AAMRL-TR-88-017

COMPONENTIAL ANALYSIS OF PILOT DECISION MAKING (U)

CHRISTOPHER D. WICKENS
ALAN STOKES
BARBARA BARNETT
TOM DAVIS, Jr.

UNIVERSITY OF ILLINOIS AVIATION RESEARCH LABORATORY

AUGUST 1988

PERIOD OF PERFORMANCE: JUNE 1986 - SEPTEMBER 1987

Approved for public release; distribution is unlimited.

ARMSTRONG AEROSPACE MEDICAL RESEARCH LABORATORY
HUMAN SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433-6573

NOTICES

Please do not request copies of this report from the Armstrong Aerospace Medical Research Laboratory. Additional copies may be purchased from:

National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22161

Federal Government agencies and their contractors registered with the Defense Technical Information Center should direct requests for copies of this report to:

Defense Technical Information Center
Cameron Station
Alexanuria, Virginia 22314

TECHNICAL REVIEW AND APPROVAL

AAMRL-TR-88-017

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

The voluntary informed consent of the subjects used in this research was obtained as required by Air Force Regulation 169-3.

This tecnical report has been reviewed and is approved for publication.

FOR THE COMMANDER

CHARLES BATES, JR.
Director, Human Engineering Division
Armstrong Aerospace Medical Research Laboratory

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

AD-A [illegible] 711

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AAMRL-TR-88-017
6a. NAME OF PERFORMING ORGANIZATION University of Illinois Aviation Research Laboratory* 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Harry G. Armstrong Aerospace Medical Research Laboratory/HEG
6c. ADDRESS (City, State, and ZIP Code) One Airport Road Savoy IL 61874	7b. ADDRESS (City, State, and ZIP Code) Wright-Patterson AFB OH 45433-6573
8a. NAME OF FUNDING/SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F33615-85-D-0514, Task Nr. 0006
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO
62202F	7184	14	23

11. TITLE (Include Security Classification)
Componential Analysis of Pilot Decision Making (U)

12. PERSONAL AUTHOR(S)
Wickens, Christopher D., Stokes, Alan, Barnett, Barbara and Davis, Tom Jr.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM Jun 86 TO Sep 87	1988 August	93

16. SUPPLEMENTARY NOTATION
*Subcontract to Southeastern Center for Electrical Engineering Education, Central Florida Facility, 1101 Massachusetts Avenue, St Cloud FL 32769

17. COSATI CODES

FIELD	GROUP	SUB-GROUP
05	08	
23	02	

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Decision Making
Pilot Performance
Simulation
Aviation Research

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This report describes the development of a computerized pilot decision-making simulator/trainer known as MIDIS, and its utilization as a research tool in the validation of an information-processing model of pilot decision making. Efforts in this project followed two parallel but interacting tracks: development of decision scenarios for the MIDIS program, following the sequence of a realistic IFR flight, and compilation of a cognitive test battery, based on an information processing model of decision making, and designed to assess individual differences in those cognitive attributes determined to be important in effective decision making.

Subjects consisted of thirty eight instrument rated pilots subdivided into two groups on the basis of reported hours of flight experience. The experiment consisted of four parts: administration of the cognitive test battery, pre-flight planning, a practice flight, and the actual MIDIS run. Subjects were scored as to the optimality and latency of their choices, and their rated confidence. (Continued)

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Gary B. Reid	22b. TELEPHONE (Include Area Code) (513) 255-8749 22c. OFFICE SYMBOL AAMRL/HEG

Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

BEST AVAILABLE COPY

19. Abstract (Cont'd)

The results indicated that low and high experienced pilots did not differ from each other in terms of their judgment performance, but that high experienced pilots expressed slightly greater confidence in their decisions. ~~Both groups became equally overconfident on their responses to more difficult decision problems.~~ The two groups however did differ in terms of what problem variables degraded decision performance, and what individual abilities affected that performance. In particular, novice decision performance was partially predicted by information processing tests related to spatial abilities, working memory capacity, mathematical ability and by tests of declarative knowledge. ~~However, these tests had little predictive abilities for the more experienced pilots.~~ The implications for future research that focuses on capturing this source of prediction of experienced pilot judgment are discussed. (SCW)

BEST AVAILABLE COPY

SUMMARY

This report describes the development of a computerized pilot decision-making simulator/trainer known as MIDIS, and its utilization as a research tool in the validation of an information-processing model of pilot decision making. Efforts in this project followed two parallel but interacting tracks: development of decision scenarios for the MIDIS program, following the sequence of a realistic IFR flight, and compilation of a cognitive test battery, based on an information processing model of decision making, and designed to assess individual differences in those cognitive attributes determined to be important in effective decision making.

Subjects consisted of thirty eight instrument rated pilots subdivided into two groups on the basis of reported hours of flight experience. The experiment consisted of four parts: administration of the cognitive test battery, pre-flight planning, a practice flight, and the actual MIDIS run. Subjects were scored as to the optimality and latency of their choices, and their rated confidence.

