NASA/TM—2003–209612

# Do You See What I See? Effects of Crew Position on Interpretation of Flight Problems

*Ute Fischer, Judith Orasanu*
*Ames Research Center, Moffett Field, California*

---

April 2003

# The NASA STI Program Office ... in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA Scientific and Technical Information (STI) Program Office plays a key part in helping NASA maintain this important role.

The NASA STI Program Office is operated by Langley Research Center, the lead center for NASA's scientific and technical information. The NASA STI Program Office provides access to the NASA STI Database, the largest collection of aeronautical and space science STI in the world. The Program Office is also NASA's institutional mechanism for disseminating the results of its research and development activities. These results are published by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers, but having less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services that complement the STI Program Office's diverse offerings include creating custom thesauri, building customized databases, organizing and publishing research results ... even providing videos.

For more information about the NASA STI Program Office, see the following:

- Access the NASA STI Program Home Page at ***http://www.sti.nasa.gov***

- E-mail your question via the Internet to help@sti.nasa.gov

- Fax your question to the NASA STI Help Desk at (301) 621-0134

- Telephone the NASA STI Help Desk at (301) 621-0390

- Write to:

NASA STI Help Desk
NASA Center for AeroSpace Information
7121 Standard Drive
Hanover, MD 21076-1320

NASA/TM—2003–209612

# Do You See What I See? Effects of Crew Position on Interpretation of Flight Problems

*Ute Fischer*
*Georgia Institute of Technology, Atlanta, Georgia*

*Judith Orasanu*
*Ames Research Center, Moffett Field, California*

National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California 94035

April 2003

# Acknowledgments

Ute Fischer, Program in Cognitive Science, School of Literature, Communication, and Culture, Georgia Institute of Technology; Judith Orasanu, NASA Ames Research Center.

Part of this work was presented at the Annual Meeting of the Society for Judgment and Decision Making in St. Louis, November 1994, and the 8th International Symposium on Aviation Psychology, Columbus, OH, April 1995. The research described in this paper was made possible through the support of NASA-Code UL and the FAA-ARD-210. The opinions expressed in this paper are the authors' and should not be construed as official NASA or FAA policy. We wish to thank the participating airlines, especially the pilots, who generously shared their time as subjects. We are also indebted to Don Bryant who helped us as subject matter expert throughout the project, to Jerry Jones, Corwin Logsdon, Mietek Steglinski, and Barbara Sweet who volunteered as preliminary subjects, to Michael Wich who helped with the data collection and analysis, and to Kathleen Mosier and to Elizabeth Veinott who commented on earlier versions of this paper.

Available from:

NASA Center for AeroSpace Information
7121 Standard Drive
Hanover, MD 21076-1320
301-621-0390

National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
703-605-6000

## Summary

The set of experiments reported in this paper is concerned with whether pilots who differ in levels of experience and job-specific roles also differ in their interpretations of situations that demand judgments about flight-related problems. In addition to this rather applied question, these studies also address more fundamental issues concerning differentiation of expertise within a rich domain. In a series of sorting studies involving aviation incidents, we compared the judgments of captains, first officers and second officers. Multidimensional scaling analyses indicated that these three groups framed aviation incidents differently. In making a decision, pilots focused on those aspects of the situation that were consistent with responsibilities associated with their particular crew roles. Moreover, differences in feature selection were found to reflect the effect of crew role, not years of aviation experience or absence of relevant knowledge.

## Introduction

Consider the meeting of three strangers, pilots aged 30, 40, and 50 years, on the flight deck of a modern jet transport. They are, figuratively, given the keys to a $50+ million aircraft and a plan of where to fly that day. With no opportunity to train together as a team, they are expected to slip smoothly into coordinated action to fly a planeload of people safely to their destination. Common training in the domain (e.g., aircraft systems, company procedures, FAA regulations) as well as in seat-specific duties allows the individuals constituting the team to perform this complex task.

But when a problem occurs and the entire crew must contribute to its solution, the question arises as to whether all crew members will perceive the situation similarly and will entertain similar solutions, regardless of their experience and roles in the cockpit (i.e., as captain, first officer, and second officer). Because of their common domain training and relatively high levels of experience, we might expect all crew members to interpret problem situations similarly, reflecting shared mental models (Cannon-Bowers, Salas, & Converse, 1990; Orasanu, 1994)[1]. On the other hand, systematic variations in their aviation experience such as crew role may lead to qualitative or quantitative differences in knowledge that are reflected in different interpretations of problems and subsequent decisions.

The set of experiments reported in this paper is concerned with whether pilots who differ in levels of experience and job-specific roles also differ in their interpretations of situations that demand judgments about flight-related problems. In addition to this rather applied question, these studies also address more fundamental questions concerning differentiation of expertise within a rich domain.

Much of what we know about cognitive principles and mechanisms underlying expertise stems from studies comparing the behavior of experts and novices (Glaser & Chi, 1988; Cooke, 1992). Originally applied to chess masters and novices (Chase & Simon, 1973; de Groot, 1965), expert-novice comparisons were soon employed in a host of other domains, including electronics (Egan & Schwartz, 1979), physics (Chi, Feltovich, & Glaser, 1981; Larkin, McDermott, Simon, & Simon, 1980), mathematics (Schoenfeld, 1983), computer programming (Adelson, 1981; McKeithen, Reitman, Rueter, & Hirtle, 1981; Soloway, Adelson, & Ehrlich, 1988) , accounting (Bouwman, 1982), medicine (Lesgold, Rubinson, Feltovich, Glaser, Klopfer, & Wang, 1988; Patel &

Frederiksen, 1984), political science (Voss, Lawrence & Engle, 1991), and writing (Bryson, Bereiter, Scardemalia & Joram, 1991).

Expert-novice contrasts typically portray experts within a domain as a rather homogenous group that share certain key characteristics such as deep, principle-based problem analysis and representation (Glaser & Chi, 1988). Since expertise is defined with reference to novices, researchers typically select a level of analysis that highlights those characteristics of the two groups that are most distinct. Focusing on cognitive processes that are shared by experts but uncommon to novices inadvertently creates the impression that experts within a given domain invariably exhibit similar behavior. However, contrasts between experts *within* a given domain suggest that expertise may indeed be quite differentiated and characterized by systematic variability.

Recent work comparing domain experts has found differences with respect to both levels of performance (Bédard & Biggs, 1991; Ericsson, Krampe, & Tesch-Römer, 1993) and underlying cognitive processes (Patel & Groen, 1991; Patel, Groen & Arocha, 1990). For example, among expert musicians and athletes, performance has been shown to vary with amount of deliberate practice (Ericsson, Krampe, & Tesch-Römer, 1993). In contrast, auditors' and physicians' diagnostic accuracy has been found to depend on distinct aspects of professional experience. Bédard and Biggs (1991) noted that auditors with recent experience auditing manufacturing businesses or businesses with inventory were more likely to detect an error in allocating inventory costs than their colleagues who did not have this specific experience. Bédard and Chi (1993) relate this finding to the distinction between generic and specific expertise that Patel and Groen (1991) proposed for medical expertise. Patel and her collaborators (Patel & Groen, 1991; Patel, Groen & Arocha, 1990) compared the diagnostic reasoning of specialists with that of subexperts (i.e., physicians who had specialized in an area of medicine different from the specialists' field); in a second study they compared practitioners and researchers (i.e., physicians who were primarily engaged in medical research). Both specialization in medical area as well as in occupation affected expert performance. When physicians analyzed a case within their medical field, they used pure forward reasoning (from clinical data to a hypothesis) to arrive at a diagnosis. In contrast, for cases outside their specialization, they used a mix of forward-directed and backward-directed strategies. In a similar vein, practitioners were found to rely on their knowledge of clinical cases whereas researchers based their diagnoses on biomedical scientific knowledge.

Patel and Groen (1991) introduced two types of medical expertise - generic and specific - to account for these specialization effects. Generic expertise in medicine refers to knowledge and skills that individuals acquire as part of their general training as physicians. Specific expertise involves additional specialized knowledge, for instance about a particular subdiscipline in medicine. This perspective on specialized expertise within a domain breaks with the traditional view of domain expertise as a unified form of competence.

The need to distinguish between different types of expertise has previously been pointed out by Shanteau (1988). However, unlike Patel and Groen who focus on differentiation of experts within a domain, Shanteau was interested in differences between domains of expertise, and accordingly proposed to categorize domains with respect to their cognitive requirements. For example, he distinguished domains that emphasize perceptual skills from those that stress problem-solving abilities. A third type of expertise relates to perception-based expertise as noted by Shanteau (1988), but also involves a large psychomotor component. These include musical skills such as playing the violin, athletic skills such as tennis, or occupations such as cigar rolling (C.f., Ericsson, Krampe, and Tesch-Römer, 1993). Thus, what is meant by expertise may differ across domains and may show different courses of development.

It is important to note that the distinction between generic and specific expertise refers to a qualitative difference in cognitive processes that Patel et al. (Patel & Groen, 1991; Patel, Groen & Arocha, 1990) link to systematic variations in professional experience rather than to differences in amount of experience. For instance, what distinguishes an internist from a cardiologist is not years of experience per se, but the kind of experience. On the other hand, quantitative measures of experience, e.g., number of classes or years in a profession, have traditionally been used in cognitive science studies on expertise to classify subjects as experts or nonexperts and to explain differences in their cognitive processes (cf., Chi, Feltovich, & Glaser, 1981; Federico, 1995; Schmidt & Boshuizen, 1993; Schoenfeld & Herrmann, 1982). Quantitative approaches may be reasonable when comparing novices with experts, but may be inadequate when examining differentiation of expertise within a domain. Quantitative approaches may also be more appropriate to certain types of expertise than others (Shanteau, 1988). Specifically, perceptual and psychomotor skills may benefit from continued “deliberate practice” (Ericsson, et al. 1993; Newell & Rosenbloom on the power law of skill acquisition, 1985); however, it is not clear whether more cognitive types of expertise such as aviation decision making show similar patterns of continuous improvement with practice.

In the present study we sought to address both quantitative and qualitative aspects of expertise in professional airline pilots. Experience in civil aviation means not only years of pilot practice, but also has a more specific sense--experience in a particular crew position. Captains, first officers and second officers fulfill different roles and have different responsibilities during both normal and abnormal situations. Within a particular airline the two aspects of experience tend to be highly correlated. Captains are generally also the most senior pilots, followed by first officers and then second officers. Airlines, however, differ in the breath of their seniority spectrum, so that first officers in one airline may be as experienced as captains in another company. Thus, across airlines it is possible to decouple crew role and years of pilot experience.

In a series of studies, we investigated how pilots who differed in years as well as kind of aviation experience interpret aviation events that require decisions to be made. Decision making is an important aspect of aviation ranging from fairly routine during normal flight to highly complex when abnormal events occur. In keeping with recent approaches to decision making that emphasize application of domain knowledge (Flin et al., 1997; Klein, Orasanu, Calderwood, & Zsambok, 1993; Zsambok & Klein, 1996), Orasanu (1994) developed a descriptive process model of decision making by expert pilots. Like other naturalistic decision models (e.g., Klein’s Recognition Primed Decision theory, 1989, 1993), it posits two major components, situation assessment and selection of a course of action. Essential elements of situation assessment in aviation include defining the nature of the problem, determining time available, and assessing the risks. Selecting a course of action in aviation depends on the responses afforded by the specific problem: either following a specified procedure, choosing among options, or structuring a novel response to an unanticipated problem (c.f. Rasmussen, 1985). In the present study we examined whether these model elements indeed underlie pilots' classifications of decision problems and whether their salience varies by crew position or years of pilot experience.

In order to determine the features or dimensions that organize pilots' interpretations of decision events, a sorting task was used. Hierarchical clustering indicated which events were judged as similar, and multidimensional scaling helped identify decision-relevant dimensions underlying the categories (Chi et al., 1981).

At least three hypotheses are possible concerning experience effects. H1 is the null hypothesis. It states that amount and type of aviation experience will have no effect on pilots' behavior. That is, all commercial airline pilots flying for major carriers, regardless of experience level and crew role, will categorize decision events in identical ways. This hypothesis is derived from expert-novice studies that demonstrated the consistency of expert behavior. Pilots flying Part 121[2] air transports are all expert pilots who have met experience and performance requirements set by the Federal Aviation Administration. H2 states that years of professional experience will determine pilots' categorizations of problem situations. That is, although all professional pilots may be considered experts, some of them are more senior than others and should thus interpret decision events differently than less experienced pilots. This hypothesis is consistent with Ericsson, Krampe, and Tesch-Römer (1993) who found that performance varies with amount of deliberate practice, even among experts. H3 states that role on the flight deck will determine pilots' interpretations of decision events, regardless of years of aviation experience. Flight crews on modern air transports include two or three crew members with different roles and different responsibilities: captains have legal responsibility for all flight relevant decisions and manage flight safety; first officers are the co-pilots and as such are second-in-command; second officers monitor the aircraft systems and may not hold a pilot license for the specific aircraft. If crew role shapes pilot expertise, then captains, first officers, and second officers should focus on distinct aspects of problem situations. The differences, moreover, should be consistent with their respective crew functions.

Three studies were conducted to address these hypotheses. In studies 1 and 2 we examined whether captains, first officers and second officers categorize decision events similarly or whether there are systematic differences between crew positions. Study 3 was designed as a follow-up to study 1 to tease apart possible effects of level of experience and crew role on pilots' interpretations of decision events.

