Newman, R.L. and Foyle, D.C. (2003). Proceedings of the Twelfth International Symposium on Aviation Psychology, 873-882. Dayton, OH: Wright State University.

Test Scenarios for Rare Events

Richard L Newman* and David C Foyle†

Department of Safety Science
Embry-Riddle
Prescott, Arizona

Ames Research Center
NASA
Moffett Field, California

Abstract

During testing of cockpit systems and displays, a frequent objective is to evaluate a system's ability to maintain the pilot's "situation awareness" (SA). That maintenance of situation awareness is critical to flight safety is a truism. While SA is, no doubt, required for pilots and other crewmembers to make correct operational decisions, it has been difficult to identify precise objective test criteria.

The approach taken is to compare experimental studies over the past several years with various human error models. This review shows four different experimental approaches to rare event test and evaluation. These match up well with the various situation awareness and human error taxonomies. Several proposed test plans are reviewed and changes in the scenarios made to match the appropriate human error model. The goal is to develop experimental scenarios that will yield suitable data and, at the same time, make efficient use of experimental facilities.

Background

During testing of cockpit systems and displays, we are often called upon to evaluate a systems ability to maintain the pilots "situation awareness" (SA). That maintenance of situation awareness is critical to flight safety is a truism. While SA is, no doubt, required for pilots and crewmembers to make correct operational decisions, it has been difficult to find precise definitions for use in developing test criteria. It is important to keep our discussion of SA in an operational setting with the aim of developing valid test criteria.

What we are really interested in, from an operational point-of-view is what the effect is on operational decisions — What does the pilot do when confronted as circumstances develop? Does the pilot make correct decisions? Does he/she have the correct information to make these decisions? Besco (1996) described the attributes of good pilots by asking what should they do. A similar approach should be our criteria for developing SA test criteria — Are the decisions made by the pilot correct?

Situation awareness (SA) may be defined as the correct perception of the environment, the current situation, and the projection of future status. Jones and Endsley (1996) describe three levels of Loss of Situation Awareness (LOSA):

- ✈ Level 1: Fail to perceive information or misperception of information
- ✈ Level 2: Improper integration or comprehension of information
- ✈ Level 3: Incorrect projection of future actions of the system.

Perhaps the most fundamental form of LOSA happens when the pilot becomes unaware of his geographical position. In the extreme form, this can be described as "lost." However, in a less extreme form, the pilot may simply be unaware of the aircraft's proximity to terrain. This can lead to controlled flight into terrain (CFIT) in which the pilot files the airplane into the ground under complete control.

---

* Associate Professor, Department of Safety Science, Embry-Riddle Aeronautical University, 3200 Willow Creek Road, Prescott, Arizona 86301; currently on sabbatical at University of Iowa, Iowa City, Iowa 52242; telephone: 319-335-5857; e-mail: rlnewman@ccad.uiowa.edu

† Human-Automation Integration Research Branch, Ames Research Center, National Aeronautics and Space Administration, Moffett Field, California 94035; telephone: 650-604-3053; e-mail: dfoyle@mail.arc.nasa.gov

A loss in awareness of aircraft position relative to the track indicated by instrument flight procedures is also important. Coyle and Krolak (1999) discuss the issues related to maintenance of geographical situation. In addition to citing an array of LOSA accidents, they indicate that 30% of initial instrument rating flight checks are failed, due to loss of situation awareness, mostly while flying published procedures.

The pilot may have good terrain awareness, but not know his/her location relative to obstacles. This type of accident is referred to as "collision with obstacle." Collisions with terrain or obstacles may also be a result of the pilot's inability to assess the aircraft's flight path relative to the obstacle. In these accidents, the pilot may be well aware of the presence of the obstacle (or terrain), but either believes the aircraft will clear the ridge or is unaware of aircraft performance.

An incorrect flight path assessment can also lead to a landing short of the runway. This is often true when adverse conditions make such assessment difficult. Typical adverse conditions include darkness, restrictions to visibility, or precipitation. Night circling approaches are particularly vulnerable. In incorrect performance assessment accidents, the pilot is aware of the terrain or obstacle, but simply misinterprets the aircraft trajectory relative to the obstruction.