The results indicated that low and high experienced pilots did not differ from each other in terms of their judgment performance, but that high experienced pilots expressed slightly greater confidence in their decisions Both groups became equally overconfident on their responses to more difficult decision problems. The two groups however did differ in terms of what problem variables degraded decision performance, and what individual abilities affected that performance. In particular, novice decision performance was partially predicted by information processing tests related to spatial abilities, working memory capacity, mathematical ability and by tests of declarative knowledge. However, these tests had little predictive abilities for the more experienced pilots. The implications for future research that focuses on capturing this source of prediction of experienced pilot judgment are discussed.

BEST AVAILABLE COPY

PREFACE

This work was performed under subcontract No. SCEEE-HER/86-6 on contract F33615-85-D-0514. The work has been carried out at the Aviation Research Laboratory, University of Illinois, Institute of Aviation. Presentations based upon this work have been made to Association for the Development of Computer Based Instructional Systems in Washington DC, to the Ohio State University Conference on Aviation Psychology, to the Division 21 meeting of the American Psychological Association, and to the meeting of the Human Factors Society.

The authors would like to acknowledge the invaluable efforts of Rob Rosenblum in programming MIDIS, and of Tak Ming Lo in programming SETSCENE, the assistance of Dr. Fred Hyman, a flight instructor in coding the attributes, the helpful editorial comments of Gary Reid, and the word processing skills of Karen Ayers.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1 Overview

Engineering Psychology has provided a number of useful and sophisticated models of human performance in certain aviation-relevant areas. Most prominent among these concerns is the extensive work that has been done on modeling manual control. The programmatic efforts to develop the quasi-linear Crossover Model undertaken by McRuer, Jex, and their colleagues (McRuer, 1980; McRuer & Jex, 1967), and the efforts of Levison (1982) to develop the Optimal Control Model have both achieved a great deal of success in predicting quantitatively how human manual control performance can be modeled, will degrade under stress, and will improve with training. While it is anticipated that manual control will continue to be a critical component in aviation, with increasing aircraft sophistication the pilot is now called upon to become more and more a systems manager and executive decision maker. Certainly a pilot's judgment and decision making abilities are critical to air safety. Analysis of FAA aircraft reports by Jensen and Benel (1977) suggested that errors in pilot judgment accounted for over 50% of pilot fatalities during the period from 1970-74. Yet despite this importance, pilot decision making has received only a minimum degree of research interest (for exceptions see Buch & de Bagheera, 1985; Buch & Diehl, 1984; Jensen, 1981; Lester, Diehl, & Buch, 1985). Nor has pilot judgment benefited from the sophisticated modeling approaches characteristic of the manual control field. This neglect is even more surprising in light of the growing amount of solid theory-based research in the psychology of decision and choice (see Einhorn & Hogarth, 1981; Kahneman, Slovic, & Tversky, 1982, Pitz & Sachs, 1984; for recent reviews), and the limited understanding of decision making on the flight deck (Stone, Babcock, & Edmunds, 1985).

The first section of this report will focus upon conclusions regarding human strengths and limitations in decision making that have been drawn from general research. Where possible these factors will be illustrated in the framework of aviation-related tasks, but for the most part their actual investigation in an aviation context has not been carried out. The second section describes a pilot decision making simulation system known as MIDIS used for validating an information processing model of pilot judgment, and a cognitive test battery developed for the purposes of testing individual differences in that model. The third section describes an experiment in which the relation between the model, the battery, and the simulation is validated.

1.2 Pilot Decision Making

There are three general characteristics that define the decision-making paradigm. First, the pilot must evaluate several sources of information in assessing the situation, or understanding the current state of the "world." This assessment forms the basis for choosing an appropriate action. Second, the information the pilot deals with is probabilistic. The cues used for situation assessment may be unreliable (e.g., a weather forecast predicts a 20% chance of thunderstorms), and the projected consequences of an action into the future are uncertain. This probabilistic element means that the right decision can often produce an unfortunate outcome ("bad luck") and the wrong decision can often "luck out." Third, the elements of value and cost underlie most decisions. For example, the pilot may have to balance the benefit of continuing a flight through bad weather and satisfying the passengers' need to reach their destination on time, against the potential greater cost of an accident.

