PILOT DECISION MAKING STYLES: DEFINING THE UNSAFE PILOT

Richard J. Adams
School of Aeronautics
Florida Institute of Technology
Melbourne, Florida

ABSTRACT

This paper describes a cognitive psychology based program designed to improve aviation safety. The program was founded on over 20 years of research, development, and experimental validation in the area of pilot decision-making. It was determined in the late 1970s that 51.6% of the fatal general aviation accidents were due to faulty decision-making. This lead to an effort to determine if decision-making could be taught, if so, how it could be taught, and then, how its impact could be assessed. The purpose of the current research was to characterize the decision-making styles of accident free and accident-prone pilots. The psychological basis of the Decision-Making Styles (DMS) instrument is discussed, including evidence for a variety of host factors that influence decision-making.

BACKGROUND

The paper is an extension of decision-making research and decision-making training development sponsored by the Federal Aviation Administration (FAA) since 1973. The following sections provide the background that lead to this research, DMS instrument development, data analysis, exploratory results and musings about potential conclusions.

National Transportation Safety Board (NTSB) statistics for 1994 to 1999 show that total general aviation accidents decreased 4.8% from the period 1994-1996 to the period 1997-1999 (FAA Statistical Handbook, 1999). During the same two time periods, fatal accidents decreased 3.5%. Three year averages were used to mitigate year to year variations. Historically, the broad cause/factor of ‘pilot error’ accounted for the majority of all aviation accidents and approximately half of the fatalities (Alkov, 1991, Adams and Thompson, 1986, and Jensen and Benel, 1977). Eighty percent of reports to the Aviation Safety Reporting System (ASRS) during the 1970s and 1980s included human errors and 50% were flight crew errors (Billings and Reynard, 1984). The ‘human factors’ that lead to pilot error included how people interacted with their environments. In aviation, it has included the study of how pilot performance is influenced by the design of cockpits, avionics software, displays, controls, the functions of the organs of the body, the effects of emotions, and the interaction with the other participants in the aviation environment. These included other crew members, maintenance, dispatch, and air traffic control personnel (Jensen and Trollip, 1991). At the same time, NTSB studies found that “human performance” factors accounted for 47% of all accidents and for 50% of the fatalities in 1988. Human performance deals with measuring the limitations of human abilities in terms of information processing, decision-making, workload, attention sharing, and response generation (Wickens, 1992). The ubiquitous human error was a causal factor in 78% of all accidents and 22% of fatal general aviation accidents in 1992 (NTSB, 1991, 1994). These percentages have remained the same even considering the improvement in aircraft, air traffic control services, avionics and pilot training.

The FAA recognized the vagueness of the term “pilot error” in the early 1970’s and funded research the underlying causes behind this significant percentage of accidents. The seminal study by Jensen and Benel (1977) is the result of this research. They found that the majority of the general aviation, non-fatal accidents (56.3%) were the result of faulty perceptual-motor behavior. In contrast, a majority of the fatal pilot error accidents (51.6%) were the result of faulty decision-making. It was quite apparent from these statistics that these two areas required additional attention in the training and testing of pilots. The large percent of fatal accidents due to faulty decision making lead to the initiation of the FAA’s Aeronautical Decision Making (ADM) training program. This Decision Making Styles DMS) research has been a part of that program since 1994. One result of this research is the development of a proposed ADM training instrument, The DMS self-test.

The DMS self-test relies on new measures of expert decision making styles (Ericsson and Charness, 1994; Adams and Ericsson, 1992; Ericsson and Smith, 1991; Chi, Glaser and Farr, 1988). These include physical and mental conditions, type of knowledge used in decision-making, type of information processing, familiarity with the situation, time pressures, gut reactions or "inner signals", and flight experience information. These new scales can be used to try to identify pilot Decision Making Styles that could lead to accidents.