Loss of geographic awareness can lead to increased pilot workload during diversions or other flight plan changes. If the pilot is confused to his/her geographical location or is not clear where his destination or proposed route is, he/she may make an incorrect decision for his initial heading or altitude. This was evident in the Boeing 757 accident near Cali in late 1995 (ACRC, 1997). In this accident, the crew accepted a revised clearance to a different runway than originally planned. The increased workload during the reprogramming of the flight management system led to an incorrect turn away from the fix, rather than toward the fix.

It can be difficult in many aircraft for the pilots to monitor their system's status during high workload portions of the flight. In particular, failures of the flight instruments themselves may lead to misinterpretation and subsequent loss of control. In a 1974 accident, the crew inadvertently left the pitot heat off. When the pitot tubes iced over during the climb, the airspeed indications became erroneously high. Unfortunately the crew failed to detect the discrepancy and ultimately stalled the airplane resulting in a fatal crash (NTSB, 1975).

It can be very difficult to deal with failed or invalid flight instruments even if the pilot realizes that there is a problem (Coyle and Krolak, 1999). In addition to detecting a failure, it is important for displays to provide sufficient cues to allow crewmembers to diagnose and correct system failures (Singer and Dekker, 2000).

Spatial disorientation (SDO) of pilots has been a persistent problem in aviation since the first flight into (and probable spiral dive out of) a cloud. It has long been known that pilots are unable to maintain straight and level flight without visual cues (Ocker and Crane, 1932). Benson (1965) defines SDO as "an experience occurring inflight in which there is a defect in the aviator's perception of the attitude or position of his aircraft* or where conflicting perceptions give rise to confusion or uncertainty†."

Test and Evaluation (T&E)

There are several elements of situation awareness that flight tests must be designed to evaluate. These elements can be divided into continuous events and rare events. Continuous events are those that happen throughout the flight, such as

- ✈ Monitoring position/trajectory relative to flight plan
- ✈ Assessing aircraft performance

Continuous event evaluation, particularly awareness of position/trajectory relative to complex flight paths is a common experimental scenario for SA evaluations. Several recent published results are available (Snow and Reising, 1999; Williams, 2000).

* Equivalent to Jones and Endley's (1996) Level 1 LOSA

† Could be either Level 1 or Level 2 LOSA

Rare events include occurrences that do not happen often, but have the potential for major accidents, such as

- ✈ Detecting proximity relative to terrain
- ✈ Detecting external threats (obstacles or traffic*)
- ✈ System failure detection

Experimentally, these three are more difficult to handle. The crux of the problem is "how do we test for pilot response to rare events?" Two issues are key: pilot recognition and pilot reaction. Many T&E studies have concentrated on pilot reaction (such as the typical unusual attitude recovery studies); fewer have examined the ability of pilots to recognize a situation. Even fewer have examined pilots' reaction to a "rare event."

In a way, studies of these rare events are similar to vigilance studies, where subjects are exposed to hours of boredom waiting for a singular event. As such, they are not very efficient in terms of time. Unlike vigilance studies, it can be very expensive to place pilots in a simulator for many trials until the "rare event" happens. Pilot recognition to a system failure is usually conducted by asking the subject pilot to fly a particular pattern and subjecting them to some pre-determined failure. Normally, the subject pilot is not aware of the real purpose of the test, and may be supplied an alternative reason.

Newman and Quam (1982) examined general aviation pilot reaction to an unexpected vacuum gyro failure. The subject pilots were told the purpose of the flight experiment was to measure their tracking performance using conventional VOR navigation. The real purpose was to determine if they could recognize and cope with a vacuum system failure. At a predetermined (but unknown to the subject pilot) time, the safety pilot closed a valve shutting off the vacuum system and causing an attitude and directional gyro spin-down. In a post-experiment debriefing, the subject pilots reported that they were surprised by the failure and were not expecting it.