Figure 1 presents a general model of human decision making that highlights the information processing components which are relevant to decision-making. To the left of the figure, environmental cues are sampled to

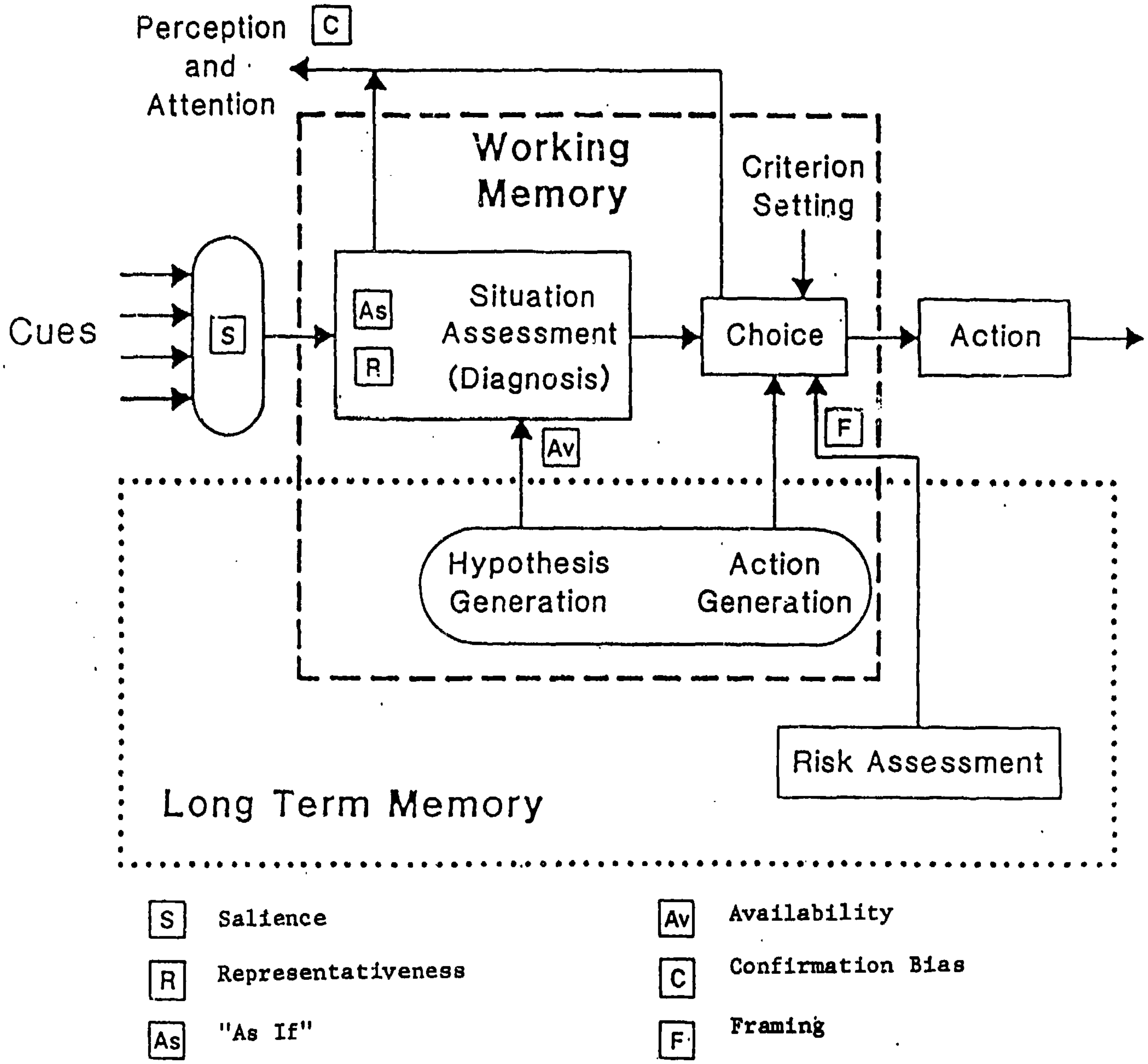


Figure 1. An information processing model of pilot decision making. Biases and heuristics are surrounded by a square and coded at the bottom of the figure.

obtain a "situation assessment" or diagnosis of the state-of-the-world that calls for a decision. An accurate assessment often requires perception of a large number of cues--radar pictures, weather forecasts, visual topographic features, fuel consumption, engine status, airport capabilities and so forth. These cues in turn must be interpreted against a knowledge base in long-term memory to accurately construct a mental model or diagnosis of the situation. Possible alternative hypotheses that describe the situation are generated from long term-memory, held in working memory, and compared against the cues. As we shall see, this construction process is hampered both by limits of attention (are relevant cues processed?), and by biases in long term memory.

Assuming that the assessed situation is identified as a problem that requires some action, the pilot must then generate plausible alternative courses of action to take. For example, the pilot may ask, "Do I continue my approach, fly around while seeking more information, or turn back to an alternate airport?" Each proposed course of action may have a different anticipated set of possible outcomes, depending upon the diagnosed state-of-the-world. Furthermore, all of these outcomes will have potential values associated with them (or costs, which may be defined as negative values). The expected value of an outcome is its true value multiplied by the probability that it might occur. When values cannot be expressed in monetary terms they are called utilities. The pilot's choice or decision should be that which produces the most positive, or least negative expected utility. As indicated in the figure, this critical choice point involves the process of risk assessment--the subjective evaluation of the probability of different outcomes--and the assessment of the utilities of these outcomes, as this information is retrieved from long term memory.