DMS INSTRUMENT DEVELOPMENT

The further development of the DMS instrument is the focus of this study. Psychological factors were emphasized in the DMS instrument due to the previously discussed, large contribution of decision-making errors to fatal accidents. There is extensive evidence for a variety of host factors' influence on decision-making, but the expected relationships are usually small (Steier & Mitchell, 1996, Singleton & Hovden, 1987, Hogarth, 1987, and Dreyfus & Dreyfus, 1986). For example, pilot's physical and mental condition such as vision, fatigue, stress, current or recent illness, medication, and alcohol have been related to impaired judgment and decision-making (Hartley & Hassani, 1994, Alkov, et al., 1985, Ross & Ross, 1991). Similarly, demographics, such as age, gender, marital status, education, flight certificate(s) held, total flight time, recent flight time and flying in certain geographic regions, e.g., Alaska, Hawaii, New York, show significant but sometimes complex relationships to having an accident or near accident (Wickens, et al., 1987, Lubner, et al., 1991).

The DMS instrument attempts to measure how people, who usually arrive at the correct solutions, actually solve problems and make decisions based on cognitive strategies, experience, type of knowledge, procedures, attitudes, "gut reactions" (or more formally, somatic markers) and possibly emotions (Damasio, 1994; Goleman, 1995). There are also occasions when pilots make erroneous decisions due to exogenous variables. They may rely on biased cognitive reasoning strategies due to: mis-application of a procedure or heuristic; external or environmental factors such as time pressures, high workload, poor communications, or deteriorating weather; and, internal factors such as stress, fatigue, and health already discussed. Finally, erroneous decisions can result from lack of appropriate skills, e.g., no instrument training when flying marginal VFR, lack of recent experience, i.e., currency, lack of familiarity with the aircraft being flown, and/or insufficient knowledge to solve an ill-defined problem never before encountered. Many of these negative factors were queried in the DMS instrument. The usefulness and popularity of the original decision-making training manuals lead to the retention of the critical incident format for a major portion of the DMS instrument. The instrument attempts to understand pilot decision-making styles by examining the individual's experience on mental, emotional and physical levels, all within an aviation context.

DMS Self-test Content

The psychometric or test construction procedures for the development of the DMS followed established principles for survey research methods (Fowler, 1993) and rating scale construction (Aiken, 1996). The critical incident format was used to measure cognitive and non-cognitive variables indicative of the decision-making processes. Pilots were asked to recall an aviation experience or hazardous situation that they personally experienced and then answer questions about their decision-making during that experience. The goal of this approach was to have the pilots respond realistically, that is, make the same decisions they would make in the actual situation. There is a rich methodological literature on the reporting of factual material (Fowler, 1993). Reporting has been compared against records in a variety of areas, in particular the reporting of economic and health events (Cannell, Marquis, & Laurent, 1977 in Fowler, 1993). Respondents in the Cannell, et al. study answered many of the questions realistically. For example, more than 90% of overnight stays in hospitals within a six month period were reported (Cannell & Fowler, 1965 in Fowler, 1993). How realistically people reported depended on both what they were being asked and how it was asked.

Respondents were given detailed instructions on criteria to follow in selecting their critical situation as well as on how to continue if they never had such a hazardous situation. This technique assumes that their answers indicate how they would respond in general. The following list summarizes some of the DMS variables. The complete list will be published in the final report.

1. Flying skills, training, and experiences
   a. number of precautionary of forced landings on and off of an airport
   b. number of times an engine had quit because of an improper pump or fuel tank selection
   c. number of inadvertent aircraft stalls
   d. number of inadvertent IMC encounters without an instrument rating
   e. number of FAA sponsored safety seminars attended in the last 12 months

   f. number of hours of in-flight training received from a certified instructor in the last 12 months.

2. Physical and mental condition
   a. upset stomach
   b. tired, fatigued
   c. anxious or worried
   d. affected by recent stressful events

3. Type of knowledge (declarative vs. procedural)
   a. basic pilot training
   b. advanced pilot training
   c. preparation for an FAA examination
   d. aviation safety articles, videos, of accident reports
   e. hangar flying

4. Type of processing (classic analytical vs. intuition)
   a. reviewed several possible diagnoses
   b. reviewed several possible solutions
   c. used heuristics or other memory aids to guide my decisions
   d. decided to implement a procedure, without debating other possibilities
   e. acted on my gut reactions
   f. behaved almost automatically and acted without much consciousness or awareness

5. Familiarity with the situation
   a. familiar due to past training
   b. familiar due to past personal experiences
   c. familiar with the terrain, weather and aircraft
   d. familiar because of discussions with aviation colleagues and friends