In a simulator experiment, Smith and Chappelow (1995) initiated total engine failure to a group of helicopter pilots. The subject pilots were informed that the exercise was a training sortie. The simulator instructor initiated the failure at unannounced times and the recovery was monitored. Smith and Chappelow measured reaction times and total recovery times, thus separating recognition from recovery. They did not mention a debriefing question relating to the surprise of the failure, but it is likely that pilots undergoing a simulator training exercise would expect engine failures.

In both cases, indeed in most failure detection experiments, only one valid data point per subject pilot is possible.

Singer and Dekker (2000) studied multiple failures presented to subject pilots during high workload instrument approaches. Their study was less concerned with failure detection than with the ability of the various displays tested to assist the pilots with diagnosis and corrective actions. The scenario was an instrument approach in a part-task simulator. Each subject pilot flew sixteen approaches with four different systems display suites. During each approach a systems failure was presented and the subject pilot was tasked with taking appropriate corrective action.

Perhaps the most common "rare event" tested is the aircraft unusual attitude (UA) recognition and recovery. However, most, if not all, evaluations place the subject pilot in a series of UAs in a series of rapid-fire recognize and recover with no element of surprise. Simulator evaluations normally reset the initial conditions in one of a series of extreme attitudes and require the subject pilot to recover using standard techniques.

During flight evaluations, the evaluation display is blanked (or the subject pilot asked to close his/her eyes) and the safety pilot flies the aircraft into a predetermined unusual attitude at which point the display is restored and the flight controls are given back to the evaluation pilot for recovery. Newman and Greeley (2001) out-

---

* Detection of airborne traffic in visual conditions is a continuous event. Detection of airborne traffic during instrument conditions is a rare event. Runway incursions are rare events. This may partially explain differences in traffic detection between Fisher, Haines, and Price (1980) and Anderson and Newman(1994).

line conventional test techniques for UA recovery.

Such evaluations are useful for comparing variations in flight displays for recovery, but do not measure the ability of the pilot to monitor aircraft attitude or recognize a rare upset. Gallimore, Patterson, Brannon, and Nalepka (2000) modified the entry conditions by having the subject pilot fly as a wingman. During the run, the lead aircraft flew into an unusual attitude and was removed from the visual scene. At this point, the subject pilot was to recover from the unusual attitude.

Most reports of experiments dealing with unusual situations have embedded one or two rare events into another study. Often the evaluation of unusual events is the primary purpose for the study and, for example, tracking performance measurement a fall-out of the study. Sometimes the reverse is true.

Typical experiments, described below, embedded the rare event into another study. In other words, the subject pilot thought he was being examined for one purpose and was unaware of the real purpose of the experiment.

<u>Runway incursions</u>: In the late 1970s, Fischer, Haines and Price (1980) examined the ability of pilots to detect runway incursions while using head-up displays. In this experiment, each subject pilot flew 19 ILS approaches for data in a simulator. Twelve of these approaches were flown using a HUD, seven without a HUD. During one of the HUD approaches and one of the no-HUD approaches the visual scene showed an airplane on the runway. Each pilot experienced the simulated incursion twice, once with a HUD and once without. There was a difference in subject pilot reaction with faster reaction times during the second exposure.

In Fischer *et al.* (1980), the difference in reaction times between first and second encounter makes using more than one encounter highly questionable.

In a civil HUD certification, attention was paid to traffic detection. During this flight test (Anderson and Newman, 1994), the pilot flying by reference to the HUD tended to spot traffic before the safety pilot (who had been assigned responsibility for looking for traffic). While looking for traffic is not a rare or unexpected event, there is apparently a difference in pilot behavior between simulated and actual flight.

<u>HUD misalignment</u>: Naish (1964) conducted a HUD tracking experiment in a simulator. During one run, the runway and the HUD were purposely misaligned. All subjects ignored the HUD commands and flew visually to the runway.

During a flight evaluation of a HUD to aid in midair retrieval, a gyro error created an inadvertent misalignment between the HUD symbology and the real world. In this case, the subject pilot ignored the real world and flew by reference to the HUD causing a hazardous situation (Newman, 1977). In a follow-on study, the gyro was purposely misaligned, creating similar situation. Again, the subject pilots flew using the display, not the real world (Ross, 1976).