Finally, the figure indicates that at any stage these operations may trigger the decision to seek more information in order to better assess the situation and/or evaluate the consequences of an action. In the following pages, we shall outline some of the behavioral findings with regard to human strengths and limitations in this interactive process. The squared letters within the figure indicate particular sources of bias or "heuristics" that will be identified and discussed as the elements of the decision process are described in more detail below.

1.2.1 Situation assessment. In setting the stage for our discussion of cue perception and situation assessment, it is appropriate to consider two different aviation scenarios. In the first, a pilot flying IFR has become completely disoriented. Not only are glimpses of the now unfamiliar terrain below intermittent and cloud obscured, but the navigational information itself may be unreliable because of a suspected instrument malfunction. The situation to be assessed is "Where am I?" In the second scenario, the pilot senses, through a series of warning indicators and gauge readings, that one of the engines may be malfunctioning, but the nature of the malfunction is not an obvious one with which he is familiar. The situation to be assessed here is the diagnosis of what is wrong with the engine.

In situation assessments such as these, psychologists have found that problem solvers and trouble shooters often engage in "heuristics" or mental rules of thumb that are intended to reach a diagnosis without expending too much mental effort (Kahneman, Slovic, & Tversky, 1982; Rasmussen, 1981). While such heuristics often work adequately, the fact that they are shortcuts may prevent the decision-maker from obtaining the most accurate information. They may, therefore, sometimes lead the decision-maker to a false understanding. These sources of bias and error in situation assessment will be the focus of the following discussion.

1.2.1.1 Cue seeking. Searching the environment for critical cues in the first stage of situation assessment is limited by characteristics of human attention. It is apparent, for example, that decision makers do not necessarily process all of the information that is available to them (Wickens, 1984), particularly under time stress (Wright, 1974). Given that stress often causes a "tunneling of attention" when monitoring multi-element displays (Hockey, 1970), it is reasonable to assume that this tunneling would have the same restricting influence on the processing of multiple cues to assess the situation. For example, in attempting to diagnose a faulty engine, a pilot may focus on only a small number of physically salient symptoms, ignoring critical cues that might provide even more diagnostic information concerning the nature of the fault (such perceptual tunneling has been considered responsible, in part, for the disaster of Three Mile Island). This bias toward salience, at the expense of information content is indicated by the S in Figure 1.

Research has also found that the decision maker's cue seeking behavior is heavily guided by the hypothesis that may already have been tentatively chosen. This tendency, known as the confirmation bias (C in Figure 1), describes the bias to seek (and therefore find) those sources of information that confirm what we already believe to be true (Mynatt, Doherty, & Tweney, 1977; Wason & Johnson-Laird, 1972). Thus, the disoriented pilot who is trying to establish his location over the ground may first posit that he is in a certain location and then focus attention on ground features that are consistent with that location, while ignoring (or discounting) those that are inconsistent. As Wason and Johnson-Laird have noted, the best way to establish that a hypothesis is true is to seek information which, if found, will falsify the hypothesis rather than confirm it.

While, in general, people process only a limited number of sources of independent information when testing or confirming a hypothesis, these limitations are removed (or at least are greatly lessened) to the extent that the information sources are correlated. Thus, the skilled pilot can rapidly diagnose the current state of an aircraft from the six crucial instrument readings because of the typical pattern of correlation that is observed between these readings. For example, a positive rate-of-climb is correlated with an increase in altimeter reading; a change in attitude predicts a change in heading and so forth. In the same way, extensive familiarity with the patterns of symptoms produced by particular aircraft malfunctions will allow the pilot to interpret rapidly the potentially large number of cues indicating their status. For example, the failure of the suction pump will cause a failure of gyro instruments (altitude and heading indicator), resulting in a correlated change in these two instrument readings.

One general characteristic of cue seeking and information integration is its apparent dependence upon facilities of selective and divided attention, particularly to the visual environment (Moray, 1986). This will represent an important component of our experimental approach.

1.2.1.2 Hypothesis formulation and testing. People typically try to understand a situation by matching in working memory the pattern of cues seen in the environment with a mental representation of the typical or representative pattern for a particular situation as recalled from long-term memory (R in Figure 1). We may think of this memorized pattern as a hypothesis of the proposed state. If the hypothesis matches the data, then the situation is diagnosed (Tversky & Kahneman, 1974). A limitation of this heuristic results from the fact that a particular pattern of cues may not be a perfectly diagnostic indicator of the true state of the world. For example, to the lost pilot, the 60° intersection of a freeway with a road below may be consistent with several different ground locations, just as a pattern of low oil pressure and high engine temperature could be symptomatic of any number of different engine failures.

To ensure an accurate diagnosis, the decision-maker should first think of a reasonable number of possible hypotheses, in order to make sure that as many situations are covered as possible. An extensive program of research by Gettys and his colleagues (summarized in Gettys, 1983) suggests however that faced with problem-solving situations, people generate only a small fraction of possible hypotheses (relative to the number of plausible ones) even as they remain overly confident that their list is exhaustive.