## STUDY 1

A free sorting task analogous to Chi et al. (1981) was employed to determine the kinds of structure imposed on decision events by captains, first officers, and second officers. A second purpose was to determine whether the features pilots use for categorization map onto components of a previously developed model of aviation decision making (Orasanu, 1994).

## Method

### Participants

Twenty-eight pilots from a major U.S. airline volunteered as subjects. Ten were captains, 10 were first officers and eight were second officers. The captains had an average of 27.2 years of experience in Part 121 aircraft (SD = 2.6 yrs.), first officers had 14.9 years (SD = 8.1 yrs.), and second officers had 4.7 years (SD = 1.5 yrs.).

Eight captains, eight first officers and seven second officers were also ex-military pilots. Military experience of the captains averaged 7.3 years (SD = 2.8 yrs.), of first officers 9.2 years (SD = 6.3 yrs.), and of second officers 8 years (SD = 3.2 yrs.).

### Material

Ninety-seven incident reports on crew decision making were retrieved from the Aviation Safety Reporting System (ASRS) data base[3]. A stratified random sample of 22 scenarios was selected to reflect phases of flight in the report set. The following distribution of scenarios was obtained: Preflight = 3; Taxi = 2;. Take-off = 4; Climb = 4; Cruise = 4; Descent = 2; Approach = 3.

The scenarios were shortened to include only the first decision event mentioned. All the information that the reporter had provided regarding flight context and precipitating event was recorded. The actual decision, however, was not revealed. For instance, subjects saw the following description: *A wide bodied aircraft weighing 43,000 lbs. experiences a #3 engine compressor stall at 138 knots on takeoff; V1 is 142 knots.* Each description was printed on an 5 X 8 inch index card. The full set of scenarios can be found in Appendix A.

### Design and Procedure

Subjects received the stack of 22 index cards with the following instruction: "Your task is to sort the 22 cards into *piles of scenarios that you think are alike with respect to the type of major decision* involved. Each pile that you create should consist of scenarios that require similar decisions. You may use a minimum of 2 piles and a maximum of 22 piles." Subjects were also told that they could sort repeatedly as long as they sorted the cards into piles of scenarios that involved similar types of decision. The instructions were identical for captains, first officers and second officers; i.e., no reference was made to sort from the point of view of a particular crew role. Subjects were tested individually and were allowed to work at their own pace. A different random ordering of the scenarios was provided to every subject and, in case of multiple sorts by the same participant, for subsequent passes.

Upon completion of the task, the sortings were recorded and subsequently handed back to subjects, one pile at a time. Subjects were then asked to explain why they had put the scenarios together in each pile; i.e., to provide a rationale for their categories. Subjects who had gone through more than one sort were asked to choose the one they preferred.

### Analyses

#### *Sorting Data*

Multidimensional Scaling (using ALSCAL) and Hierarchical Clustering (average linkage procedure) were conducted on subjects' sortings. Only data from single or preferred sorts were used.

Separate analyses were performed on the groupings of captains, first officers and second officers. Since Multidimensional Scaling and Clustering analyses are designed to reflect the psychological distance between items, they require as input data some form of distance measurement. Accordingly we converted the obtained similarity ratings into disassociation scores as has been proposed by Rosenberg, Nelson, and Vivekananthan (1968). Disassociation scores refer to the inverse of how often two scenarios, A and B, have been placed into the same pile *and* how often each of them has been put into a pile with scenario C. That is, disassociation scores reflect the degree to which A differs from B and from both B and C.

### *Verbal descriptions*

Interpretation of the resulting dimensions and clusters was aided by the verbal descriptions that the subjects provided to their sorted categories. Descriptions were assigned to the following eight categories: *Type of Problem* summarizes statements that characterize the nature of the incident. Examples are “Communication problem” or “Mechanical problem.” Statements such as “Take-off problem” or “In the air” specify when the incident occurred and were classified as *Phase of Flight*. The category *Course of Action* includes statements that mention specific responses to an incident such as “Abort take-off.” Also included were statements that described options available to a crew in a given situation (”Continue or return?”), as well as references to flight manuals (“Check the book”), other parties (“Get maintenance to help you.” “Let maintenance handle it.”), or remarks that simply stated that the crew need to seek further information. *Decision Complexity* refers to statements that evaluate how difficult it is to resolve the incident. Examples are “Cut and dried,” “Require more thought.” Statements such as “Highest attention” as well as evaluations of how important it is for the crew to react to an incident - for instance, “low priority decision” - were assigned to the category *Priority*. Statements that addressed implications of the incident for the safety of the flight (“No risk,” “Serious problem,” ”Could turn into a problem”) were summarized under the heading *Risk*. References to the existence or absence of time constraints in decision making were classified as *Time Criticality*.

Subjects frequently mentioned several of the above categories in describing common elements of the scenarios they had placed in a pile. Within a description we counted all references to different categories; multiple references to the same category were counted only once. Thus, the description “most critical, not cut and dried, got to make decision quickly” refers to three categories, namely to Risk, Decision Complexity and Time Pressure. Reference to only one category - Decision Complexity - was noted in the following description: “Decision is made. Definite things you do, easy decision.”

## Results and Discussion

### *Number of Piles*

A one-way analysis of variance on the number of piles per subject revealed no significant differences between the captains, first officers, and second officers ($F(2,25) = 1.38$; *ns*.). Pilots sorted the 22 scenarios on average into 5.9 piles.

### *Verbal Descriptions*

Two raters independently coded the verbal descriptions. Inter-rater reliability was high (Cohen's Kappa = .93[5]). Figure 1.0 indicates that captains, first officers, and second officers emphasized different types of descriptors when referring to the 22 decision events ($\chi^2(12, N = 1{,}071) = 78.91$, $p < .001$).

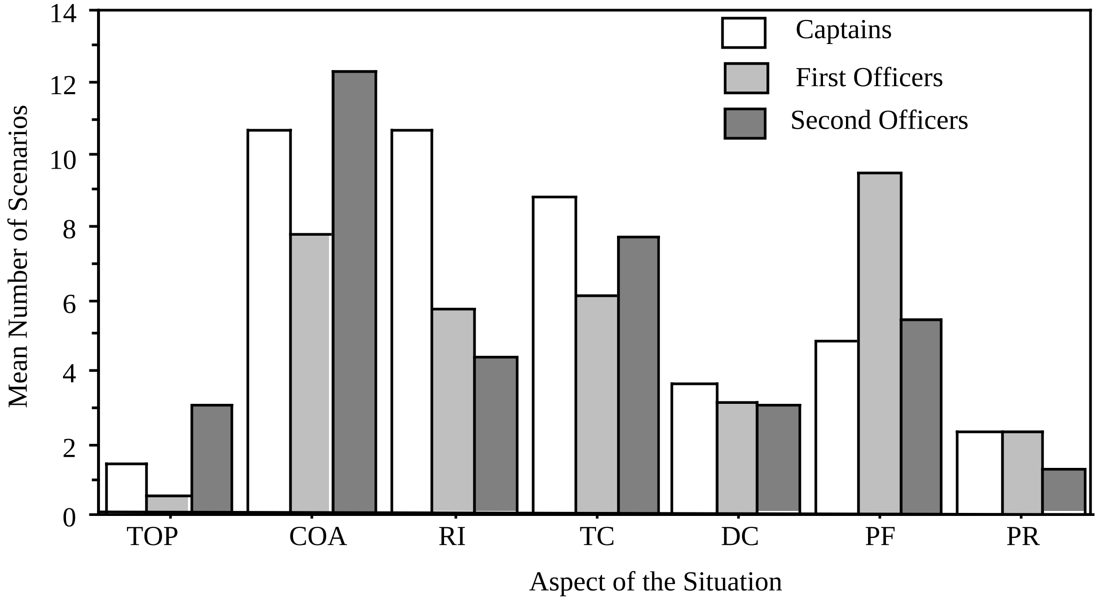

**Figure 1.0.** Mean frequencies of various situation descriptors provided by captains, first officers and second officers in categorizing aviation incidents (N = 22). Descriptors are: Type of Problem (TOP), Course of Action (COA), Risk (RI), Time Criticality (TC), Decision Complexity (DC), Phase of Flight (PF), and Priority (PR).

On average, captains stressed course of action and implications for flight safety in response to 11 of the 22 scenarios[6]. Time constraints were mentioned with reference to 9 scenarios. First officers categorized the incidents most frequently in terms of phase of flight (M = 10), followed by course of action (M = 8), time criticality (M = 6) and risk (M = 6). Important descriptors for second officers were course of action (M = 12), time criticality (M = 8) and phase of flight (M = 5).

### *Sorting Data*

Multidimensional scaling analyses for 1-, 2-, 3-, and 4-dimensional solutions were conducted on the sorting data of captains, first officers and second officers. Kruskal's Stress and the RSQ values for these analyses plotted against the number of dimensions are depicted in Figure 1.1. Both measurements reflect how well a particular solution accounts for the data. A flattening in their graphs indicates that additional dimensions do not substantially decrease the variance in the data. The elbow thus points to the maximum number of dimensions that should be used in interpreting the data (Schiffman, Reynolds, & Young, 1981).

Inspection of Figure 1.1 suggests that the 2-dimensional solution is optimal for the three data sets. Hierarchical clustering analyses were performed to confirm this inference. The optimal number of clusters was determined using Milligan and Cooper's (1985) refinement of Mojena's (1977) stopping rule. This rule yields a confidence limit for the clustering process. A fusion coefficient exceeding this value indicates the level in the hierarchy at which the clustering process combined entities that are too heterogeneous relative to the variability present in the whole set. The number of clusters at the previous level then is taken to be the optimal partitioning. When we applied this rule to

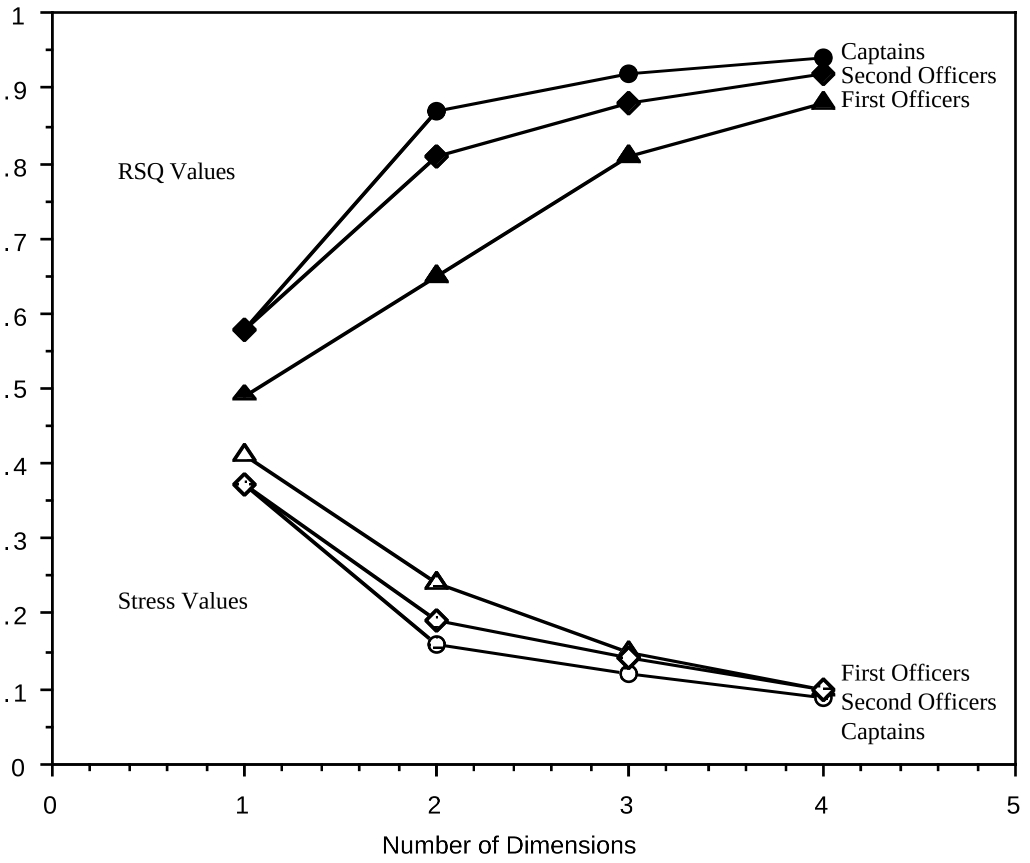

**Figure 1.1.** Kruskal's Stress and RSQ values for 1-, 2-, 3-, and 4-dimensional solutions obtained in Multidimensional Scaling Analyses of captains', first officers', and second officers' categorizations of the aviation incidents. RSQ values are given in the upper portion of the figure; Stress values in the lower portion.

our data sets, we found that the sortings of fall into three clusters. In each instance, the three clusters also mapped well onto the two dimensions derived in the multidimensional scaling analysis. Figures 1.2, 1.3 and 1.4 combine the results of the multidimensional scaling and hierarchical clustering procedures for captains, first and second officers, respectively. As can be seen, different solutions were obtained for the three groups.

***Captains' Dimensions***

For the captains, scenarios 7, 8, 13, 15, and 19 contrast with all the other scenarios along the first dimension (x-axis). All these situations take place either prior to or during taxiing. Participants classified these captains, first officers, and second officers incidents as no risk situations that involve few time constraints and that require the crew to return to the gate and to get help from ground personnel such as maintenance. The remaining scenarios fall into two clusters and contrast along the second dimension. Scenarios 2, 4, 12, 14, 16, 17, 20, and 21 describe incidents that pose an immediate threat to flight safety and leave very little time for decision making. Incidents that involve some risk and time constraints form cluster 3 and include scenarios 1, 3, 5, 6, 9, 10, 11, 18, and 22.