While these are HUD-related issues, the point for immediate discussion is the difference in pilot response when a misalignment is encountered in a simulator and in-flight. In the simulator, Naish's subject pilots ignored the HUD and followed the external cues. In flight, Newman's subject pilots ignored the external cues and followed the HUD cues.

In both cases, each subject pilot encountered the situation once during a series of several experimental runs.

<u>Unexpected threats</u>: There have been a couple of attempts to overcome the difficulty of only allowing one (or, at most, two) "events" per subject. Hofer, Braune, Boucek, and Pfaff (2000) had twelve subject pilots fly a series of instrument takeoffs and instrument approaches. During each takeoff/approach sequence, each pilot experienced an "event" for a total of 16 "events" per pilot. The events were divided into display events (such as a frozen airspeed indication), a scene event (such as a runway incursion), or a display+scene event. Of the 192 events (for all pilots), 72 were serious enough to be considered a hazard. Overall, 36% of the HUD and 26% of the HDD events were missed.

More recently, McKay, Guirguis, Zhang, and Newman (2002) evaluated enhanced vision systems to aid in pilot detection of "anomalies." Their anomalies were arranged in similar fashion to those of Hofer *et al* (2000). Again, significant "anomalies" were missed, in spite of the pilots being warned that there would be one per approach.

Human Error Taxonomy

Endsley (1988a) describes situation awareness as the perception of the elements in the environment, their comprehension, and the projection of their status in the near future. She developed a model where the environment and objectives are fed through a situation awareness filter to lead to a decision. Table I summarizes the SA Error Taxonomy.

The recommended evaluation called for frequent interruptions with pilot perception of the elements, comprehension of the elements, and projection determined from specific questions.

Rasmussen (1982) developed an approach for diagnosing cognitive failures. He considered information errors, diagnostic errors, goal errors, and strategy errors. Rassmussen's taxonomy is shown in Table II.

Reason (1990) classifies human action into three levels: skill-based, rule-based, and knowledge-based actions. Skill-based actions are routinely practiced and automatic, such as keeping a car in its lane. Rule-based actions are a combination of conscious and unconscious processes, such as routine takeoffs and landings. Knowledge-based actions involve cognitive conscious thought to accomplish a process.

Approach to "Rare Event" Evaluations

What do all of these evaluations have in common? The experiments were based on operational tasks. The tasks were generally aimed at recognition, not on procedures following recognition.* From this generalization, we can develop a generalized approach to situation awareness studies.

During the development of flight displays, mission task elements (MTEs) should be flown first. MTEs are stylized maneuvers that are similar to maneuvers used during actual missions. MTEs include instrument tasks, such as vertical S's, Yankee patterns, and instrument tracking, and visual tasks such as air-to-ground visual tracking. These maneuvers are designed to ensure that the various building blocks of maneuvers used during operations fit with the display and aircraft characteristics. MTEs are primarily concerned with aircraft/display dynamics.

Mission related tasks put these building blocks together to form sequences of tasks that will be flown in an operational setting. This increases the cognitive demands on the pilot. The assessment of these increased cognitive demands is a major objective of these evaluations.†

Generally, these tasks will be developed from the intended mission for the aircraft/display being tested. It is not possible to list all possible mission related tasks for all aircraft.

During the course of these evaluations, embedded tasks may be included at unexpected (to the evaluation pilot) times. These embedded tasks will include engine and system failures, unplanned diversions and threats. Sufficient numbers of such embedded tasks should be included to allow for some statistical validity. The need for statistical significance, however, must be weighed against creating an artificial number of "rare events" with the effect of reducing the realism of the environment.

It is important to develop test objectives reflecting one of the Human Error Taxonomies (*see above*). For example, if the SA Taxonomy is used in developing the display model, the dis-

---

* Most UA recoveries and some system failures (Singer and Dekker, 2000) experiments are exceptions to this generalization

† To distinguish between MTEs and mission tasks, consider instrument approaches. Flying a simple instrument approach from intercept to tracking down to minimums would be considered an MTE. This task is primarily concerned with aircraft/display interactions. If the task includes self-navigation to the final approach and requires a cognitive decision to land or go-around, then it becomes a mission task.

play might be designed to assist with Level 1 errors (failure to perceive). In this case, the experiment would examine the ability of the pilot to detect, for example, external threats. The experiment would concentrate on detection, less on decisions following detection.