In the second place, those hypotheses that people do generate should be the most probable or likely ones. For example, suppose a pilot has formed two alternative hypotheses concerning the diagnosis of an electronic system failure, one of which occurs ten times more frequently than the other. In such a case, the pilot's initial hypothesis concerning the cause of the malfunction should indicate the more frequently occurring failure. Yet, people do not accurately use the probability or "base rate" frequency information to guide their choice in this way (Kahneman, Slovic, & Tversky, 1982). Instead, when generating the few hypotheses from memory, they use what is described as the _availability_ heuristic (Tversky & Kahneman, 1974). A hypothesis is considered most likely if it is most available in memory. However, the most available hypothesis in memory may not be the most probable, but rather the one that was most recently experienced, or the simplest one, since simple hypotheses are easier to remember than complex ones (Fontenella, 1983; Tversky & Kahneman, 1974).

There is a second sense in which people fail to use probability information appropriately in diagnosis, and this relates to _cue reliability_.

Clearly some cues are quite reliable: the visual sighting of a distinct ground landmark, or the smell of smoke in the cockpit. For others the reliability may be somewhat less: instrument readings, or views of the same landmarks through the haze. Still other cues may have a reliability that is at best marginal--a message spoken by another pilot heard through static, an instrument reading that is notoriously unstable, or the sense of vertical obtained through vestibular cues. Yet, when integrating a number of information sources that vary in their reliability, people follow what is sometimes referred to as the "as if" heuristic (_"As"_ in Figure 1, Wickens, 1984). In the extreme, this amounts to treating all information sources _as if_ they were of equal reliability or, to a lesser degree, failing to "devalue" those information sources of lower reliability to an extent that is optimal (Johnson, Cavanagh, Spooner, & Samet, 1973; Kahneman & Tversky, 1973; Schum, 1975).

Instead of using cue reliability as a basis for choosing their hypothesis people more often focus attention most heavily on those cues that are physically _salient_ (loud, bright, recent, centrally visible and easy to interpret; Wallsten & Barton, 1982), and those that are likely to confirm the hypothesis that was already tentatively formed (_S_ and _C_ respectively in Figure 1). If those cues, by chance or by design, also happen to be quite reliable, then the assessment of the situation will likewise be accurate, but if not, and their indicated diagnosis is wrong, then even the best-intended decision of what action to take may lead to disaster because it will be based on a faulty assessment of the world.

An important characteristic of all human information processing dealing with hypothesis entertainment and selection is the heavy dependence on the mental "workbench" of working memory (Baddeley & Hitch, 1974; Wickens, 1984).

The assessment and prediction of working memory strength represents another important component of our experimental validation.

1.2.2 Decision formulation. Once an assessment of the situation is made, a decision must then follow as to what action to take. Of course, the decision may simply involve the choice to seek still more information, as indicated by the top loop in Figure 1. In all cases, the decision maker should choose the course of action with the most favorable expected outcome-- the highest expected utility. Sometimes this course of action is simple, if the situation is diagnosed with certainty (I'm sure that my fuel is about gone), and there is no question about the best action (land in the nearest field below rather than going further). However, at other times the choice of possible actions is far less clear cut. This may either be because the situation assessment leaves some uncertainty to be resolved in the pilot's mind (There is a 80% chance that my fuel is gone, but because I haven't flown very far since I refueled, there is a 20% chance that my fuel gauge may be in error), or because the consequence of one's choice of actions cannot be predicated with certainty (If I try an emergency landing here, I believe my chances of survival are high but I am not certain).

Formally, this state of affairs may be represented in terms of the decision tree shown in Figure 2. In this example, two states of the world, with different subjective probabilities are shown across the top and two potential courses of action are shown down the sides. (Of course, in a real world decision problem, there may be a greater number of both states of the world and of potential actions.) The decision-maker should optimally assign probabilities to each state of the world as we have seen above. Each action then, when taken in the presence of one or the other states of the world, can generate one or more potential outcomes. In the case of the example here, the possible outcomes of a decision to land might be a safe landing in the nearby

State of the World (S.O.W.)

CHOICE OF ACTION	Fuel Gone (p = .80)	Fuel Available (p = .20)	
Continue to Airport	O_1: Disastrous Landing (U = -100, p = .60) (p* = .80 x .60 = .48) O_2: Safe Landing (U = -10, p = .40) (p* = .4 x .8 = .32)	O_3: Reach Airport (U = 0, p = 1.0) (p* = 1 x .2 = .20)	.48X(-100) + .32X(-10) + .2X(0) = -48 - 3.2 + 0 = -51.2
Emergency Landing in Field Below	O_4: Disastrous Landing (U = -80, p = .20) (p* = .20) O_5: Safe Landing (U = -10, p = .80) (p* = .80)		.2X(-80) + .8X(-10) = -16 - 8 = -24

p = Absolute probability of S.O.W.
p = Absolute probability of outcome.
p* = Contingent probability of outcome given the state of the world.

