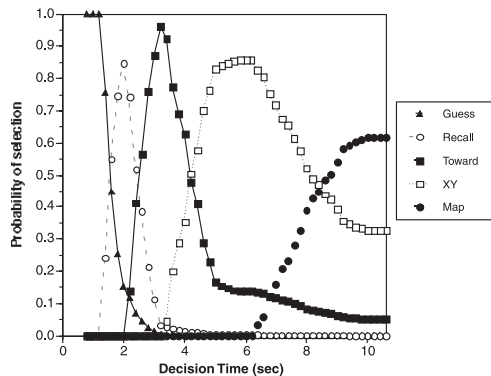


FIGURE 3 Selection probability for each decision strategy by the decision-time horizon.

current interest (see, e.g., Leiden, Laughery, & Corker, 2001), and it is fair to say that there are no unanimously agreed-on criteria in this area. 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. Because each route was run six times, a total of 582 intersection crossings occurred in the baseline trials. As mentioned earlier, 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 intersections crossed, the clearance indicated that crews should have proceeded in a direction toward the destination gate in 534 cases (91.8%), whereas the clearance directed crews in directions away from the gate in only 48 cases (8.2%). On 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 the 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 discussed in the following section, 4 of the 12 taxi errors were attributed not to decision making but rather to a loss of SA (i.e., losing track of one's position on the airport surface, see Goodman, 2001; Hooey & Foyle, 2001), a cognitive phenomenon beyond the scope of this modeling. Our modeling approach assumed that location knowledge (loss of SA) was not the primary factor in contributing to taxi error, but instead time-stressed decision making combined with what might be called *counterintuitive intersection* and *clearance pairs*, that is, 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 following, we briefly consider each error in turn. When we use the term *classifi-*

cation, we refer to the terms adopted by Hooley and Foyle (2001) and have bolded errors we believe to provide evidence for our model, especially for the fast and frugal decision heuristics used to make decisions under time stress. Italics are used to indicate errors due to loss of SA, as such are beyond the purview of our research, which thus provide neither support for or against our model, given our initial modeling focus. In the following, all quotations are from Goodman:

Error 1: This error was classified as a decision (as opposed to planning or execution) error, and it confirmed our modeling as the crew turned toward the gate when the clearance indicated a turn away from the gate.

Error 2: This error was also classified as a decision error associated with “lack of awareness of airport layout and concourse location” (p. 5). We thus considered this error due to a loss of SA.

Error 3: This error was classified as a planning error, in which the “crew verbalized that Tango didn’t seem to make sense because it was a turn away from the concourse” (p. 7). They thus turned in a direction toward the destination gate.

Error 4: This error was classified as an execution error due to perceptual confusion over center lines; the crew, nonetheless, prematurely turned in the direction of the concourse.

Error 5: This error was classified as an execution error, as vocalizations indicated the crew was aware of the proper clearance. However, they made a premature turn toward the gate.

Error 6: This error was classified as an execution error, as the captain stated that the lines were confusing but made a premature turn into the ramp area near the destination gate.

Error 7: This error was classified as a planning error, as the FO verbally omitted an intermediate turn in the clearance to Foxtrot. However, “the turn to Foxtrot would have led pilots away from concourse—Instead, FO suggested turning toward concourse on Alpha” (p. 15).

Error 8: This error was classified as a decision error, as the crew immediately made a turn toward the gate after exiting the runway, whereas the clearance prescribed a turn away from gate.

Error 9: This error was classified as an execution error, as the FO voiced confusion over center lines. Crews made a (one-gate) premature turn into the concourse area, whereas the clearance indicated they should proceed ahead further prior to turning into the concourse.

Errors 10, 11, and 12: Each of these errors was classified as a due to a loss of SA, due to the FO being “head down with Jepp chart, [and] didn’t realize where they were on the airport surface” (p. 21; Error 10), the crew’s “demonstrated lack of awareness of airport layout” (p. 23; Error 11), and “FO lacked awareness of their location on the airport surface” (p. 25; Error 12).

Although several of these errors were not, strictly speaking, classified as decision errors, we think it is revealing to note that the bulk of the errors classified as planning and execution errors were consistent with the same decision-making heuristics.