The three clusters thus map onto a *risk dimension* as well as a *time dimension*. Interestingly, a time factor is also implied in the evaluation of danger. The scenarios are not aligned in an increasing order ranging from no to high risk, but adhere to a temporal order instead: no risk situations, then situations involving an imminent threat to flight safety, followed by situations which might turn dangerous at some future time. This finding suggests that the captains did not conceive of risk as a quality that situations either have or do not have. Rather the threat imposed by an incident was seen to vary as time constraints on generating a response tighten or loosen. Examples of each category are:

*(#15) No risk:* An aircraft fueler advises the crew of a medium large transport that he can't get the numbers to balance between the fuel off loaded from the truck and the fuel on the aircraft. He works on it for an hour and gets nowhere.

*(#20) Immediate risk:* A wide bodied aircraft weighing 43,000 lbs. experiences a #3 engine compressor stall at 138 knots on takeoff; V1 is 142 knots.

*(#18) Potential risk:* According to its logbook, a medium large transport had not been flown for 10 days; during that time maintenance was done and a new interior put in. On an approach at BUR the right main gear "unsafe" red light comeson and stays on during gear extension with no green down and locked light. At 8-10 miles out, flight level 3000-4000', 180 knots, IAS with 5 degree flaps, the crew recycle the gear but still get the same result.

The *time dimension* reflects the captains' concern with the issue of how much time they have left for solving a problem and averting negative consequences. Scenarios along this dimension range from situations

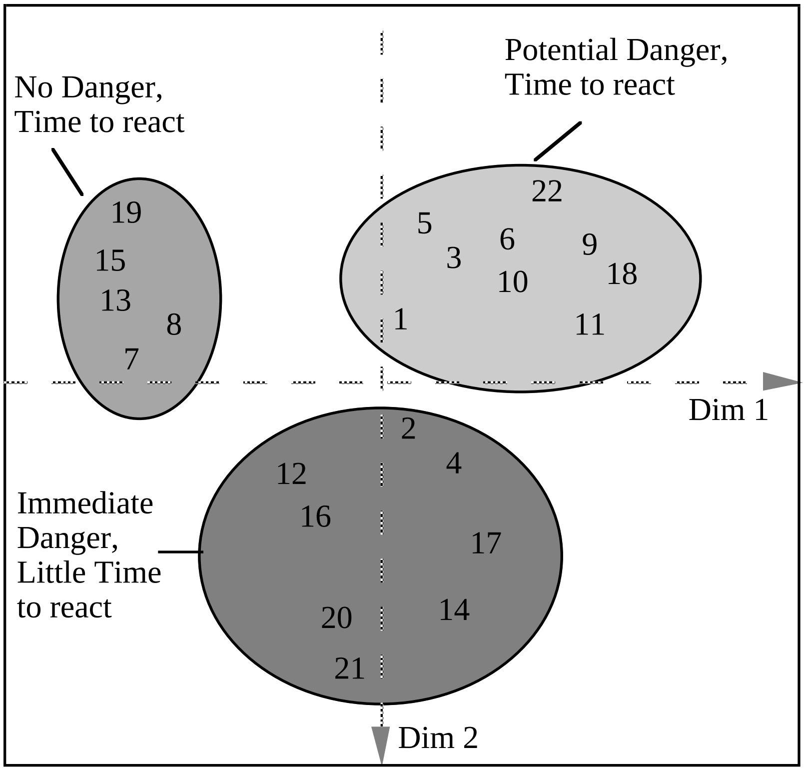

**Figure 1.2.** Summary depiction of the 2-dimensional and 3-cluster solutions of captains' categorizations of the aviation incidents. Numbers (1-22) refer to individual scenarios as listed in the appendix.

in which there is very little time available for making a decision to situations in which time is not a critical factor.

*(#21) Time critical:* A medium large transport is cleared for takeoff on runway 13C. Power is advanced and takeoff roll commences with the crew setting takeoff power. At approximately 100 knots, a crew member notices a light twin engine taxing at a fairly high speed toward 13C. At the same time, he hears the tower controller making repeated unsuccessful attempts to contact the small plane.

*(#19) Not time critical:* After all the passengers board a medium large transport at CHA, a ramp agent reports to the flight crew that the forward cargo door will not lock closed. CHA is a non-maintenance base. A mechanic contracted from the airfield, however, is able to get the cargo door latch mechanism to function marginally--i.e., it can only be closed/opened once and the “Fwd Cargo Door” light stays on.

### *First Officers' Dimensions*

The first cluster obtained for first officers comprises scenarios in which the incident occurred during take-off (12, 16, 20, 21), cruise (4), descent (2, 14) and approach (17). First officers described these incidents as serious or potentially serious and requiring fast decision making, and contrasted them with incidents on the ground (7, 8, 13, 15, 19) or in cruise (5, 10, 11) which they classified as less time critical. Dimension 1 for the first officer group is thus best characterized as *time dimension*. The alignment of scenarios along this dimension, moreover, was found to correspond closely to the ordering of scenarios along the captains' time dimension. A strong correlation was found between the values of each scenario along the first officers' and the captains' time dimension (Pearson's Correlation Coefficient: $r = .88$; $p < .0001$).

The low time criticality situations are further partitioned into two clusters that contrast along Dimension 2. One group involves incidents prior to or during taxiing (7, 8, 13, 15, 19), or during cruise (5, 10, 11). The other group includes incidents either during the climb-out phase of the flight (1, 3, 9, 22) or during the approach phase (6, 18). What distinguishes these two clusters is not readily evident from the first officers' verbal descriptions. It appears though that they contrast with respect to *response determinacy*. The incidents on the ground or during cruise elicited specific responses suggesting that it was apparent to our subjects what should be done in these circumstances. The incidents during climb-out or during approach were less clear-cut in this respect. Although the available response options were known - either, for instance, you continue with the climb-out, or you return - subjects seemed uncertain as to what option should be taken. That is, the scenarios vary along Dimension 2 with respect to how straightforward it is to decide on a course of action. On one extreme are situations in which the decision is clear because there is a rule, a procedure, or a policy in place that tells the crew what to do. For instance, subjects indicated that they would return to the gate if they were confronted with the following situation:

*(#8) Decision clear-cut:* A large transport is preparing to depart DCA for a flight to MSP. The flight is running late and the DCA curfew is nearing. After push back, it is found that there are 185 passengers aboard, but only 184 seats.

At the other extreme are situations which leave options and may require additional information.

*(#22) Decision unclear:* A medium large transport climbing through FL 280 at .82 mach experiences a high frequency flutter through the air frame. Upon visual inspection, it is discovered that the right outboard aileron balance tab is fluttering 1-2” up and down. The flutter stops some time during deceleration to .74 mach and is stable on second inspection. A company mechanic on board indicates there are no visual signs of delamination or structural failure.

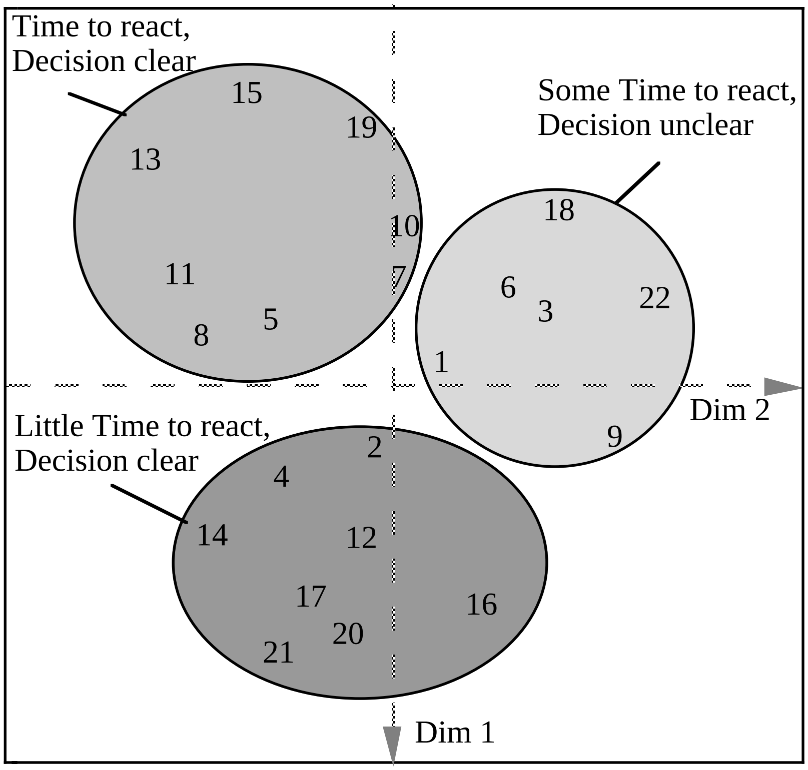

**Figure 1.3.** Summary depiction of the 2-dimensional and 3-cluster solutions of first officers' categorizations of the aviation incidents. Numbers (1-22) refer to individual scenarios as listed in the appendix.

***Second Officers' Dimensions***

The primary dimension obtained for second officers appears to be similar to the _response_ _determinacy_ dimension of the first officers (Pearson Correlation Coefficient: $\underline{r} = .50$; $\underline{p} < .05$).

However, unlike the first officers, second officers apparently relate response determinacy to time criticality and problem understanding. As can be seen in Figure 1.4, situations in which decision making is straightforward involve two types of incidents. One cluster consists of incidents in which decision making is highly time-constrained (2, 12, 16, 17, 20, 21). With no time to spare on deliberating options, the crew simply takes some prescribed course of action.

> _(#20) Decision clear/time-critical_: A wide bodied aircraft weighing 430,000 lbs. experiences a #3 engine compressor stall at 138 knots on take-off; V1 is 142 knots.

The other cluster is made up of ground situations that are not time-critical and in which the crew can rely on procedures or on ground personnel for further input (7, 8, 13, 15, 19). For second officers, the appropriate decision in these instances was straightforward: For instance, simply get maintenance involved.

> _(#13) Decision clear/not time-critical_: The crew of a small transport find nothing out of the ordinary during a preflight walk-around. Then while inside the plane awaiting clearance, they feel the plane move slightly. They immediately look outside and see a big cart wheel near the airstair door. When they ask the driver whether he hit the airplane, he says he may have struck the

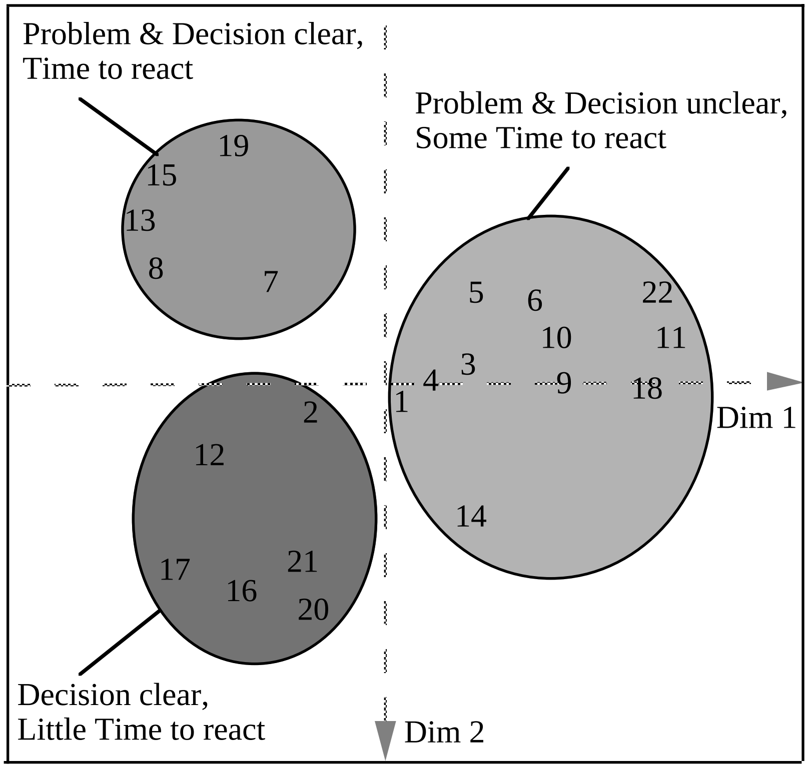

**Figure 1.4.** Summary depiction of the 2-dimensional and 3-cluster solutions of second officers' categorizations of the aviation incidents. Numbers (1-22) refer to individual scenarios as listed in the appendix.

door but he doesn't know. The crew inspect the side of the aircraft and test the door but find nothing unusual.

Time-critical situations and incidents on the ground contrast with situations that second officers felt left them with options. Specifically these situations involved the decision of whether or not to divert (1, 3, 4, 5, 9, 10, 11, 22), or in case of a diversion (14), the choice of an alternate airport that would be best in accommodating passengers and resolving the problem. Uncertainty about the appropriate course of action in these situations was apparently increased by the ambiguous nature of the problem. For instance, in the following situation, the crew could not determine the cause of the problem:

*(#11) Decision unclear/problem unclear*: A medium large transport is in cruise at flight level 310, westbound over the North Atlantic, when the crew discover that the crew oxygen bottle is reading 800 psi. The bottle had been changed in Frankfurt due to an incoming crew write-up and had been definitely noted as full. The crew shut off the regulators to see if a leak point can be determined; none is found but the supply decreases to 750 psi.