Figure 2. A decision tree representation of pilot decision making.

field with the unpleasant aspects of getting it out again, or a disastrous landing in the same place; a decision to continue might result in a safe flight to the final destination, or the potential disaster of running out of fuel, short of the field with a less feasible landing place. Each of these outcomes has a utility, a positive or negative consequence to the decision maker that can be assigned some relative value, and a probability, or expected frequency of occurrence. Together, the utility and the probability serve to define the risk, and the human should optimally choose that action with the lowest expected risk. Formally, the expected risk of an action is computed as the expected risk of each outcome--its utility times its probability--summed across actions. These calculations are shown to the right of Figure 2, in which it is clear that the emergency landing has the lowest expected risk, and hence is the decision that should be made. Here again, human performance has been found to be adversely affected by certain biases and limitations.

To begin with, even the basic rows and columns in the decision matrix may not be set up optimally. As we have noted, the diagnosis estimating the possibility of the possible system states may be in error. Because of the confirmation bias, the diagnosis will probably show a far greater confidence or estimated probability of the most likely hypothesis than is warranted. Secondly, Gettys (1983) has found that, as in hypothesis generation, people generate only a small fraction of the feasible problem-solving actions that may be appropriate in a given situation.

Even assuming that an adequate matrix is set up, arriving at an optimal decision still requires that the risks (probability x value) of the different outcomes be accurately assessed. Here again, experiments show that people are not skilled at assessing the probability of different outcomes and their resulting risks (Fischoff, 1977; Kahneman, Slovic, & Tversky, 1982; Slovic, 1984), although it is not entirely clear what kind of biases these problems will demonstrate. On the one hand, people clearly overestimate the frequency of very rare positive events (Pitz, 1965). This bias explains why gambling and lotteries are pursued--because the low probability payoffs are perceived as occurring more frequently than they do. On the other hand, peoples' estimates of the frequency of different kinds of unpleasant or negative events appear to be influenced very much by the availability heuristic described above (Tversky & Kahneman, 1974). Highly available events, because they are salient and well publicized are overestimated (fatal aircraft accidents fall into this category), while less salient ones are greatly underestimated (near misses, or non-fatal accidents; Slovic, 1984). Collectively, the effect of these biases on the decision matrix such as that shown in Figure 2 cannot be entirely predicted.

To this analysis, two further important findings should be added. The first is based on a general theory of choice, put forward by Tversky and Kahneman (1981) which describes the influence of problem framing. While the entire theory is relevant to the concept of risky decision-making, its most critical aspect for this discussion is the assertion that the choice between two actions, one a risk and the other a "sure thing," will depend very much upon whether the problem is framed as a choice between gains or between losses. Of course, in our critical analysis of pilot decision-making, the choice is often between losses. Here Tversky and Kahneman observe that people are biased to choose the risky loss rather than the certain loss even when the expected loss resulting from the former is greater. For example, consider the pilot who must choose between turning back in the face of potentially bad weather (with the certainty of missing a critical appointment and disappointing his passengers), and continuing on (with a chance of getting through safely and on time, but also a small chance of suffering a major disaster). The choice is clearly one between negatives: a sure loss versus an uncertain probability of disaster, and Tversky and Kahneman have shown that people have a bias to favor the risky choice. This risk-seeking tendency is reversed however when the choice is framed as one between gains, and here the "sure thing" alternative is favored. In the previous example, we might suppose that if the pilot could frame the same decision as one between ensuring that lives are saved (the option to turn back) and probably keeping an appointment (the option of going ahead), the bias would swing toward the "sure thing" turn-back option.

The second bias that is relevant in choosing actions is a well-documented tendency toward overconfidence in forecasting. In a general sense, people overestimate the likelihood that their predictions of the future will be correct. Here again, one may account for the "can do," or "it won't happen to me" bias of a pilot, choosing to undertake a risky option. Studied repeatedly by Fischoff (1977), this bias is accounted for by peoples' inherent dislike of uncertainty.

A general conclusion emerging from the previous section is the strong dependence of good judgment on the accurate, calibrated assessment of risk and probability. Hence, another major component of the experimental approach will focus on risk and probability assessment.

The previous section has focused on generic limitations that would be applicable across a wide variety of decision tasks. In addition, the flight environment highlights two specific characteristics that must be present for effective decisions: well developed spatial abilities, and a strong knowledge base of facts and information. Both of these will enter into our evaluation and prediction of pilot judgment.

2. METHODOLOGICAL APPROACH: THE MIDIS TASK

2.1 Logic of the Approach

Given the general background reviewed above, we propose to validate the model presented in Figure 1 as a tool for examining pilot judgment using the following logic. If effective pilot judgment in fact depends upon avoiding the biases and pitfalls encountered in Figure 1, then those individuals who possess processing characteristics that minimize those biases and limitations should make good decisions. Correspondingly, those individuals who are deficient in relevant attributes should perform poorly. But different decisions may place greater or lesser demands on different attributes. The decision to abort a takeoff following engine failure, for example, may involve processing just two cues of information, one's airspeed and position on the runway, but will require processing those cues in a rapid manner. But diagnosing an instrument failure may require integration of a large number of cues with less time pressure, but heavy reliance on working memory.