6. Time pressure
   a. had enough time to make my decisions
   b. decisions were affected by time pressures
   c. some reactions were reflexes more than considered decisions
   d. felt as if time had slowed down

7. Inner signals (reactions, thoughts, and feelings)
   a. confident and sure of myself
   b. frustrated
   c. worried
   d. sudden flash of insight
   e. trouble concentrating
   f. fearful
   g. peaceful and calm while making decisions

8. Flight experience
   a. total number of years actively flown
   b. total number of flying hours logged during all the years actively flying
   c. total flying hours logged each year during the past four years
   d. highest pilot certificate held
9. Demographics
   a. marital status
   b. highest grade or year in school completed
   c. number of friends who are pilots

METHOD

This section presents a review of the methods used for analyzing the DMS instrument data. Selection of participants for the mailing, their classification as cases or controls, and the mailing procedures are described. Mailing procedures to ensure a high response rate are outlined, and data collection procedures are summarized. The proposed measures and methods for data analysis are described. These include univariate, bivariate, and multivariate analyses. Due to the predominance of Lickert scale type of data, the primary analyses will focus on non-parametric Discriminant Analysis.

Participants

A total of 2000 cases and 2000 controls were selected for the initial mailing of 4000 DMS self-tests. Case and control respondents were randomly selected from NTSB and FAA data bases respectively. Cases were classified as currently active US pilots who had an accident, incident or near accident in the year prior to the survey. Controls were defined as currently active pilots who had no accident, incident or near accident in the past one to five years.

To achieve and maintain active pilot status on the FAA Airmen's Registry, pilots are required to have recently passed standard FAA medical exams. Airmen include pilots and non-pilots such as control tower operators. Only pilots from the Airmen's Registry were studied in this survey. Depending on the type of Airmen's certificate held, medical examinations must be passed every six months to two years. Shorter time between examinations usually indicates a higher level of pilot certificate.

The available FAA Accident/Incident data bases contained lists of pilots who had an accident, incident or near accident in the five years 1994-1998. Events recorded on the data bases have been investigated by at least an FAA inspector and possibly by other agencies including the National Transportation Safety Board (NTSB). Cases were selected from this data.

Controls were selected from the part of the FAA's Airmen's Registry of all US Airmen which did not appear in the Accident/Incident data base. Therefore, controls were active pilots who had no reported accident, incident or near accident in at least the year prior to the survey.

Data Collection

The DMS questions, format and length were completed as a part of the FAA sponsored research after completion of the subject matter expert interviews, the convenience sample, and the pre-test. The analysis of the DMS data is being completed as a part of doctoral degree requirements at the University of Central Florida.

A multi-faceted data collection procedure was utilized (Fowler, 1993). First, a letter was sent to respondents to notify and inform them of the forthcoming case-control survey. Next, the FAA's Office of Aviation Medicine mailed the questionnaires, together with a cover letter re-introducing the survey and guaranteeing anonymity. About 10-14 days after the first mailing, the non-respondents were sent a reminder postcard emphasizing the importance of the survey and of a high response rate. About 10-14 days after the post card reminder, non-respondents were sent a second letter again emphasizing the importance of a high rate of return and including another DMS self-test for those who had discarded the first one.

Data Analysis

SPSS version 10.0 (1999) was used for the entire data analyses. Study variables were tested using univariate, bivariate and multivariate techniques.

Univariate Analysis

Each respondents' entire questionnaires is reviewed, checking for proper completion of the response sheets, missing data or incomplete data, and reading the accident or incident selected by the respondent for self-analysis of decision-making styles.

The second step in analysis of the responses was to run descriptive statistics to identify missing data and outliers. Frequency distributions were checked for outliers, missing values, ranges, and normal distributions. The minimum, maximum, median, mean and standard deviation were examined for all risk taking, personality factor, decision making styles, accident, incident and near accident, flight experience, highest rating held and demographic data.

Based upon the responses and the descriptive statistics, out of range values or unreasonable values were identified as follows:

1. If more than 50 responses are missing for a questionnaire, it was not used.
2. Next, questionnaires with less than 50 missing responses were examined, for significance with respect to impact on the dependent variables (number of accidents, incidents, and near accidents) and on the impact of data for flying skills, training and experiences, e.g., fuel starvation, inadvertent stall, inadvertent IMC, fatigue, novice vs. expert decision making or time pressure.
3. A general rule was used to classify as outliers, any variable values that exceed 3 standard deviations
4. Individual question responses were analyzed to determine the applicability of this rule especially for the primary dependent variables of accidents, incidents, and near accidents.