Similarly, in the following situation it is unclear whether or not the crew had a problem. Cues available to the crew were contradictory:

*(#6) Decision unclear/problem unclear*: A large transport, on a flight from DFW, begins a visual approach at RNO. Turning crosswind, 2 deg flaps are selected, but the inboard flap indicator shows a split: L=0 degree extension, R=2 degrees. To the crew the aircraft “feels” like both sides are extended because there are no control problems. The crew attempt to position the flaps back up; but the indicator continues to show a split (L=0. R=2).

The second dimension that was obtained for second officers concerns *time constraints* in decision making. This dimension corresponds closely to the time dimension discerned for captains and first officers (Pearson's Correlation Coefficients for dimensional values: $r = .865$, $p < .0001$; $r = .929$, $p < .0001$, respectively).

In summary, a two-dimensional, three cluster solution accounted best for the data from the three crew positions. However, our analyses suggest that the dimensions differed across positions. The captains apparently asked two questions of a situation: (1) Does the situation pose an immediate threat to flight safety, a potential threat, or no threat at all? And (2) How much time do we have to decide? Except for time criticality, first officers and second officers showed other priorities. First officers and second officers shared the time constraint dimension, but instead of risk focused on response determinacy. For first officers, "Time Criticality" was their primary dimension, followed by "Response Determinacy." For second officers, the reverse order was observed: "Response Determinacy" was followed by "Time Criticality."

# STUDY 2

A second study was conducted to confirm our interpretations of the Multidimensional Scaling and Clustering analyses of the free sorting data from Study 1. A *directed* sorting task was employed. In contrast to Study 1, we now provided the participants with specific criteria for sorting the 22 scenarios. The criteria that subjects were instructed to use in their sortings were the dimensional labels inferred in Study 1. Ratings obtained through directed sorting could then be used to predict the alignment of the scenarios along the dimensions observed in the free sorting task. Thus, if our previous interpretations were correct, then subjects who were directed to rate the scenarios, for instance, according to time criticality, should align them similarly to the "Time Criticality" dimension identified in Study 1.

## Method

### Participants

Twenty-seven pilots from the same airline as in Study 1 volunteered as participants. Nine were captains, 10 were first officers and eight were second officers. The captains had on average 24 years of experience in Part 121 aircraft (*SD* = 4.6 yrs.); first officers had 9.2 years (*SD* = 4.6 yrs.), and second officers 5.3 years (*SD* = 3.6 yrs.). Six of the captains also had experience flying military aircraft averaging 7 years (*SD* = 2.4 yrs.) and one captain had seven years of experience flying Part 135 aircraft.[7] All first officers and all second officers were ex-military pilots. The average military experience for first officers was 9.2 years (*SD* = 2.9 yrs.), and for second officers 10.9 years (*SD* = 4.6).

### Material

Participants received the same 22 scenarios that were used in Study 1.

### Design and Procedure

Subjects were told that they would be asked to categorize 22 aviation decision events four times, using four different sorting criteria. The four criteria they were instructed to use were as follows: "how risky is a given situation," "how much time pressure is there," "how clearly defined is the problem," and "how straightforward is it to decide on a course of action." The orders of the four sorting bases were randomized across subjects. For each sort they were asked to create no more than 10 different piles and to align the piles from the one having the least of a particular quality to the one having the most of it.

### Analyses

Subjects' ordered sorts were translated into values on a scale ranging from -1 to +1. For instance, when a subject created 5 piles, the following mapping was implemented: Scenarios that were judged to be lowest on a given dimension received the score -1; those judged to be highest on the dimension were scaled at +1; intermediately rated scenarios were assigned -.5, 0, and +.5, respectively. Mean rating scores were computed for each scenario on risk, time criticality, problem definition and response determinacy. Separate mean scores were calculated for captains, first officers, and second officers. These rating scores were then transformed to z-scores and used as independent variables in multiple regression analyses to predict scenario values along the dimensions obtained for each group in Study 1.

## Results

### Validation of the Captains' Dimensions

Captains created on average five piles when sorting for risk and time criticality ($SD$ = 1.5; $SD$ = 1.7, respectively), and four piles when sorting for problem definition and response determinacy ($SD$ = 1.2 in both instances). From these sortings, mean rating scores for each scenario concerning its risk, time criticality, problem definition and response determinacy were derived as outlined above. Since the rating scores were subsequently to be used as predictor variables in multiple regression analyses, we assessed their fit with the assumptions underlying multiple regression. All assumptions were met except for multicollinearity. Significant correlations were observed between captains' ratings of risk and time criticality ($r$ = .74; $p$ < .001) as well as between their ratings of problem definition and response determinacy ($r$ = .84; $p$ < .001). To avoid the problems arising from multicollinearity of predictor variables, we adopted the following strategy: From two correlated predictors, only the variable that had the higher correlation with the dependent variable would enter the regression equation.

#### *Captain Dimension 1 as Potential Risk*

Table 2.0 displays the simple regressions between the scenarios' values along the *Potential Risk Dimension* from Study 1 and their mean rating scores of risk, time criticality, response determinacy and problem definition of Study 2.

**Table 2.0.** *Simple Regressions between Captains' Mean Ratings of Scenarios and their Values Obtained in Multidimensional Scaling ($N$ = 22).*

| CAPTAINS' RATINGS OF | CAPTAINS' DIMENSIONS | | | |
|---|---|---|---|---|
| | Dimension 1 (= Potential Risk?) | | Dimension 2 (= Time Criticality?) | |
| | $R$ | $R^2$ | $R$ | $R^2$ |
| Risk | .40$^{\dagger}$ | .16 | .64*** | .41 |
| Time Criticality | .41$^{\dagger\dagger}$ | .17 | .84**** | .71 |
| Response Determinacy | .19 | .04 | .54** | .29 |
| Problem Definition | .07 | .005 | .52* | .27 |

*Note*: $^{\dagger}p$ = .07. $^{\dagger\dagger}p$ = .06. *$p$ < .05. **$p$ < .01. ***$p$ < .001. ****$p$ < .0001.

The alignment of problems according to Risk was marginally significant compared to the Potential Risk Dimension of Study 1. Reflecting the collinearity between the two rating scores, Time Criticality was also marginally related to Dimension 1.

Using multiple regression, we further explored the validity of our interpretation of Dimension 1 as a risk dimension. In addition to the mean ratings of risk and of response determinacy, the squared mean rating of time criticality was entered as a predictor variable for values along this dimension. Ratings for time criticality were squared since initial data screening revealed a quadratic relation with Dimension 1. Note that a quadratic relationship between this dimension and the time variable is in line with our initial interpretation of the dimension as *potential* risk; i.e., with the distinction between situations in which a threat to flight safety is imminent versus those involving no risk or those in which a threat may occur at some future point. $R$ (.783) for this regression was significantly different from zero, $F(3,18) = 9.486$; $p < 0001$. As can be seen in Table 2.1, the only significant predictors were the variables Risk and Time Criticality squared. Together they explained 57% of the variability in Dimension 1 values from Study 1.

### *Captain Dimension 2 as Time Criticality*

As shown in the right hand column of Table 2.0, simple regression analyses between values along Dimension 2 of Study 1 and the four rating scales indicated that ratings of time criticality provided the best fit for Dimension 2. Similarly, ratings of response determinacy were slightly better than those obtained for problem definition and were subsequently entered along with ratings of time criticality into a multiple regression analysis to predict values along the *Time Dimension* of Study 1.

**Table 2.1.** *Summary of Standard Multiple Regression Analysis for Variables Predicting Captains' Risk Dimension (= Dimension 1) (N = 22).*

| Variables | B | SE B | ß |
|---|---|---|---|
| Risk | .62 | .17 | .56** |
| (Time Criticality)$^2$ | -.76 | .19 | -.62*** |
| Response Determinacy | -.18 | .18 | -.16 |

*Note*: $R^2 = .61$; adjusted $R^2 = .55$; unique variability (of Risk and Time Criticality squared combined) = .57; shared variability (of all variables) = .02.
**$p < .01$. ***$p < .001$.

**Table 2.2.** *Summary of Standard Multiple Regression Analysis for Variables Predicting Captains' Time Dimension (= Dimension 2) (N = 22).*

| Variables | B | SE B | ß | $sr^2$ (unique) |
|---|---|---|---|---|
| Time Criticality | -.69 | .11 | -.74**** | .47 |
| Response Determinacy | -.25 | .11 | -.27* | .06 |

Note: $R^2 = .76$; adjusted $R^2 = .74$; shared variability (of all variables) = .23.
*$p < .05$. ****$p < .0001$.

Table 2.2 shows that both Time Criticality and Response Determinacy significantly explained values along Dimension 2. Time Criticality, however, was found to be a stronger predictor than Response Determinacy. Its unique contribution to the variability in dimensional values was .47; the contribution of Response Determinacy was .06. The overall fit of the model involving both variables was significantly different from zero ($R = .872$; $F(2,19) = 30.10$; $p < .0001$) and accounted for 76% of the variance (74% adjusted).

## Validation of the First Officers' Dimensions

First officers sorted the incidents on average into five piles when rating risk, time criticality and response determinacy ($SD = 1.3$; $SD = 1.7$; $SD = 1.3$, respectively), and into four piles when rating problem definition ($SD = 1.0$). Ratings for risk and time criticality were found to be related. The scenarios' mean rating scores for these two variables were significantly correlated ($r = .92$; $p < .001$). No other simple correlation approached significance. As previously with the analysis of the captains' data, we decided to include either risk or time criticality in a regression equation dependent on which one of them had the highest simple correlation with the dependent variable.

### *First Officer Dimension 1 as Time Criticality*

Table 2.3 lists the simple regressions between first officers' mean rating scores for the 22 scenarios and their corresponding values along the *Time Criticality Dimension* that was hypothesized for the first officer group in Study 1.

Ratings of time criticality, problem definition, and response determinacy were subsequently used as independent variables in a multiple regression analysis for predicting each scenario's value along Dimension 1 of Study 1. Results of this regression analysis are summarized in Table 2.4. The overall regression model was statistically significant ($R = .888$; $F(3,18) = 22.28$; $p < .0001$). The three variables explained 79% of the variability in values along this dimension.

However, neither Problem Definition nor Response Determinacy added significantly to the prediction of dimensional values over and above the contribution of Time Criticality. This variable alone accounted for 62.6% of the variance.

**Table 2.3.** *Simple Regressions between First Officers' Mean Ratings of Scenarios and their Values Obtained in Multidimensional Scaling (N = 22).*

| FIRST OFFICERS' RATINGS (Study 2) | (Study 1) FIRST OFFICERS' DIMENSIONS | | | |
|---|---|---|---|---|
| | Dimension 1 (= Time Criticality?) | | Dimension 2 (= Response Determinacy?) | |
| | $R$ | $R^2$ | $R$ | $R^2$ |
| Risk | .70*** | .49 | .09 | .01 |
| Time Criticality | .88**** | .77 | .12 | .01 |
| Response Determinacy | .02 | .0004 | .15 | .02 |
| Problem Definition | .40† | .16 | .04 | .002 |

*Note*: †$p$=.07. ***$p$<.001. ****$p$<.0001.

**Table 2.4.** *Summary of Standard Multiple Regression Analysis for Variables Predicting First Officers' Time Dimension (= Dimension 1) (N = 22).*

| Variables | B | SE B | ß |
|---|---|---|---|
| Time Criticality | -.97 | .13 | -.86**** |
| Response Determinacy | -.09 | .13 | -.08 |
| Problem Definition | -.11 | .13 | -.10 |

*Note*: $R^2$ = .79; adjusted $R^2$ = .75; unique variability (of Time Criticality) = .63; shared variability (of all variables) = .14.
****$p$ < .0001.

### *First Officer Dimension 2 as Response Determinacy*

The correlation coefficients for Dimension 2 and each of the four rating scales are displayed in the right hand column of Table 2.3. As can be seen, no significant linear relations were observed. Inspection of the residual plots of the simple regressions revealed a quadratic relation between Dimension 2 and Time Criticality as well as between Dimension 2 and Risk. However, since neither Problem Definition nor Response Determinacy had significant simple correlations with this dimension, no meaningful regression model could be constructed.

Our interpretation of the first officers' second dimension as Response Determinacy was thus not supported. This finding may imply that we did not infer the appropriate label when we interpreted the multidimensional scaling analysis of the free sorting data for this group. Alternatively, it may mean that we selected the incorrect number of dimensions, and that we should adopt the computationally less parsimonious one-dimensional solution in the multidimensional scaling analysis. We will return to this issue in Study 3.

## Validation of the Second Officers' Dimensions

Second officers created on average five piles when their judgments concerned risk and time criticality ($SD = 1.6$; $SD = 1.5$, respectively), three piles for problem definition ($SD = 1.4$) and four piles concerning response determinacy ($SD = 1.1$). Since ratings for risk and time criticality correlated significantly ($r = .93$; $p < .001$), only one of the two variables could be included in subsequent multiple regression analyses. As previously, strongest association with the dependent variable was the criterion for selecting risk or time criticality as predictor variable.

### *Second Officer Dimension 1 as Response Determinacy*

Table 2.5 shows the simple regressions between second officers' mean rating scores and values along Dimension 1 of Study 1. Ratings of Response Determinacy were significantly aligned with scores on Dimension 1.