Hence each decision can be characterized by a "profile" of demanded attributes as shown in Figure 3; those decisions that have high demands on an attribute that is relevant to decision making should be performed poorly. Finally, each pilot will generate a corresponding "profile" of available attributes. We hypothesize that to the extent that a pilot profile of attribute strength matches (or exceeds) the decision profile, the decision will be fast and accurate. To the extent that a mismatch occurs, performance will be less optimal. An incorrect decision may be reached, or the correct decision may be made only after a long time. Hence, an interaction between pilot abilities and decision type is predicted. Given this characterization of pilot abilities and demand attributes, a second thrust of the study examines how decision performance differs as a level of pilot experience. Does performance simply improve? Or does it change in a qualitative fashion?

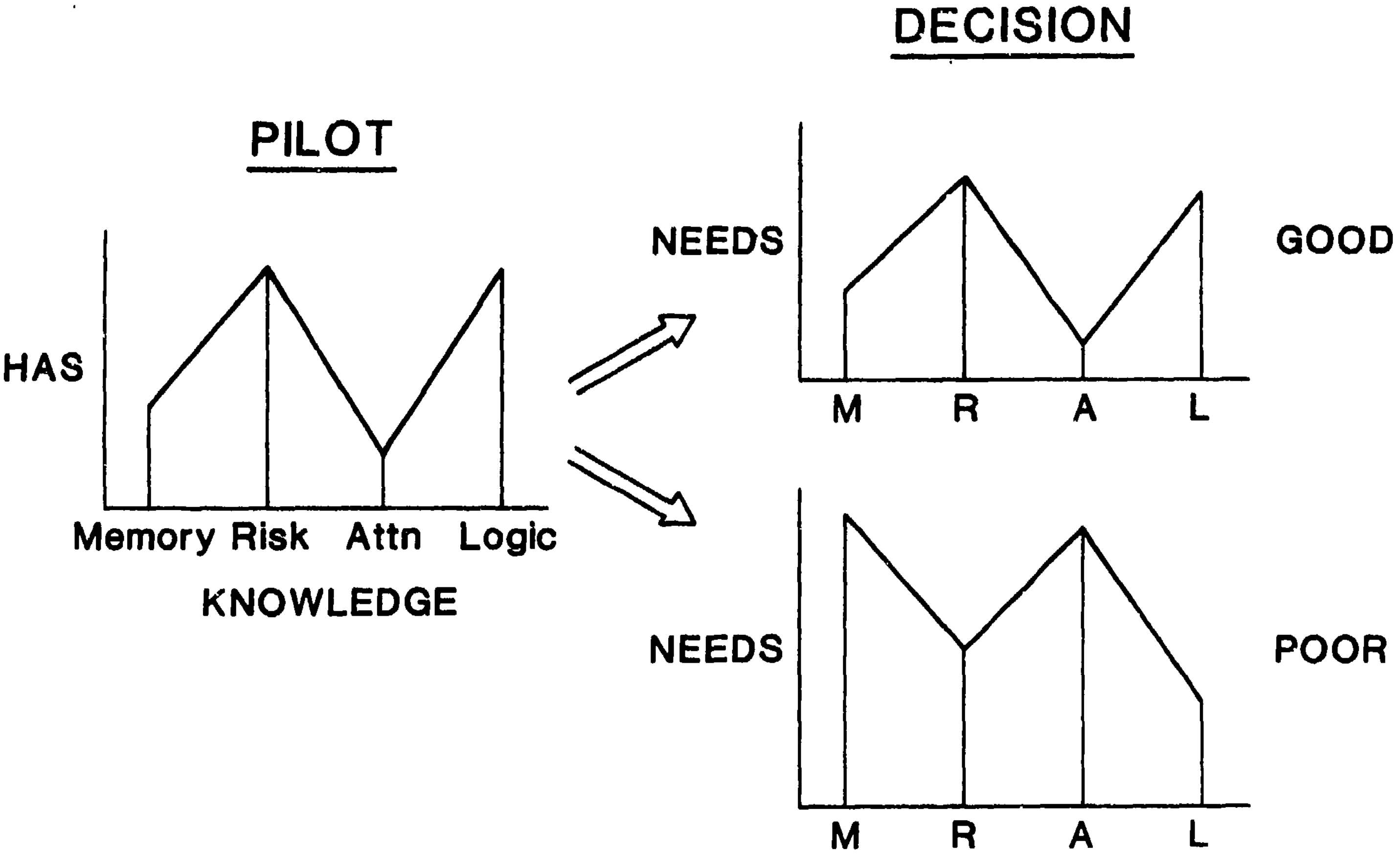


Figure 3. Profile of pilot's cognitive attributes, along with two representative scenario profiles. To the extent that the demands of the scenario match the pilot's profile, good performance is predicted. To the extent that a mismatch occurs, poorer performance is expected.