Upon completion of the determining and disposition of outliers, the data (all independent variables) were analyzed using frequency data and histograms to determine if the distributions were normal, and, if not, the values of their skewness and kurtosis. All normally distributed variables were used in their "raw" data collection state. All skewed variables were considered candidates for some type of transformation. Finally, minimum and maximum values of all screened variables were checked to verify that they were within the specified range.

Bivariate Analysis

Since the majority of the data collected consisted of Lickert scale data (i.e., ranking an individual's responses to a series of questions from strongly agree (0) to strongly disagree (4) or from not at all (0) to a very great extent (4), nonparametric (or distribution free) statistics were the primary analytical techniques used. Cross tabulations and measures of association were used as the next level of data screening. Variable count, expected count, percent within each variable score, and percent of total scores were used to assess the strength and weakness of within variable scores as well as between categorical variables. For example, in the flying skills, training, and experiences, the cross tab values of expected count and % of total were used to validate the conclusions based upon means for cases and controls for such survey items as "times you have inadvertently stalled an aircraft". At this stage, Chi-Square ratios were used to assess the expected vs. obtained values for each variable.

Multivariate Analysis

The primary data analysis technique used was Discriminant Analysis. This technique provides a linear combination of variables (or Discriminant Function) that looks like the right side of a multiple regression equation because it sums the products of variables multiplied by coefficients. The technique estimates the coefficients, and the resulting function can be used to classify new cases. Combining information from two or more variables can greatly enhance the separation of groups.

In addition to the Discriminant Function, this analysis technique was used to explore or describe:

- Which variables among many were most useful for discrimination among groups,
- If one set of variables performs equally as well as another,
- Which groups are most alike
- Which cases are outliers (differ markedly from others within their groups).

When there are more than two groups, canonical variables become the focus of the analysis. The first canonical variable is the linear combination of the variables that maximizes the differences between the means of the groups in one dimension. The second canonical variable represents the maximum dispersion of the means in a direction orthogonal to the first direction, and the the third represents the dispersion in a dimension independent of the first two, etc. (much like Principal Components Analysis). Canonical variables are *factors* that discriminate optimally among the group centroids relative to the dispersion within the groups.

Finally, Discriminant Analysis provides a table of within-groups correlations of each predictor with the canonical variable. This provides another way to study the usefulness of each variable in the discriminant function.

EXPLORATORY RESULTS

A total of 1346 questionnaires were returned. Serendipitously, 653 were cases, 664 were controls (29 questionnaires had to be eliminated based on the descriptive statistics, out of range values, or unreasonable values (e.g., 70,000 flight hours). Upon completion of the univariate and bivariate analyses, exploratory data sets were compiled and run. These were:

1. Individual runs with all 136 survey questions (with and without missing data replaced by means),
2. Single groups of variables (see below) combined with select individual variables ,
3. Fourteen groups of variables without any individual variables, and,
4. Fourteen groups of variables with select individual variables.

Table 1 Fourteen Variable Groups

DMS Groups	Variable Name
Thrill and Adventure Seeking	TASGRP
Cattell Personality Factors (6)	CATPERF
Flying skills, training & experience	FLYEXP
Physical & Mental condition	PHYCOND
Type of knowledge (during accident)	KNOWL
Type of information processing	DECMAK
Familiarity with the situation	SITAWAR
Time pressure	TIME
Inner signals	INNERSIG

The results of running these 14 groups with and without select individual variables to augment them found a 62.9% - 71.3% correct classification of cases and controls. Since a simple coin flip would provide a 50% chance correct classification, these preliminary, exploratory results are encouraging. Frequently occurring canonical variables were: TASGRP, FLYEXP, Gut Reactions, Automaticity, TIME, Disbelief.

MUSINGS ON POTENTIAL CONCLUSIONS

Based on these early, exploratory results, it appears that the Discriminant Analysis technique can be used to differentiate between accident prone and non-accident prone pilots. The rate of prediction is somewhat better than chance, about 70%. Addidtional work needs to be done to try to achieve an 80% or greater correct classification of cases and controls. This should involve transforming the data by reflecting the scales (were necessary), rechecking the skewed variables and transforming (logarithmically) those variables that are highly skewed, defining variable groups using all of the questions within each group category, and running the Discriminant Analysis with these new groups.