**Table 2.5.** *Simple Regressions between Second Officers' Mean Ratings of Scenarios and their Values Obtained in Multidimensional Scaling ($N = 22$).*

| SECOND OFFICERS' RATINGS (Study 2) | (Study 1) SECOND OFFICERS' DIMENSIONS | | | |
|---|---|---|---|---|
| | Dimension 1 (= Response Determinacy?) | | Dimension 2 (= Time Criticality?) | |
| | $R$ | $R^2$ | $R$ | $R^2$ |
| Risk | .36 | .13 | .70** | .49 |
| Time Criticality | .21 | .04 | .83**** | .69 |
| Response Determinacy | .44* | .19 | .09 | .01 |
| Problem Definition | .11 | .01 | .11 | .01 |

*Note*: *$p < .05$. **$p < .01$. ****$p < .0001$.

A multiple regression analysis was conducted involving second officers' ratings of response determinacy, problem definition and risk as independent variables and values along Dimension 1 as the dependent variable. The overall $\underline{R}$ (.543) for this regression was marginally significant ($\underline{F}(3,18) = 2.51$; $\underline{p} = .09$) with Response Determinacy as significant and Risk as marginally significant predictors. We repeated the analysis, but dropped Problem Definition from the equation, to see whether we could improve the fit of the model. As can be seen in Table 2.6, the resulting $\underline{R}$ (.538) was significantly different from zero ($\underline{F}(2,19) = 3.87$; $\underline{p} < .05$). Risk and Response Determinacy together accounted for 29% of the variability in the dimensional values; but only Response Determinacy contributed significantly to the prediction ($sr^2 = .16$).

### *Second Officer Dimension 2 as Time Criticality*

The simple regressions between Dimension 2 and second officers' ratings are also given in the right hand column of Table 2.5. The subsequent multiple regression analysis for Dimension 2 involved three predictor variables: Time Criticality, Response Determinacy, and Problem Definition. As can be seen in Table 2.7, the three variables together provided a good fit for the dimensional values ($\underline{R} = .84$; $\underline{F}(3,18) = 14.6$; $\underline{p} < .0001$), but most of its predictive power was due to Time Criticality. 69% of the dimension's variance could be explained by this variable.

**Table 2.6.** *Summary of Standard Multiple Regression Analysis for Variables Predicting Second Officers' Response Dimension (= Dimension 1) ($\underline{N} = 22$).*

| Variables | B | SE B | ß |
|---|---|---|---|
| Risk | .35 | .22 | .31 |
| Response Determinacy | -.47 | .22 | -.41* |

*Note*: $\underline{R}^2 = .29$; adjusted $\underline{R}^2 = .21$; unique variability (of Response Determinacy) = .16; shared variability (of all variables) = .04.
*$\underline{p} < .05$.

**Table 2.7.** *Summary of Standard Multiple Regression Analysis for Variables Predicting Second Officers' Time Dimension (= Dimension 2) ($\underline{N} = 22$).*

| Variables | B | SE B | ß |
|---|---|---|---|
| Time Criticality | .74 | .11 | .84**** |
| Response Determinacy | .13 | .11 | .15 |
| Problem Definition | .02 | .11 | .02 |

*Note*: $\underline{R}^2 = .71$; adjusted $\underline{R}^2 = .66$; unique variability (of Time Criticality) = .69.
****$\underline{p} < .0001$.

## Discussion

With the exception of the first officers' second dimension, Study 2 confirmed our interpretations of the multidimensional scaling analyses of the free sorting data. Together, studies 1 and 2 suggest that pilots categorize decision events by assessing how much risk there is in a given situation, how much time they have for making a decision and whether there is only one, clear-cut answer to a problem, or whether the situation leaves options. Risk and constraints on available time and responses are three of the four aspects Orasanu (1994) proposed to be crucial to aviation decision making. Problem definition, the fourth aspect in Orasanu’s model, was not validated in these studies.

Studies 1 and 2 also indicate systematic differences in the categories created by captains, first officers and second officers. Captains in the free sorting condition (Study 1) were primarily sensitive to risk and time criticality. First officers focused on time criticality, and second officers categorized situations in terms of available response options, in addition to time criticality.

Two questions remain unanswered by these findings. (1) Are the differences between crew positions in feature selection related to years of aviation experience, or do they reflect role-specific biases? (2) Do the differences in feature selection reflect differences in what pilots know about decision-relevant components? Our analyses thus far do not permit us to address the first question since years of experience and pilot role were confounded variables in Study 1. This issue will be addressed in Study 3. Question 2, on the other hand, can be answered within the context of Study 2.

Recall that all pilots in Study 2 judged the decision scenarios with respect to risk, time criticality, problem definition, and response determinacy. If the different sortings by captains, first officers and second officers in Study 1 reflect differences in what they know about these features with respect to specific aviation decision situations, then the pilot groups in Study 2 should differ in their ratings; i.e., there should be no significant agreement between captains', first officers' and second officers' ratings of risk, time criticality, problem definition and response determinacy in Study 2. If, on the other hand, the groups differ not in what they know but in the relative importance they assign to the four dimensions in aviation decision making, then their ratings in the directed sorting task should be highly correlated.

To test this hypothesis, correlation coefficients were computed between the captains', first officers', and second officers' mean rating scores for each scenario on each of the four sorting dimensions. As can be seen in Table 2.8, the groups agreed strongly in their ratings of the scenarios on all four dimensions. This indicates that expert pilots in different crew positions apparently have similar understandings of which events do or do not impose a threat to flight safety, which situations involve time pressure, which problems are well-defined, and which have clear-cut responses. Despite this common understanding, they spontaneously focus on different dimensions.

**Table 2.8.** *Intercorrelations Between Captains', First and Second Officers' Mean Ratings of Scenarios (N = 22).*

| Crew Position | 1 | 2 | 3 |
|---|---|---|---|
|  |  | Ratings of Risk |  |
| 1 Captains | -- | .85**** | .83**** |
| 2 First Officers |  | -- | .96**** |
| 3 Second Officers |  |  | -- |
|  |  | Ratings of Time Criticality |  |
| 1 Captains | -- | .98**** | .83**** |
| 2 First Officers |  | -- | .94**** |
| 3 Second Officers |  |  | -- |
|  |  | Ratings of Problem Definition |  |
| 1 Captains | -- | .65*** | .59** |
| 2 First Officers |  | -- | .77**** |
| 3 Second Officers |  |  | -- |
|  |  | Ratings of Response Determinacy |  |
| 1 Captains | -- | .63** | .59** |
| 2 First Officers |  | -- | .76**** |
| 3 Second Officers |  |  | -- |

Note: **$p$ < .01. ***$p$ < .001. ****$p$ < .0001.

## STUDY 3

The third study was designed to explore further the contributions of years of experience and crew role to pilots' spontaneous categorization of decision situations. Specifically, we addressed the question left unanswered by Study 1: Are the differences between crew positions in feature selection related to years of aviation experience, or do they reflect role-specific biases? These two aspects of aviation experience were confounded in Study 1, since the captains were more experienced pilots than the first officers, who, in turn, were more experienced than the second officers. To assess the relative importance of years of aviation experience versus crew role, we repeated Study 1 with a new sample of pilots who had many fewer years of aviation experience than those in Study 1. In fact, the average experience level of captains in Study 3 was comparable to that of the first officers in Study 1. If crew role determined the pattern of effects found in Study 1, then this group of captains should categorize aviation incidents similarly to the more experienced captains in the first study. On the other hand, if years of experience determines what aspects pilots consider in their classifications, then this new group of captains should use the dimensions observed for the first officers in Study 1.

## Method

### Participants

Twenty-nine pilots from a different U.S. airline volunteered as subjects. Thirteen were captains with an average of 14.9 years of experience (*SD* = 5.5 yrs.) in Part 121 transport. This level of experience was identical to that of the first officers in Study 1. The 16 first officers had on average 9.8 years of Part 121 aviation experience (*SD* = 5.6 yrs.).

Four captains also had experience in Part 135 aircraft averaging 5.3 years (*SD* = 2.5 yrs.); eight others had on average 10.4 years of experience flying military aircraft (*SD* = 6.4 yrs.). Of the first officers, 12 had additional Part 135 experience averaging 6.3 years (*SD* = 4.3 yrs.), and seven had military aviation experience averaging 11.8 years (*SD* = 7.5 yrs.).

### Material

The same 22 incident scenarios as in the previous studies were used.

### Design and Procedure

Design and procedure of this study were identical to those employed in Study 1, the free sorting task.

### Analyses

Identical to Study 1, analyses were based on data from single or preferred sorts. Disassociation scores were calculated for pairs of scenarios and taken as input into multidimensional scaling and hierarchical clustering analyses. Data for captains and first officers were analyzed separately.

The dimensional values obtained for the scenarios with multidimensional scaling were then used as dependent variables in standard multiple regression analyses. The predictor variables in these analyses were the captains' (or the first officers') ratings of risk, time pressure, problem definition and response determinacy that were collected in Study 2.

## Results

### Number of Piles

Captains sorted the 22 scenarios on average into 6 piles (*SD* = 2.1); first officers into 5 piles (*SD* = 1.5).

### Number of Dimensions

Multidimensional scaling analyses for 1-, 2-, 3-, and 4-dimensional solutions were performed on the sorting data of captains and first officers. The optimal number of dimensions was determined applying the rule detailed in the results section of Study 1. A two-dimensional solution was found to account best for the captains' as well as the first officers’ sorting judgments. Subsequent hierarchical clustering analyses confirmed this solution for both data sets. Figure 3.0 depicts the results obtained for the captain data. As can be seen, a three-cluster solution mapped well onto the two dimensions except for two

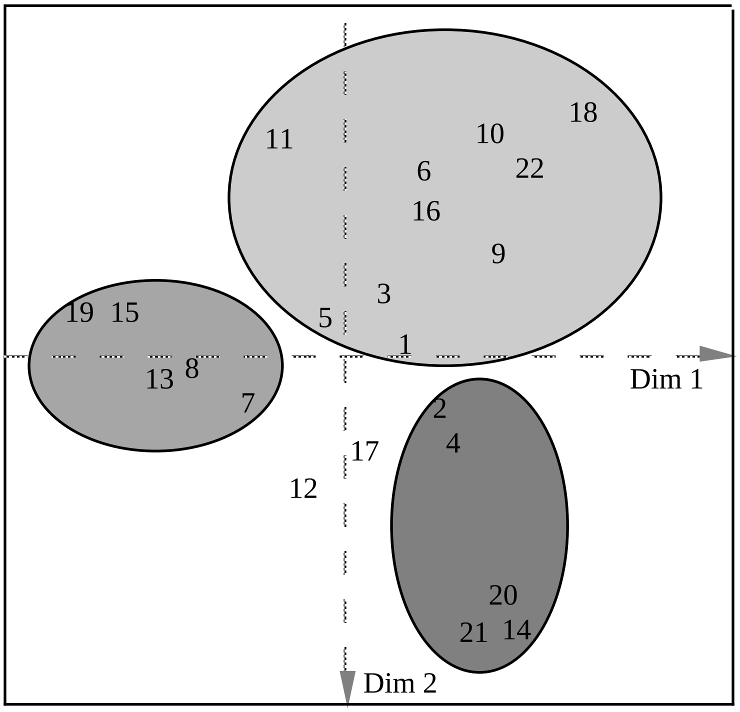

**Figure 3.0.** Summary depiction of the 2-dimensional and 3-cluster solutions of junior captains' categorizations of the aviation incidents. Numbers (1-22) refer to individual scenarios as listed in the appendix.

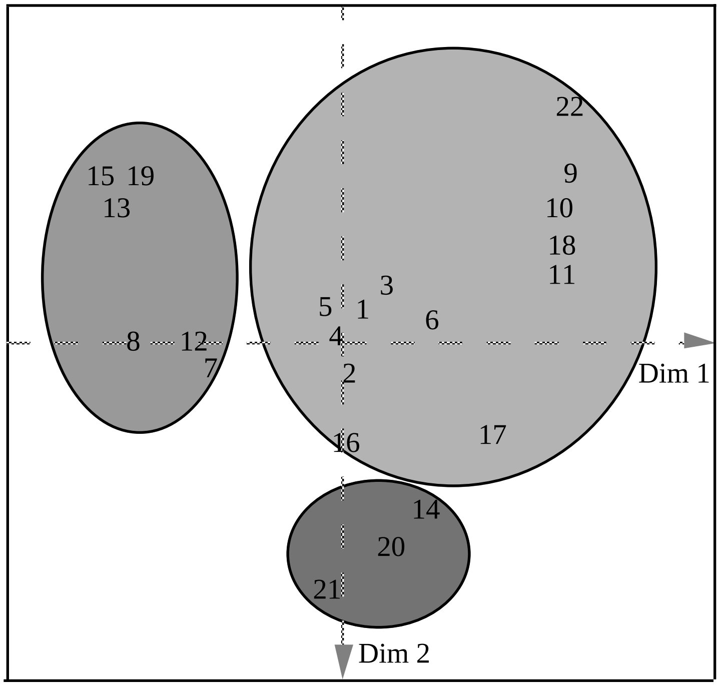

**Figure 3.1.** Summary depiction of the 2-dimensional and 3-cluster solutions of junior first officers' categorizations of the aviation incidents. Numbers (1-22) refer to individual scenarios as listed in the appendix.

scenarios (12 and 17). Results for the first officers' sorting judgments are shown in Figure 3.1.