If, on the other hand, the display were designed to assist with Level 2 errors (failure to assess), the experiment would concentrate on the decision following detection. Similarly, displays addressing Level 3 errors would concentrate on the actions following the detection/decision.

Objective test metrics should be developed. Such metrics could include reaction time, percent correct decisions, etc. Sometimes, only quasi-objective data can be obtained, such as Cooper-Harper (1969) ratings or SAGAT responses (Endsley, 1988b).

Sample Experimental Scenarios

Two sample scenarios were obtained from NASA Ames, and additional scenarios were obtained from published sources or experiments known to the author. The following discussions do not deal with equipment suitability or experimental design, but concentrate on the "appropriateness" of the scenarios.

We will use Endsley's (1988a) situation awareness (SA) taxonomy for consistency. Either Rasmussen's (1982) or Reason's (1990) taxonomies could be used, although they are probably more appropriate for studying the ability of pilots to use displays to recover from situations.

Commercial CFIT Scenario

This scenario (Newman, 2002) involves a non-scheduled flight from Montreal to a fictitious Caribbean island. The scenario is suggestive of the Korean Air Lines accident at Guam (NTSB, 2000). The scenario description suggests that a number of system failures can be introduced to ensure a CFIT situation.

The proposed scenario is certainly appropriate and would be effective in evaluating enhanced/synthetic vision (E/SV) and other systems designed to prevent CFIT accidents. The crew workload is kept high by a combination of time pressure and poor weather. Fatigue is likely to be a factor. However, it is not clear how crew fatigue can be kept high without requiring an excessive non-productive simulator flight time.

Such a scenario can only be used once since surprise is a major factor. Further, to be completely realistic, the approach should be flown once to replicate non-familiarity with the approach and terrain.

The proposed scenario appears to be suitable for CFIT related studies. However, we would recommend that flying the entire flight would be an inefficient use of resources (simulator, experimenters, and participants). We would recommend using the terminal area operations only as part of a series of approaches.

Highway-in-the-Sky Scenario - I

In this scenario, reported by Williams (2000), the subject pilots flew a typical closed traffic pattern using one of several advanced displays, including a highway-in-the-sky (HITS). During three of six patterns, airplanes flew within the field-of-view (FOV) and the subject pilot was instructed to make a radio call identifying the aircraft. These aircraft appeared during course intercept (high workload) and during straight course tracking (low workload).

The runs were terminated in midair by blanking the displays. At the end of two of six patterns, the subject pilots were asked to estimate airspeed, heading, and altitude.

Williams states three types of SA as applying to HITS. The first is knowledge of position of the aircraft's flight path relative to the intended flight path. This would be a continuous event.

The second type of SA is the ability to detect other aircraft. This is a "rare event," albeit a rare event of relatively high probability. The third type of SA, according to Williams, is knowledge of secondary information, such as airspeed, altitude, and heading.

All three types of SA are Level 1 in scope, although the flight path information could be considered Level 2 or 3.

The experiment used latency (reaction time) as the test criteria for traffic detection. A SAGAT questionnaire was used to determine perception of aircraft state.

The scenario chosen is marginally appropriate for the study. There are two difficulties. First, the flight path was fairly simple, consisting of a standard traffic pattern, not some of the more complicated flight paths proposed by other experimenters. Second, the workload was fairly low. Including marginal weather would have increased the workload and could have been used as an additional independent variable.

Having half of the runs with "rare events" is suitable for detection of other aircraft in flight. Such a proportion would not be appropriate for normal rare events, such as runway incursions or uncharted hazards.

Highway-in-the-Sky Scenario II

This scenario was used by the NASA Synthetic Vision Program at Dallas-Fort Worth Airport in October 2000. The subject pilots flew approaches using various synthetic vision concepts.* Most of the "data" taken was in the form of preference scales. However, one approach per pilot involved a sidestep maneuver from one runway to another. The subject pilot was briefed on this maneuver during the pre-experiment briefing.