REFERENCES

Adams, R. J. & Ericsson, K. A. (1991). Introduction to cognitive process of expert pilots. (Federal Aviation Administration Contract no. DTFA01-90-C-00042). Washington, D.C.

Adams, R. J. & Thompson, J. L. (1986). Aeronautical decision making for helicopter pilots (Federal Aviation Administration Rep. No. DOT/FAA/PM-86/45). Washington, D.C.

Aiken, L. R. (1996). Rating Scales and Checklists: Evaluating behavior, personality, and attitudes. New York: John Wiley & Sons, Inc.

Alkov, R. A. (1991). U.S. Navy aircrew Coordination Training. Sixth International Symposium on Aviation Psychology. Columbus, OH.

Alkov, R. A., Gaynor, J. A. & Borowsky, M. S. (1985). Pilot error as a symptom of inadequate stress coping. Aviation Space environmental Medicine. 56, 244-7.

Billings, C.E. and Reynard, J. D. (1984). Human factors in aircraft incidents: Results of a seven year study. Aviation Space Environmental Medicine. 55: 960-965.

Institute for Personality and Ability Testing.

Chi, M. T. H., Glaser, R. and Farr, M. J. (Eds.) (1988). The nature of expertise. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Damasio, A. R. (1994). Descartes' error: Emotion, reason, and the Human Brain. New York: Avon Books.

Dreyfus, H. L.,& Dreyfus, S. E. (1986). Mind over machine. New York: The Free Press.

Ericsson, K. A. and Charness, N. (1994). Expert performance its structure and acquisition. American Psychologist. 49(8), 725-747.

Ericsson, K. A. and Smith, J. (Eds.) (1991). Toward a general theory of expertise: prospects and limits. New York: Cambridge University Press.

Federal Aviation Administration (1999). Statistical Hanbook of Aviation. Washington, DC.

Fowler, F. J. (1993). Survey Research Methods. Newbury Park, London: Sage Publications.

Goleman, D. (1995). Emotional intelligence: Why it can matter more than IQ. New York: Bantam Books.

Hartley, L. R. & El Hassani, J. (1994). Stress, violations and accidents. Applied Ergonomics. 25(4), 221-230.

Hogarth, R. M. (1987). Judgement and choice: the psychology of decision. New York: John Wiley & Sons Ltd.

Jensen, R. S. & Benel, R. A. (1977). Judgment evaluation and instruction in civil pilot training (Federal Aviation Administration Rep. No. FAA-RD-78-24). Washington, D.C.

Jensen, R. S. & Trollip, S. R. (1991). Human Factors for General Aviation. Englewood, Co: Jeppessen Sanderson, Inc. p. 1-2.

Lubner, M. E., Mardowitz, J. S., & Isherwood, D. A. (1991). Rates and risk factors for accidents and incidents versus Violations for U.S. airmen. International Journal of Aviation Psychology. 1(3), 231-243.

National Transportation Safety Board. (1991). Annual Review of Aircraft Accident data: U.S. General Aviation Calendar Year 1988. Washington, DC: Report No. NTSB/ARG-91/01. US Government Accession No. PB91-156851.

National Transportation Safety Board. (1994). Annual Review of Aircraft Accident data: U.S. General Aviation Calendar Year 1992. Washington, DC: Report No. NTSB/ARG-94/02. US Government Accession No. PB94-181054.

Singleton, W. T., & Hovden, J. (Eds.). (1987). Risk and Decisions. New York: John Wiley & Sons Ltd.

SPSS 10.0 (Computer software). (1999). Chicago, IL: SPSS Inc.

Steier, D. M., & Mitchell, T. M. (Eds.). (1996). Mind matters: a tribute to Allen Newel. Mahwah, NJ: Lawrence Earlbaum Associates, Inc.

Wickens, C. D. (1992). Engineering Psychology and Human Performance, 2nd Ed. New York: HarperCollins Publishers Inc.

Zuckerman, M. (1994). Behavioral expressions and biosocial bases of sensation seeking. New York: Cambridge University Press.