## Interpretation of Captains' Dimensions

As in Study 2, we used ratings of risk, time criticality, response determinacy and problem definition to approximate the alignment of the scenarios along the dimensions. Recall that for the captains, ratings of risk and time criticality were significantly correlated, as were ratings of response determinacy and problem definition. Thus only one variable of each pair could be entered into multiple regression analysis. As previously, this decision was based on their simple regressions with a criterion variable. Table 3.0 displays simple regressions between the four variables and Dimension 1 as well as Dimension 2.

Risk, Problem Definition, and the square of Time Criticality were used in a standard multiple regression analysis to predict values along *Dimension 1*. The square of the variable Time Criticality was entered since the simple regression of Dimension 1 by ratings of time criticality indicated a quadratic relation between the two variables.

**Table 3.0.** *Simple regressions between Junior Captains' mean ratings of scenarios and their values obtained in multidimensional scaling (N = 22).*

| CAPTAINS' RATINGS OF | CAPTAINS' DIMENSIONS | | | |
|---|---|---|---|---|
| | Dimension 1 | | Dimension 2 | |
| | $R$ | $R^2$ | $R$ | $R^2$ |
| Risk | .65*** | .42 | .44* | .19 |
| Time Criticality | .64*** | .41 | .57** | .32 |
| Response Determinacy | .11 | .01 | .64*** | .41 |
| Problem Definition | .20 | .04 | .55** | .30 |

*Note*: *$p < .05$. **$p < .01$. ***$p < .001$. ****$p < .0001$.

**Table 3.1.** *Summary of Standard Multiple Regression Analysis for Variables Predicting Junior Captains' Dimension 1 (N = 22).*

| Variables | B | SE B | ß |
|---|---|---|---|
| Risk | .78 | .16 | .70**** |
| (Time Criticality)$^2$ | -.61 | .18 | -.50** |
| Problem Definition | .21 | .16 | .19 |

*Note*: $R^2$ = .65; adjusted $R^2$ = .60; unique variability (of Risk and Time Criticality combined) = .61; shared variability (of all variables) = .01.
**$p < .01$. ****$p < .0001$.

Table 3.1 summarizes how well each of the three independent variables predicted values along Dimension 1. R (.81) based on all three predictor variables was significantly different from zero, F(3,18) = 11.30, $p < .001$); but only Risk and Time Criticality squared made significant contributions to the prediction of values along Dimension 1. These two variables combined contributed .61 to the overall prediction of .65.

The regression equation modeling values along Dimension 2 involved two predictor variables, Response Determinacy and Time Criticality. As shown in Table 3.2, this model was significantly different from zero ($R = .73$, $F(2,19) = 10.70$, $p < .001$) and both variables made significant contributions to the prediction ($R^2 = .53$, adjusted .48). The unique contribution of Response Determinacy was .21. Time Pressure uniquely contributed .13.

## Interpretation of the First Officers' Dimensions

Predictor variables were first officers' ratings of risk, time criticality, response determinacy, and problem definition obtained from the directed sorting of study 2. Risk and Time Criticality could not be used in the same regression model since they were found to be significantly correlated. Again we selected either variable depending on the size of its correlation with a criterion variable. Table 3.3 displays the simple regressions between the four variables and Dimension 1 as well as Dimension 2.

**Table 3.2.** *Summary of Standard Multiple Regression Analysis for Variables Predicting Junior Captains' Dimension 2 (N = 22).*

| Variables | B | SE B | ß | $sr^2$ (unique) |
|---|---|---|---|---|
| Time Criticality | -.35 | .16 | -.38* | .13 |
| Response Determinacy | -.46 | .16 | -.50** | .21 |

Note: $R^2 = .53$; adjusted $R^2 = .48$; shared variability (of all variables) = .19.
*$p < .05$. **$p < .01$.

**Table 3.3.** *Simple regressions between Junior First Officers' mean ratings of scenarios and their values obtained in multidimensional scaling (N = 22).*

| FIRST OFFICERS' RATINGS OF | FIRST OFFICERS' DIMENSIONS | | | |
|---|---|---|---|---|
| | Dimension 1 | | Dimension 2 | |
| | $R$ | $R^2$ | $R$ | $R^2$ |
| Risk | .48* | .23 | .68*** | .46 |
| Time Criticality | .40† | .16 | .83**** | .68 |
| Response Determinacy | .48* | .23 | .18 | .03 |
| Problem Definition | .21 | .04 | .30 | .09 |

Note: †$p = .07$. *$p < .05$. **$p < .01$. ***$p < .001$. ****$p < .0001$.

**Table 3.4.** *Summary of Standard Multiple Regression Analysis for Variables Predicting Junior First Officers' Dimension 1 ($N = 22$).*

| Variables | B | SE B | ß |
|---|---|---|---|
| Risk | .40 | .24 | .34 |
| Response Determinacy | -.50 | .22 | -.43* |
| Problem Definition | .17 | .23 | .14 |

*Note*: $R^2 = .40$; adjusted $R^2 = .30$; unique variability (of Response Determinacy) = .17; shared variability (of all variables) = .23.
****$p < .05$.

**Table 3.5.** *Summary of Standard Multiple Regression Analysis for Variables Predicting Junior First Officers' Dimension 2 ($N = 22$).*

| Variables | B | SE B | ß | $sr^2$ (unique) |
|---|---|---|---|---|
| Time Criticality | -.74 | .11 | -.87**** | .65 |
| Response Determinacy | -.25 | .10 | -.29* | .08 |
| Problem Definition | .03 | .11 | .03 | -- |

*Note*: $R^2 = .76$; adjusted $R^2 = .72$; shared variability (of all variables) = .03.
*$p < .05$. ****$p < .0001$.

A standard multiple regression was conducted between Risk, Problem Definition, and Response Determinacy as independent variables and *Dimension 1* as dependent variable. Results of this analysis are summarized in Table 3.4. $R$ (.63) was significantly different from zero ($F(3,18) = 3.93$, $p < .05$). The only significant predictor, however, was Response Determinacy. The unique contribution of this variable to the overall prediction ($R^2 = .40$, adjusted .30) was .17.

Time Criticality, Response Determinacy and Problem Definition were used as independent variables in a standard multiple regression to predict *Dimension 2*. The overall model was significantly different from zero ($R = .87$, $F(3,18) = 19.22$, $p < .0001$). All three variables accounted for 76% of the variance in Dimension 2 (adjusted 72%). However, as shown in Table 3.5 , only two variables, Time Criticality and Response Determinacy, were significant predictors. The unique contribution of Time Criticality was .65, and of Response Determinacy .08.

## Discussion

These findings are consistent with the view that crew role affects how pilots categorize aviation decision events. Years of aviation experience per se does not seem to have a significant impact. The captains in Study 3 showed preferences similar to their more senior colleagues in Study 1 and differed from the first officers in Study 1 who were their equals in years of aviation experience. Risk was again the captains' primary dimension. The second dimension, as before, reflected concern for both time constraints and determinacy of available responses. The relative importance of the two variables, however, was more balanced for this group than for those in the earlier study.

The sortings by the first officers in Study 3 corresponded to the dimensions observed for second officers in the previous study: response determinacy, followed by time criticality. This result, although at first glance contradictory, may indeed indicate role-specific experience effects. The first officers in Study 3, who in terms of experience levels fall between the first and second officers in Study 1, flew aircraft types that involve only two crew members. In contrast, the majority of first officers in Study 1 (60%) came from three-person crews. How responsibilities are divided up on the flight deck is partly determined by the size of the crew. On a two-person flight deck, tasks that have traditionally been part of the second officer's role are frequently allocated to automated systems or are shared by first officers and captains.

Some of the differences between findings from Study 1 and Study 3 may thus be accounted for by crew size. The first officers in Study 3 may have sorted the decision events more like the second officers than the first officers in Study 1 because their crew role included some of the responsibilities of a second officer. Similarly, the fact that the captains in Study 3 came exclusively from two-person crews whereas only half of the captains in Study 1 did may explain why response determinacy was more salient in the classifications of the former than the latter captain group.

Differences in company policy concerning allocation of crew responsibilities or in company climate may also have contributed to the different emphases found in Studies 1 and 3. The pilots in Study 3 were from a smaller airline than those in Study 1. Smaller airlines may adopt a more collegial, less hierarchical crew structure than large corporations. It is thus possible that captains and first officers in Study 3 were used to sharing responsibilities that were more clearly divided up among the crew members in Study 1.

## Overall Discussion

The studies reported here were designed to address several questions, one quite applied, the others more basic. The applied question was, Do pilots who sit in different seats on the flight deck “see” problems differently? If so, what is the nature of those differences and what might account for them? The more fundamental issue is what our findings contribute to an understanding of the nature and the development of expertise in complex engineered domains such as aviation.

Results from a set of sorting studies found consistent differences between crew positions in how pilots categorized events requiring flight decisions: Captains, but not first officers or second officers, focused on the severity of a problem and its potential threat to flight safety. First officers and second officers were concerned with whether a situation called for a prescribed response or for crew judgment in determining what to do. All pilots evaluated how much time they had for dealing with a problem and for deciding on a course of action. This pattern indicates that expertise within aviation is differentiated, at least with respect to how problems are interpreted. Moreover, these differences are associated with specialized roles on the flight deck, not simply with years of experience on the job. Our discussion will further address the meaning and possible source of these differences.

### Category Structures

One objective of our present research was to determine the kinds of categories expert pilots impose on decision events. Based on previous work that examined expert/novice differences in problem solving (for an overview see Glaser & Chi, 1988), we assumed that the categories would reflect fundamental components of aviation decision making rather than superficial features. An initial impression of what these basic components might be was gleaned from aviation accident investigations by the National Transportation Safety Board (NTSB). NTSB reports on accidents involving faulty crew decision making suggested four fundamental factors**:** crews misjudged the severity of a situation, mishandled or underestimated time available, misdiagnosed the problem, or neglected standard procedures or alternatives to their course of action (Orasanu, Dismukes, & Fischer, 1993). Additional analyses of aviation incident reports and observations of professional pilots' performance in simulated flights involving decision making converged on the same four components (Orasanu, Fischer, & Tarrel, 1993). These observations led to the development of a decision process model whose two primary elements are situation assessment and choosing a course of action. Situation assessment involves defining the problem, and assessing levels of risk and time pressure (Orasanu, 1994).

In the present series of studies we found that expert pilots indeed categorized aviation decision events by fundamental components rather than in terms of surface features such as phase of flight or type of aircraft system involved. The components relevant to the pilots in our studies were potential risk, time constraints, and response options afforded by the problem situation. Contrary to our expectations, pilots did not classify incidents in terms of problem definition; i.e., the extent to which the nature of the problem was understood or known.

This result was surprising. Models of human problem solving place considerable emphasis on problem understanding as a primary determinant of problem solving strategies (Newell & Simon, 1972; Lipshitz & Strauss, 1997; Orasanu, 1995; Voss et al., 1991). A problem must first be understood before a solution can be sought, or the available options determined. Recognition-Primed Decision theory relies on appropriate recognition and classification of a situation as a basis for retrieving an appropriate response (Klein, 1989, 1993). Problem understanding, moreover, carries an additional weight in domains such as aviation in which errors might have lethal consequences. A large number of aviation incident reports indicates that problems frequently are ill-defined and that pilots' understanding of the situation may be lacking (Gibson, Orasanu, Villeda, & Nygren, 1997). Moreover, recent analyses of aviation accidents indicate that inappropriate assessment of the situation has played a causal role in a large proportion of the accidents (Endsley, 1995; NTSB, 1994). For example, a flight crew shut down one engine on a two-engine aircraft in response to some vague cues without first verifying which engine had failed. In fact, they shut down the functional engine.

Yet, the categories created by pilots in this study suggest that the distinction between well-defined and ill-defined problems was not significant. More important to them was the issue of how to respond: whether the problem was covered by a rule or procedure, required choosing between options, or had no known response and required invention of a novel solution or adaptation of an old one. It may be that pilots lump together ill-defined problems and those in which responses are not determined in contrast to those in which the problem is clear and they know what to do. A significant correlation was found between problem definition and response determinacy (r=.84 for captains), suggesting that at least the captains in this study created categories of ill-structured versus well-structure problems, regardless of whether the uncertainty or ambiguity lay in problem definition or response determinacy. It may well be that response determinacy dominated in the sorting task because pilots are action oriented and focus on what they can do (even to reduce the ambiguity associated with a problem). The instructions to subjects could also have played a role. Pilots were instructed to sort on the basis of the major decision required by the scenarios, which may have implied a focus on the response component to the neglect of problem definition.

Alternatively, one could argue that we did not observe problem understanding as a dimension in pilots' categorizations because of methodological limitations of our studies. First, our sample of scenarios may not have been sufficiently varied in terms of clarity of problem definition. Lack of contrast along this dimension may have led pilots to disregard it as a meaningful basis for sorting. However, data from the directed sorting task tend not to support this explanation. Pilots did not create notably fewer piles when they sorted the scenarios for problem understanding than when they classified them according to risk, time criticality, or response determinacy. If our sample of scenarios was indeed biased, they should have made fewer distinctions pertaining to problem definition, i.e., created fewer piles.

A second possible methodological limitation is that aviation incidents are often multilayered, allowing for diverse views on the nature of their problem structure. Insufficient agreement among pilots concerning the problem structure of the incidents may have introduced too much noise into the analyses and diminished the importance of the problem definition dimension. This explanation also fails to be supported by data. In the directed sorting task pilots exhibited comparable levels of agreement when sorting on problem definition and on response determinacy. Nonetheless, response determinacy proved to be a reliable dimension in pilots' categorizations, while problem definition did not**.** Consequently, there is little reason to suspect that variability among pilots' judgments concealed problem definition as a dimension in their sortings. Thus, neither methodological consideration is as convincing as the argument that pilots treated problem understanding as part of deciding what action to take.