This particular maneuver was intended to evaluate the ability of the display to provide information to aid a pilot during a sidestep -- Level 3 SA according to Endsley (1988a). The question of surprise or detection was not an issue.

In this regard, the sidestep scenario was an appropriate maneuver being operationally significant and well fitted to the experimental objective.

Multiple Systems Failures

Singer and Dekker (2000) studied multiple failures presented to subject pilots during high workload instrument approaches. Their objective was to determine the ability of the displays tested to assist the pilots with diagnosis and corrective actions. The scenario was an instrument approach in a part-task simulator. During each approach a systems failure was presented and the subject pilot tasked with taking appropriate corrective action while maintaining a stabilized instrument approach.

The objectives match with Endsley's (1988a) Level 3 SA -- the ability to project to future times. As such, the experiment is less concerned with detection of the failure than with the corrective action. The scenario chosen was entirely appropriate for this study and maximized the data taken with respect to the equipment and personnel utilization.

It might be argued that providing one approach with no failure would provide a baseline. This, however, assumes that one of the results of the experiment would be ILS tracking performance with multiple failures. This was not a stated objective, nor is it useful data.

The criterion used was measuring the latency (reaction time) following the failure. In addition, Singer and Dekker verified the accuracy of the pilots' assessments using a questionnaire. The overall scenario was realistic and appropriate to the test objective.

Recommendations

Several recommendations for creating experimental scenarios for display and pilot-vehicle interface studies can be made.

Develop Operational Scenarios: First, it is important to develop operationally relevant experimental scenarios. These scenarios should reflect the intended operational use of the equipment under consideration. For novel navigation and SA displays, such as highway/tunnel-in-the-sky displays, the scenarios will likely be some navigation route involving convoluted routes similar to those described above, although not necessarily at low altitudes.

Model Human Error: The likely human-involved errors must be modeled and will, again, depend on the specific application. Endsley's (1998b) or Rasmussen's (1982) taxonomies are examples of how to characterize the issues. For example, Endsley lists perception, comprehen-

---

* R L Newman, Unpublished debriefing notes from NASA Synthetic Vision Tests at DFW Airport, October 2000

sion, and prediction as the key elements. Rasmussen lists detection, diagnosis, goal setting, strategy, procedure, and action as the elements. The experimental design must reflect the issues taxonomy chosen by the experimenter.

Develop Test Objectives: The test objectives should follow directly from the human error model used coupled with the intended use.

Develop Objective Test Criteria: It is important to set objective metrics to the extent possible, such as reaction time, accuracy of decisions, accuracy of tracking, etc.

References

ACRC (1997), Controlled Flight Into Terrain: American Airlines Flight 965, Boeing 757-223, Near Cali, Columbia, December 20, 1995, Aeronautica Civil of the Republic of Columbia, Bogota

M Anderson and R L Newman (1994), "HUD Flight Testing: Lessons Learned," Presented at Southeast Section Symposium, Society of Experimental Test Pilots, Stone Mountain, Georgia, May 1994

A J Benson (1965), "Spatial Disorientation in Flight," in Aviation Physiology, Gillies, JA (ed.), London: Pergamon Press, pp. 1086-1129

R O Besco (1996), "What All Good Pilots Do: Does Anyone Know Or Care?", Business and Commercial Aviation, November 1996, pp. 108-114

S Coyle and W Krolak (1999), Common Sense Navigation for the 21st Century, Transport Canada White Paper

M R Endsley (1988a), "Design and Evaluation for Situation Enhancement," Proceedings of the Human Factors Society 32nd Annual Meeting, pp. 97-101

M R Endsley (1988b), "Situation Awareness Global Assessment Technique (SAGAT)," Proceedings of the National Aerospace and Electrical Conference, pp. 789-795

E Fischer, R F Haines, and T A Price (1980), Cognitive Issues in Head-Up Displays, National Aeronautics and Space Administration TP-1711

J J Gallimore, F R Patterson, N G Brannon, and J P Nalepka (2000), "The Opto-Kinetic Cervical Reflex During Formation Flight," Aviation, Space, and Environmental Medicine: 71, 812-821