Summary

Errors in the T-NASA2 experimentation arose due to both poor SA and to turn-related decision making (Goodman, 2001). As described in an early section of this article, we decided to focus our modeling efforts on decision-related errors, thus complementing other modeling efforts that took SA-related errors to be the focus of their efforts. In summary, given the empirical results provided in this article, we conclude that there is reasonably good empirical support for our model.

CONCLUSIONS

We are encouraged by the results of this research to continue to pursue computational, cognitive models of human performance in dynamic, aviation contexts. We believe that the errors observed in the T-NASA2 scenario were consistent with the results of our analysis of information-impoverished, dynamic decision making and the mechanisms by which it was embedded in the ACT-R modeling architecture. As such, we believe that the view of cognition embodied in ACT-R, as constrained adaptation to the statistical and cost-benefit structure of the previously experienced task environment, achieves some level of support from this research.

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 & 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 was 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 hope that this research will motivate more members of the human factors and aviation psychology communities to study human performance issues with the benefits of emerging developments in computational cognitive modeling. We believe that detailed modeling of dynamic, integrated, human-machine-environment systems holds great promise for meeting the challenges posed by emerging, systems-oriented views of error etiology in complex, operational systems.

Implications

Obviously, the model presented here does not generalize directly to operational taxiing situations due to practical limitations in both the original study and the modeling effort itself. However, we believe that the ultimate lessons learned from this effort are relevant. This includes the general lesson that the details and dynamics of both the human cognitive system and the structure of the environment in which that system operates must be considered jointly, not in isolation from one another. More directly in the taxiing domain, this research suggests that taxi routes, which are inconsistent with the heuristics available to time-pressured flight crews are likely to be error prone and will continue to be so until a system that makes the correct route computable with greater speed and accuracy than those heuristics is made available to flight crews.

ACKNOWLEDGMENTS

This research was supported by NASA Ames Grants NCC2-1219 and NDD2-1321 to Rice University and NAG 2-1609 to the University of Illinois. We thank Captains Bill Jones and Robert Norris who served as SMEs, and Brian Webster, Michael Fleetwood, and Chris Fick of Rice University. We are grateful to the AvSP SWAP Human Performance Modeling Element research team, who provided their time, data, and expertise to this project, in particular, David Foyle, Tina Beard, Becky Hooey and Allen Goodman.

REFERENCES

- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Quin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *11*, 1036-1060.
- Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Bisantz, A. M., & Pritchett, A. R. (2003). Measuring judgment in complex, dynamic environments: A lens model analysis of collision detection behavior. *Human Factors*, *45*, 266-280.
- Cassell, R., Smith, A., & Hicok, D. (1999). *Development of airport surface required navigation performance (RNP)* (Report No. NASA/CR-1999-209109). Moffet Field, CA: NASA Ames Research Center.
- Cheng, V. H. L., Sharma, V., & Foyle, D. C. (2001). Study of aircraft taxi performance for enhancing airport surface traffic control. *IEEE Transactions on Intelligent Transportation Systems*, *2*(2), 39-54.
- Connolly, T. (1999). Action as a fast and frugal heuristic. *Minds and Machines*, *9*, 479-496.
- Degani, A., Shafto, M., & Kirlik, A. (1999). Modes in human-machine systems: Review, classification, and application. *The International Journal of Aviation Psychology*, *9*, 125-138.
- Deutsch, S., & Pew, R. (2002). Modeling human error in a real-world teamwork environment. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th annual meeting of the Cognitive Science Society* (pp. 274-279). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