## Why did pilots’ categories of decision events vary by crew role?

Our finding that different underlying dimensions were used by captains versus first and second officers to categorize aviation incidents is consistent with recent theories of expertise emphasizing the fit of experts’ behavior to task demands (Ericsson & Lehman, 1996; Johnson, Kochevar, & Zualkernan, 1992). Recall that in the free sorting tasks, we did not tell pilots to assume any particular perspective, but simply asked them to sort the problem events according to similarity of major decisions. Nonetheless the categories created by captains, on the one hand, and first and second officers, on the other, correspond to their differing responsibilities during both normal and abnormal events in flight. According to the Federal Aviation Agency, captains have “full control and authority in the operation of the aircraft” (Federal Aviation Regulation 121.533e). They are “during flight time in command of the aircraft and crew and (are) responsible for the safety of the passengers, crew members, cargo, and airplane” (FAR 121.533d). They alone have both the authority and responsibility for making decisions concerning all aspects of flight. It is thus not surprising that risk was most important to the captains in our studies. Evaluating the likelihood and degree of threat associated with a problem is fundamental to deciding on a course of action to ensure flight safety, while achieving company goals. It also accords with their role of managing the crew's problem solving and decision making efforts. Risk assessment is inherently strategic and fundamental to aviation decision making since the severity of a problem determines potential changes in goals and plans.

While the captain is responsible for making decisions, the other crew members are responsible for supporting the captain by obtaining decision-relevant information and options. Tactical issues such as troubleshooting systems and finding appropriate checklists are generally delegated to first and second officers. Participants in our studies from these two pilot groups categorized aviation incidents accordingly by emphasizing the extent to which problems provide for determinate responses. Monitoring the time and resources required and available are also tasks typically carried out by first and second officers, likewise reflected in the primacy of the time constraint dimension.

Teamwork in complex environments is characterized by division of labor (Hackman, 1993; Hutchins, 1995; Salas, Dickinson, Converse & Tannenbaum, 1992), with specific roles and responsibilities assigned to individual crew members. Given the team context of pilot decision making, it is tempting to suspect that role-consistent categories are the result of a specialization in pilot expertise, analogous to the specialization in medicine. Resolving this issue, however, will require an understanding of the cognitive mechanisms and processes underlying role-consistent categorization .

Specialization in expertise has so far been associated with the acquisition of specialized domain knowledge. In addition to a certain stock of knowledge that is common to all domain experts, specialists are believed to have "certain crucial components of necessary knowledge" that is missing in other experts (Patel & Groen, 1991; p.118). For instance, physicians were found to rely on specialized medical knowledge when they reasoned about cases in their area of specialization, and on generic medical knowledge when they diagnosed cases outside their specialty area (Patel & Groen, 1991).

Presence or absence of knowledge, in contrast, seems an unlikely explanation for the role-consistent categorizations by captains, first officers and second officers in our studies. As the directed sorting task demonstrated, captains not only could categorize decision events by those situational aspects that were normally used by first and second officers, but the first and second officers could classify situations according to risk just as the captains did.

The strong agreement among all three pilot groups in the directed sorting task indicates that pilots have comparable knowledge of decision-relevant components. While this result contradicts the view that role-specific categorization resulted from differences in what crew members knew, it does not preclude other knowledge-based accounts. For example, our findings could be explained in terms of knowledge structures that differentially support information retrieval. Frederick (1991) reports that expert auditors organize their internal controls knowledge into both taxonomic and schematic representations which enable fast and flexible access to relevant information.

Instead of being supported by multiple representations, knowledge retrieval can also be enhanced by processes such as proceduralization; i.e., by processes that rewrite knowledge into specialized procedures (cf. VanLehn's [1989] discussion of learning mechanisms during problem solving). A similar line of reasoning could be applied to our findings. According to this account, crew members spontaneously categorized decision events on the basis of features consistent with their particular crew role because their professional functions fostered specialized knowledge structures that made these aspects most easily accessible. Crew members' behavior in the directed sorting task would result from common but inert knowledge that needed specific prompting, such as our instructions, to be retrieved.

Alternatively, role-consistent categorization could also come from attentional processes. This argument builds on studies showing how expertise in domains such as sport is related to attentional processes efficiently adapted to task requirements. One characteristic of expert athletes is their ability to attend to certain domain-specific cues, enabling fast and highly accurate predictions concerning type, placement, direction and force of opponents' passes or serves (Abernethy, 1993). Anticipating an opponent's behavior quickly and precisely, in turn, allows a player to prepare for a fast response. In a similar vein we could argue that pilots' categorizations in the undirected sorting task reflected role-consistent perceptions that support particular crew functions. That is, the captains focused on risk and the first and second officers focused on response determinacy as a basis for sorting because each pilot group had learned to attend to those elements in routine problem situations. Those factors are particularly relevant to fulfilling their respective duties and responsibilities (cf. Allport's [1989] discussion of attentional selectivity as selection-for-action).

Explaining crew members' differing categories in terms of selective attention to job-relevant components does not presuppose differences in knowledge or knowledge structures, nor does it necessarily entail a specialization of pilot expertise. In fact, an attention-based account of our findings accords with either assumption, that role-consistent categories are a sign of specific expertise, or that they are a generic feature of pilot expertise. The specific expertise position would make the following argument: Experience in a particular crew function increases the relevance and thus the salience of certain aspects of problem events. In the undirected sorting task pilots classified decision events by role-consistent components because they viewed the events from an acquired role-specific perspective.

The generic expertise position, in contrast, would claim that a significant aspect of pilot expertise consists of using knowledge flexibly and selectively to accommodate given task requirements. According to this position, pilots do not acquire a position-specific perspective; rather pilots' spontaneous categories were role-consistent because they simply assumed the perspective of their regular crew role. However, if instructed to take on a different role in an otherwise undirected sorting task, they could shift their perspective in accordance with the pretended role and their categories would be consistent with this new role rather than with their actual crew position. However, future research will have to decide which position holds if it were found that role-consistent categories are attention-based rather than knowledge-based.

## General Conclusion

Our findings contribute to an understanding of three issues that motivated this research. First, as has been noted for experts in other domains, professional pilots categorized aviation problems on the basis of fundamental principles rather than surface features. Second, differences were observed in pilots' categorizations that are consistent with Patel and Groen's (1991) notion of specific expertise. While we found that all three groups of pilots shared knowledge of decision-relevant components, their spontaneous structuring of decision events reflected role-specific responsibilities on the flight deck. Future work will have to determine whether these differences result from a specialization of knowledge structures or from acquired role-specific perceptions. Third, our findings validated three of the four components of a previously developed model of aviation decision making. Role-consistent use of the three components by pilots suggests that alternative perspectives on problems may in fact enhance the collective expertise of flight crews in threatening problem situations.

## References

Abernethy, B. (1993). Attention (pp. 127-170). In R. N. Singer, M. Murphey & L. K. Tennant (Eds.), *Handbook of research on sport psychology*. New York, NY: MacMillan Publishing Company.

Adelson, B. (1981). Problem solving and the development of abstract categories in programming languages. *Memory and Cognition*, *9*, 422-433.

Allport, A. (1989). Visual attention (pp. 631-682). In M. I. Posner (Ed.), *Foundations of cognitive science*. Cambridge, MASS: The MIT Press.

Bakeman, R., & Gottman, J. M. (1986). *Observing interaction: An introduction to sequential analysis*. Cambridge: Cambridge University Press.

Baillet, S. D., & Keenan, J. M. (1986). The role of encoding and retrieval processes in the recall of text. *Discourse Processes*, *9*, 247-268.

Bédard, J. C., & Biggs, S. F. (1991). The effect of domain-specific experience of management representations in analytical procedures. *Auditing*, *10* , Supplement, 78-90.

Bédard, J., & Chi, M. T. H. (1993). Expertise in auditing. *Auditing*, *12*, Supplement, 21-45.

Bouwman, M. J. (1982). The use of accounting information: Expert versus novice behavior. In G. R. Ungson, & D. N. Braunstein (Eds.), *Decision making: An interdisciplinary inquiry* (pp. 134–167). Boston, Mass: Kent Publishing Company.

Bryson, M., Bereiter, C., Scardemalia, M., & Joram, E. (1991). Going beyond the problems as given: Problem solving in expert and novice writers. In R. J. Sternberg & P. A. Frensch (Eds.), *Complex problem solving: Principles and mechanisms* (pp. 61–84). Hillsdale, NJ: Erlbaum.

Cannon-Bowers, J. A., Salas, E., & Converse, S. (1990). Cognitive psychology and team training: Training shared mental models of complex systems. *Human Factors Society Bulletin*, *33*, 1-4.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215-281). New York: Academic Press.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121-125.

Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence: Vol. 1* (pp. 17-76). Hillsdale, NJ: Erlbaum Associates.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational Psychological Measurement*, *20*, 37-46.

Cooke, N. J. (1992). Modeling human expertise in expert systems. In R. R. Hoffmann (ed.), *The psychology of expertise: Cognitive research and empirical AI* (pp. 29-60). New York: Springer.

De Groot, A. D. (1965). *Thought and choice in chess*. The Hague: Mouton.

Egan, D. E., & Schwartz, B. J. (1979). Chunking in recall of symbolic drawings. *Memory and Cognition*, *7*, 149-158.

Endsley, M. R. (1995). A taxonomy of situation awareness errors. In R. Fuller, N. Johnston, & N. McDonald (Eds.), *Human*

*Factors in Aviation Operations* (pp. 287-292). Aldershot, England: Avebury Aviation, Ashgate.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*,3, 363-406.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, *47*, 273-305.

Federico, P-A. (1995). Expert and novice recognition of similar situations. *Human Factors*, *37* (1), 105-122.

Flin, R., Salas, E., Strub, M. & Martin, L. (Eds.) (1997). *Decision making under stress: Emerging themes and applications*. Aldershot, England: Avebury Aviation, Ashgate Publishing.

Frederick, D. M. (1991). Auditors' representation and retrieval of internal control knowledge. *The Accounting Review*, *66*, 240-258.

Gibson, J., Orasanu, J., Villeda, E., & Nygren, T. E. (1997). Loss of situation awareness: Causes and consequences. In R. Jensen (ed.), *Proceedings of the Ninth International Symposium on Aviation Psychology* (pp. 1417-1422). April 23, 1997, Ohio State University, Columbus, OH.

Glaser, R. & Chi, M. T. H. (1988). Overview. In M. T. H. Chi, R. Glaser & M. J. Farr (Eds.), *The nature of expertise* (pp. xv-xxviii). Hillsdale, NJ: Erlbaum.

Hackman, J. R. (1993). Teams, leaders, and organizations: New directions for crew-oriented flight training. In E. L. Wiener, B. G. Kanki, & R. L. Helmreich (Eds.). *Cockpit resource management* (pp. 47-70). San Diego, CA: Academic Press.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MASS: The MIT Press.

Johnson, P. E., Kochevar, L. K. & Zualkernan, I. A. (1992). Expertise and fit: Aspects of cognition. In H. L. Pick Jr., P. van den Broek & D. C. Knill (Eds.), *Cognition*: *Conceptual and methodological issues* (pp. 305-331). Washington, DC: American Psychological Association.

Johnston, A.N. (1992). The development and use of a generic nonnormal checklist with applications in ab initio and introductory advanced qualification programs. *International Journal of Aviation Psychology*, 2, 323-337.

Klein, G. A. (1989). Recognition-primed decisions. In W. Rouse (Ed.), *Advances in man-machine systems research, 5* (pp. 47-92). Greenwich, CT: JAI.

Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decision making. In G. Klein, J. Orasanu, R. Calderwood, & C. Zsambok (Eds.). *Decision making in action: Models and methods* (pp. 138-147). Norwood, NJ: Ablex Publishers.

Klein, G., Orasanu, J., Calderwood, R., & Zsambok, C. E. (Eds.). (1993). *Decision making in action: Models and methods*. Norwood, NJ: Ablex Publishers.

Larkin, J. H. (1983). The role of problem representation in physics. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 75-100). Hillsdale, NJ: Erlbaum.

Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science*, *208*, 1335-1342.

Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing x-ray pictures. In M. T. H. Chi, R. Glaser & M. J. Farr (Eds.), *The nature of expertise* (pp. 311-342). Hillsdale, NJ: Erlbaum.

Lipschitz, R. & Strauss, O. (1996). How decision-makers cope with uncertainty. In *Proceedings of the Human Factors and Ergonomics Society 40th annual meeting*, 1, 189-193. Santa Monica: Human Factors and Ergonomics Society.

McKeithen, K. B., Reitman, J. S., Rueter, H. H., & Hirtle, S. C. (1981). Knowledge organization and skill differences in computer programmers. *Cognitive Psychology*, *13*, 307-325.

Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrica*, *50*, 2, 159-179.

Mojena R. (1977). Hierarchical grouping methods and stopping rules: An evaluation. *The Computer Journal*, *20*, 359-363.

National Transportation Safety Board. (1994). *A Review of Flightcrew-Involved, Major Accidents of U.S. Air Carriers, 1978 through 1990.* (PB94-917001, NTSB/SS-94/01). Washington, DC: Author.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.

Orasanu, J. (1995). Situation awareness: Its role in flight crew decision making. In R. Jensen (Ed.), *Proceedings of the Eight International Symposium on Aviation Psychology*. Columbus, OH: Ohio State University.