E F Hofer, R J Braune, G P Boucek, and T A Pfaff (2000), Attention Switching Between Near and Far Domains: An Exploratory Study of Pilots' Attention Switching with Head-Up and Head-Down Tactical Displays in Simulated Flight Displays, Boeing Report D6-36668

D G Jones and M R Endsley (1996), "Sources of Situation Awareness Errors in Aviation," Aviation, Space, and Environmental Medicine: 67, 507-512

D McKay, M Guirguis, R Zhang, and R L Newman (2002), Evaluation of EVS for Approach Hazard Detection, Presented at NATO RTO Systems Concepts and Integration and Sensors Electronics and Technology Symposium, Enhanced and Synthetic Vision Systems, Ottawa, Canada, September 2002

J M Naish (1964), "Combination of Information in Superimposed Visual Fields," Nature: 202, 641-646

R L Newman and K W Greeley (2001), Cockpit Displays: Test and Evaluation, Aldershot, UK: Avebury

R L Newman and D L Quam (1982), "Pilot Reaction to Attitude Gyro Failure," Canadian Aeronautics and Space Journal: 28, 303-310

R L Newman (1977), CH-3E (MARS) Head-Up Display Evaluation, Crew Systems TR-77-2, 1977

NTSB (1975), Aircraft Accident Report: NWA B-727-251, N274US, Near Thiells, New York, December 1, 1974, National Transportation Safety Board AAR-75-13

NTSB (2000), Aircraft Accident Report: Controlled Flight Into Terrain, KAL 801, Boeing 747-300, HL7468, Nimitz Hill, Gaum, August 6, 1997, National Transportation Safety Board AAR-00-01

W C Ocker and C J Crane (1932), Blind Flight in Theory and Practice, San Antonio: Naylor Publishing

P R Smith and J W Chappelow (1995), "Pilot Intervention Times," in Human Factors in Aviation Operations, R Fuller, N Johnston, and N McDonald, N (eds.), Aldershot, UK: Avebury

J Rasmussen (1982), "Human Errors: A Taxonomy for Describing Human Malfunctions in Industrial Installations," Journal of Occupational Accidents: 4, 311-335

R E Ross (1976), Head-Up Display Evaluation, 6594th Test Group LTR-76-8

G Singer and S W A Dekker (2000), "Pilot Performance During Multiple Failures: An Empirical Study of Different Warning Systems," Transportation Human Factors: 2, 63-76

M P Snow and J M Reising (1999), "Effect of Pathway-in-the-Sky and Synthetic Terrain Imagery on Situation Awareness in a Simulated Low-Level Ingress Scenario," Proceedings 4th Annual Symposium on Situational Awareness in the Tactical Air Environment, Piney Point, pp. 198-207

K W Williams (2000), Impact of Aviation Highway-in-the-Sky Displays on Pilot Situation Awareness, Federal Aviation Administration DOT/FAA/AM-00/31

Table I: Situation Awareness Error Taxonomy

Error Type	Description
Level 1 Failure to Correctly Perceive Element	
Data not available	Data are not available due to failure of the system design
Data hard to detect	Examples: inadequate lighting or resolution
Failure to observe data	Data not perceived due to omission, attentional limitations, distractions, etc.
Misinterpretation	Data misperceived because of prior expectations or misunderstood because of distraction
Cognitive failures	Caused by high workload
Level 2 Failure to Comprehend Elements	
Poor mental model	Poor mental model does not allow combining for information needed to meet goals
Incorrect mental model	Leads to incorrect assessment
Over reliance on default model	Routine expectation of the system is assumed
Other	
Level 3 Failure to Project Future State	
Poor mental model	Poor mental model does not allow for projection into future state
Over projection of current trends	Projection further into future than current data warrants
Other	

Table II: Internal Human Error Taxonomy

Error Type	Description
Information error	Did not detect cues
Diagnostic error	Did not accurately diagnose system state
Goal error	Did not choose reasonable goal
Strategy selection error	Did not choose appropriate strategy
Procedure error	Did not select procedures
Action error	Did not execute procedures as intended