- Endsley, M. R., & Smolensky, M. W. (1998). Situation awareness is air traffic control: The big picture. In M. W. Smolensky & E. S. Stein (Eds.), *Human factors in air traffic control* (pp. 115–154). San Diego, CA: Academic.
- Foushee, H. C., & Helmreich, R. L. (1988). Group interaction and flight crew performance. In E. L. Wiener & D. C. Nagel (Eds.), *Human factors in aviation* (pp. 189–227). San Diego, CA: Academic.
- 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*, 650–669.
- Goodman, A. (2001). *Enhanced descriptions of off-route navigation errors in T–NASA2*. Moffet Field, CA: National Aeronautics and Space Administration Ames Research Center.
- Gore, B., & Corker, K. M. (2002). Increasing aviation safety using human performance modeling tools: An air man–machine design and analysis system application. In M. J. Chinni (Ed.), *2002 military, government and aerospace simulation* (Vol. 34, No. 3, pp. 183–188). San Diego, CA: Society for Modeling and Simulation International.
- Hollnagel, E. (1998). *Cognitive reliability and error analysis method*. Oxford, England: Elsevier.
- Hollnagel, E. (2000). Looking for errors of omission and commission or *The hunting of the snark* revisited. *Reliability Engineering and System Safety*, *68*, 135–145.
- Hooley, B. L., & Foyle, D. C. (2001). A post-hoc analysis of navigation errors during surface operations. Identification of contributing factors and mitigating strategies. In *Proceedings of the 11th Symposium on Aviation Psychology*. Columbus: Ohio State University.
- Hooley, 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*, 1053–1065.
- Jones, D. (2000, February). *Runway incursion prevention system (RIPS)*. Paper presented at the SVS CONOPS Workshop, National Aeronautics and Space Administration Langley Research Center, Virginia.
- Karlin, S. (1983). *11th R. A. Fisher Memorial Lecture*. Lecture presented to the Royal Society, London.
- Kirlik, A. (1998). The ecological expert: Acting to create information to guide action. In *Fourth Symposium on Human Interaction with Complex Systems*. Piscataway, NJ: IEEE Computer Society Press. Retrieved October 2001, from <http://www.computer.org/proceedings/hics/8341/83410015abs.htm>
- Kirlik, A., Miller, R. A., & 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*, *23*(4), 929–952.
- Lebiere, C., Bielfeld, E., Archer, R., Archer, S., Allender, L., & Kelly, T. D. (2002). Imprint/ACT–R: Integration of a task network modeling architecture with a cognitive architecture and its application to human error modeling. In M. J. Chinni (Ed.), *2002 military, government and aerospace simulation* (Vol. 34, No. 3, pp. 13–19). San Diego, CA: Society for Modeling and Simulation International.
- Leiden, K., Laughery, R., & Corker, K. (2001). *Verification and validation of simulations*. Retrieved January 2002, from https://postdoc.arc.nasa.gov/postdoc/t/folder/main.ehtml?url_id=90738
- McCarley, J. S., Wickens, C. D., Goh, J., & Horrey, W. J. (2002). A computational model of attention/situation awareness. In *Proceedings of the 46th annual meeting of the Human Factors and Ergonomics Society* (pp. 1669–1673). Santa Monica, CA: Human Factors and Ergonomics Society.
- Olson, W. A., & Sarter, N. B. (2000). Automation management strategies: Pilot preferences and operational experiences. *The International Journal of Aviation Psychology*, *10*, 327–341.
- Payne, J. W., & Bettman, J. (2001). Preferential choice and adaptive strategy use. In G. Gigerenzer & R. Selten (Eds.), *Bounded rationality: The adaptive toolbox* (pp. 123–146). Cambridge, MA: MIT Press.

- Rasmussen, J. (1980). What can be learned from human error reports? In K. D. Duncan, M. M. Gruneberg, & D. Wallis (Eds.), *Changes in working life* (pp. 97–113). Chichester, England: Wiley.
- Reason, J. (1990). *Human error*. Cambridge, England: Cambridge University Press.
- Runway incursions. (2003). *NTSB Reporter*, 21(3), 4–5.
- 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.
- Wiegman, D. A., & Goh, J. (2001). *Pilots' decisions to continue visual flight rules (VFR) flight into adverse weather: Effects of distance traveled and flight experience* (Tech. Rep. No. ARL-01-11/FAA-01-3). Savoy: University of Illinois, Aviation Research Laboratory.
- Wiegmann, D. A., & Shappell, S. A. (1997). Human factors analysis of post-accident data: Applying theoretical taxonomies of human error. *The International Journal of Aviation Psychology*, 7, 67–81.
- Woods, D. D., Johannesen, L. J., Cook, R. I., & Sarter, N. B. (1994). *Behind human error: Cognitive systems, computers and hindsight*. Columbus, OH: CSERIAC.

Manuscript First Received: June 2003