Orasanu, J. (1994). Shared problem models and flight crew performance. In N. Johnston, N. McDonald, & R. Fuller (Eds.), *Aviation psychology in practice* (pp. 255-285). Hants, England: Avebury Technical.

Orasanu, J., Dismukes, R. K., & Fischer, U.(1993). Decision errors in the cockpit. In L. Smith (Ed.), *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting* (Vol. 1, pp. 363-367). Santa Monica, CA: Human Factors and Ergonomics Society.

Orasanu, J., Fischer, U., & Tarell, R. (1993). A taxonomy of decision problems on the flight deck. In R. Jensen (Ed.), *Proceedings of the Seventh International Symposium on Aviation Psychology* (pp. 226-232). Columbus, OH: Ohio State University

Patel, V. L., & Frederiksen, C. H. (1984). Cognitive processes in comprehension and knowledge acquisition by medical students and physicians. In H. G. Schmidt & M. C. DeVolder (Eds.), *Tutorials in problem based learning* (pp. 143-157). Assen, Holland: Van Gorcum.

Patel, V., L., & Groen, G. J. (1991). The general and specific nature of medical expertise: A critical look. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise* (pp. 93-125). Cambridge: Cambridge University Press.

Patel, V.L., Groen, G. J., & Arocha, J. F. (1990). Medical expertise as a function of task difficulty. *Memory and Cognition*, *18* (4), 394-406.

Rasmussen, J. (1985). The role of hierarchical knowledge representation in decision making and system management. *IEEE Transactions on Systems, Man and Cybernetics*, 2, (SMC - 15), 234-243.

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology*, *9* (4), 283-294.

Rosenbloom, P. S. & Newell, A. (1983). The chunking of goal hierarchies: A generalized model of practice. *Proceedings of the International Machine Learning Workshop*.

Salas, E., Dickinson, T. O., Converse, S. A. & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. W. Swezey & E. Salas (Eds.), *Teams: Their training and performance* (pp. 3-30). Norwood, NJ: Ablex.

Schiffman, S. S., Reynolds, M. L., & Young, F. W. (1981). *Introduction to Multidimensional Scaling*: *Theory*, *methods*, *and applications*. New York: Academic.

Schmidt, H. G., & Boshuizen, H. P. A. (1993). On acquiring expertise in medicine. *Educational Psychology Review*, *5*, 3, 205-221.

Schoenfeld, A. H. (1983). Episodes and executive decisions in mathematical problem solving. In R. Lesh & M. Landau (Eds.), *Acquisition of mathematics concepts and process*. New York: Academic Press.

Schoenfeld, A. H., & Herrmann, D. J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. *Journal of Experimental Psychology*: *Learning*, *Memory and Cognition*, *8*, 484-494.

Shanteau, J. (1988). Psychological characteristics and strategies of expert decision makers. *Acta Psychologica*, *68*, 203-215.

Soloway, E., Adelson B, & Ehrlich , K. (1988). Knowledge and processes in the comprehension of computer programs. In M. T. H. Chi, R. Glaser & M. J. Farr (Eds.), *The nature of expertise* (pp. 129-152). Hillsdale, NJ: Erlbaum.

VanLehn, K. (1989). Problem solving and cognitive skill acquisition. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 527-579). Cambridge, MASS: The MIT Press.

Voss, J. F., Wolfe, C. R., Lawrence, J. A., & Engle, R. A. (1988). From representation to decision: An analysis of problem solving in international affairs. In R. J. Sternberg & P. A. Frensch (Eds.), *Complex problem solving*: *Principles and mechanisms* (pp. 119-158). Hillsdale, NJ: Erlbaum.

## Footnotes

[1] Commercial pilots are hired only after they have achieved a minimum of 1,000 hours of flying experience and are certified for multiengine instrument flight. In fact, most new hires by US carriers far exceed the minimum standards for experience, having achieved substantial experience as military or civilian fliers. However, other carriers around the globe train pilots *ab initio*, or “from scratch” (Johnston, 1992/?).

[2] According to Federal Aviation Regulations, Part 121 air transport refer to aircraft that carry more than 30 passengers. Airlines that operate aircraft of this category are subjected to specific FAA requirements concerning aircraft equipment, maintenance and dispatch facilities, as well as crew experience and training.

[3] The Aviation Safety Reporting System is a NASA program to which pilots, air traffic controller, and other aviation personnel can voluntarily submit reports describing aviation incidents such as unsafe operations and hazardous situations. The ASRS processes and analyzes these accounts and uses the resulting information to remedy aviation safety problems.

[4] An Individual Difference Scaling (INDSCAL) was not performed on the data since we could not assume a priori, as presupposed by this analysis, that the same dimensions will be discerned for all three subject groups, albeit with different weights.

[5] Cohen's Kappa (Cohen, 1960) is a more conservative measurement of interrater reliability than percent agreement since it takes chance agreement into account. Values of .70 or larger are considered sufficiently high as to rule out that agreement was due to chance (Bakeman & Gottman, 1986).

[6] Recall that subjects provided multidimensional descriptors for a given scenario. Thus the total number of classifications is greater than 22; i.e., the number of scenarios used.

[7] Part 135 operations involve civil aircraft carrying less than 30 passengers.

# Appendix

## Complete List of Scenarios

| | |
|---|---|
| 1 | Shortly after departure from HNL on a flight to LAX, a wide body aircraft experiences a pneumatic overheat. The flight crew perform the appropriate checklist but fail to isolate the pneumatic system. The plane is above environmentally sensitive waters. |
| 2 | Coming upon a line of thunderstorms (with tops to approx 20,000’), the crew of a large transport request a deviation to the west to avoid the problematic weather. Their request is denied and they are told to deviate east; according to their aircraft’s radar, this would put them on a direct course for the thunderstorms. An aircraft in front of them on the arrival requests and is granted a deviation to the west (approx heading of 200 degrees). By this time, they (on the large transport) are descending through FL195 and are within 5 miles or so of the line of storms. |
| 3 | A medium large transport takes off from MIA enroute to EWR. Upon retraction of the landing gear, the nose gear light illuminates. The crew cycle the gear twice, but are unable to extinguish the light. |
| 4 | While enroute from LAX to HNL, shortly after the EPT (Equal Point Time), the crew of a wide-bodied transport notice a discrepancy between the fuel gauge reading and the computed fuel used. Actual fuel loss is confirmed by a crew member’s subsequent visual sighting. |
| 5 | A large transport is flying a scheduled flight from DCA to DFW. At flight level 310, the crew discover that the crew oxygen supply is indicating 200 lbs. instead of the required 1200 lbs. Weather on the route to LIT and the Blue Ridge arrival into DFW is bad enough that they are given a reroute to JAN and the Scurry arrival. |
| 6 | A large transport, on a flight from DFW, begins a visual approach at RNO. Turning crosswind, 2 deg flaps are selected, but the inboard flap indicator shows a split: L=0 degree extension, R=2 degrees. To the crew the aircraft “feels” like both sides are extended because there are no control problems. The crew attempt to position the flaps back up; but the indicator continues to show a split (L=0, R=2). |
| 7 | On the taxi out from the gate, the lower rudder of a large transport shows only partial movement during the flight control check. All other flight controls check normal. |
| 8 | A large transport is preparing to depart DCA for a flight to MSP. The flight is running late and the DCA curfew is nearing. After push back, it is found that there are 185 passengers aboard, but only 184 seats. |
| 9 | Immediately after takeoff, the crew of a medium large transport leaving DCA realize the aircraft is not pressurizing properly. The pressurization controller is functioning normally, but the aircraft will not pressurize in standby or manual. After several minutes, the #1 flight attendant reports air leakage from the forward entry door. |
| 10 | At around 40 minutes out of TYS, the crew of a large transport receive an indication that the #1 VF generator is off-line. Shortly thereafter, the #1 accessory gear box light illuminates. |
| 11 | A medium large transport is in cruise at flight level 310, westbound over the North Atlantic, when the crew discover that the crew oxygen bottle is reading 800 psi. The bottle had been changed in Frankfurt due to an incoming crew write-up and had been definitely noted as full. The crew shut off the crew regulators to see if a leak point can be determined; none is found but the supply decreases to 750 psi. |

| | |
|---|---|
| 12 | A medium large transport is cleared for takeoff. The thrust levers are put approx 3/4” forward of idle to allow the engines to accelerate and stabilize at the initial thrust setting (approx 40% N1) per current company policy. Once stabilized, the crew attempt to further advance the thrust lever to the takeoff thrust setting. As they do so, the takeoff warning horn sounds intermittently. |
| 13 | The crew of a small transport find nothing out of the ordinary during a preflight walk-around. Then while inside the plane awaiting clearance, they feel the plane move slightly. They immediately look outside and see a big cart wheel near the airstair door. When they ask the driver whether he hit the airplane, he says he may have struck the door but he doesn't know. The crew inspect the side of the aircraft and test the door but find nothing unusual. |
| 14 | At FL260 smoke from an unknown source appears in the cockpit, followed by flight attendant reports of heavy smoke filling the cabin. |
| 15 | An aircraft fueler advises the crew of a medium large transport that he can’t get the numbers to balance between the fuel off loaded from the truck and the fuel on the aircraft. He works on it for an hour and gets nowhere. |
| 16 | The crew of a medium large transport abort a takeoff because of a slight EPR fluctuation on the #2 gauge. They decide to get the plane checked for problems. The engine is run to approx 1.8 EPR at the pad and all indications are normal. The crew attempt another takeoff. At 80 knots, they notice small fluctuations in the N1 gauge for #2. |
| 17 | A medium large transport is cleared by DFW approach control to descend to 3000' and intercept the 13R ILS. Clearance is also given to contact the tower. On the approach when contact cannot be established via the frequency set on the VHF Com, the crew dial in the frequency for the west tower on the approach plate. Several attempts are made to contact the tower but due to the extreme congestion on the frequency, no landing clearance is obtained. |
| 18 | According to its logbook, a medium large transport had not been flown for 10 days; during that time maintenance was done and a new interior put in. On an approach at BUR the right main gear “unsafe” red light comes on and stays on during gear extension with no green down and locked light. At 8-10 miles out, flight level 3000-4000’, 180 knots, IAS with 5 degree flaps, the crew recycle the gear but still get the same result. |
| 19 | After all the passengers board a medium large transport at CHA, a ramp agent reports to the flight crew that the forward cargo door will not lock closed. CHA is a non-maintenance base. A mechanic contracted from the airfield, however, is able to get the cargo door latch mechanism to function marginally--i.e., it can only be closed/opened once and the “Fwd Cargo Door” light stays on. |
| 20 | A wide bodied aircraft weighing 43,000 lbs. experiences a #3 engine compressor stall at 138 knots on takeoff; V1 is 142 knots. |
| 21 | A medium large transport is cleared for takeoff on runway 13C. Power is advanced and takeoff roll commences with the crew setting takeoff power. At approximately 100kts, a crew member notices a light twin engine taxiing at a fairly high speed toward 13C. At the same time, he hears the tower controller making repeated unsuccessful attempts to contact the small plane. |
| 22 | A medium large transport climbing through FL280 at .82 mach experiences a high frequency flutter through the air frame. Upon visual inspection, it is discovered that the right outboard aileron balance tab is fluttering 1-2” up and down. The flutter stops some time during deceleration to .74 mach and is stable on second inspection. A company mechanic on board indicates there are no visual signs of delamination or structural failure. |

## Report Documentation Page

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
April 2003

3. REPORT TYPE AND DATES COVERED
Technical Memorandum

4. TITLE AND SUBTITLE
Do You See What I See? Effects of Crew Position on Interpretation of Flight Problems

5. FUNDING NUMBERS
199-06-12-37

6. AUTHOR(S)
Ute Fischer and Judith Orasanu

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
NASA Ames Research Center
Moffett Field, California 94035-1000

8. PERFORMING ORGANIATION
REPORT NUMBER
IH-014

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
National Aeronautics and Space Administration

10. SPONSORING/MONITORING
AGENCY REPORT NUMBER
NASA/TM—2003–209612

11. SUPPLEMENTARY NOTES
Point of Contact: Judith Orasanu, M/S 262-4, Ames Research Center, Moffett Field, CA 94035
(650-604-3404

12A. DISTRIBUTION/AVAILABILITY STATEMENT
Subject Category: 53-01 Distribution: Public
Availability: NASA CASI (301) 621-0390

12B. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The set of experiments reported in this paper is concerned with whether pilots who differ in levels of experience and job-specific roles also differ in their interpretations of situations that demand judgments about flight-related problems. In addition to this rather applied question, these studies also address more fundamental issues concerning differentiation of expertise within a rich domain. In a series of sorting studies involving aviation incidents, we compared the judgments of captains, first officers and second officers. Multidimensional scaling analyses indicated that these three groups framed aviation incidents differently. In making a decision, pilots focused on those aspects of the situation that were consistent with responsibilities associated with their particular crew roles. Moreover, differences in feature selection were found to reflect the effect of crew role, not years of aviation experience or absence of relevant knowledge.

14. SUBJECT TERMS
Flight crews, Group dynamics, Psychological aspects of human behavior

**15. NUMBER OF PAGES**
40

**16. PRICE CODE**

**17. SECURITY CLASSIFICATION OF REPORT**
Unclassified

**18. SECURITY CLASSIFICATION OF THIS PAGE**
Unclassified

**19. SECURITY CLASSIFICATION OF ABSTRACT**
Unclassified

**20. LIMITATION OF ABSTRACT**
Unlimited

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z-39-18
298-102