Our approach integrates three sources of data. (1) Decision performance data are collected using the MIDIS flight decision simulator incorporated on an IBM PC/AT, which will be described in section 2.2. (2) A test battery provides a psychological "profile" of cognitive abilities for each subject pilot described in section 2.3. (3) The decision situations used in MIDIS are content analyzed in terms of our hypotheses about the psychological demands made by each situation or "scenario." In the following pages, we shall first discuss the MIDIS system, then the cognitive test battery, and finally the analytical approach used to connect the two.

2.2 The MIDIS Decision Simulator

The project that we describe has followed two parallel but interacting tracks, as shown in Figure 4. On the left of the track, a team of flight instructors collaborating closely with cognitive psychologists have designed a series of flight decision problems or "scenarios" that incorporate the heterogeneous set of information processing demands that may be imposed upon

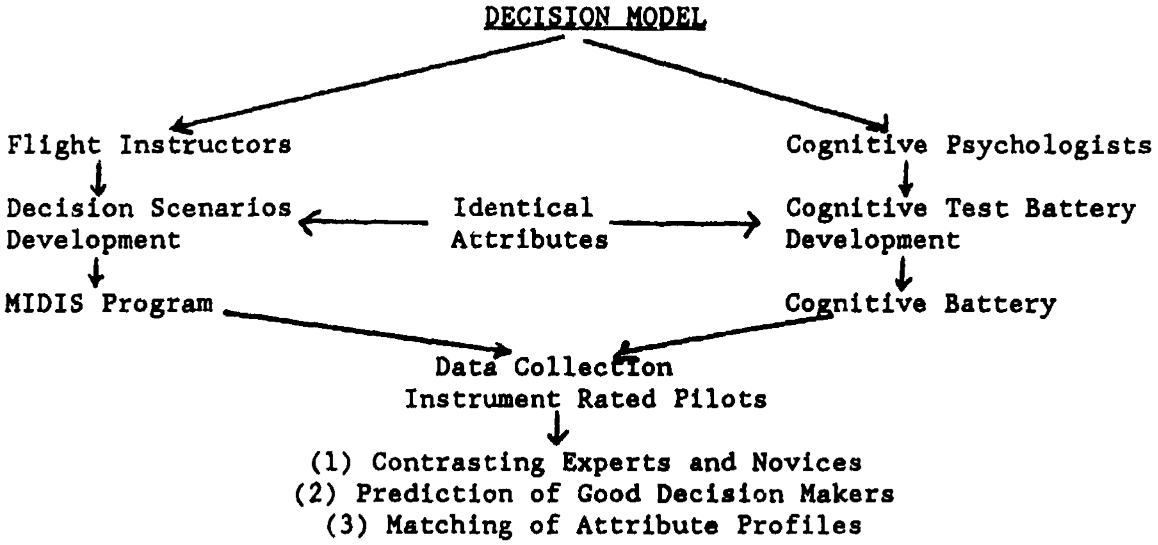

DECISION MODEL
Flight Instructors
Cognitive Psychologists
Decision Scenarios Development
Identical Attributes
Cognitive Test Battery Development
MIDIS Program
Cognitive Battery
Data Collection
Instrument Rated Pilots
(1) Contrasting Experts and Novices
(2) Prediction of Good Decision Makers
(3) Matching of Attribute Profiles

Figure 4. The MIDIS Project.

the pilot. Generation of these scenarios has depended both upon an understanding of the model in Figure 1, and years of expertise in instrument flying. Certain decision problems will require a breadth of attention, others will require that hypotheses be revised in light of new data, and still others may require an accurate assessment of risk. While incorporating these attributes, an effort has also been made to present the series of decision-situations as discrete events in a single coherently flowing flight from an origin to a destination. To enhance experimental validity, MIDIS has a number of simulator-like qualities (it provides a continuous "engine" sound cue, for example, and permits route deviations or reversals).

The MIDIS system itself consists of two programs, SETSCENE 2 and MIDIS 2, written in PASCAL and running on the IBM AT. The first program, SETSCENE, is an editor that facilitates the preparation of "flights" by the experimenter/flight-instructor. SETSCENE provides input to MIDIS, which controls a text and instrument panel display. The general structure of the MIDIS system places it in a class of programs referred to as "Graph Traversers" (Doran & Mitchie, 1966). Graph traversers are applicable to situations where a number of states are connected by a set of transformations or "operators." This can be represented as a branching tree-structure graph in which the nodes represent the states and the operators linking them are transitional probabilities. The states in MIDIS take the form of descriptions of realistic in-flight situations referred to as "scenarios." These are similar in concept to the SET (Situational Emergency Training) scenarios developed at Luke AFB for F15 pilot training, i.e., simulations of real situations requiring decision-making skills. Unlike SET, however, a MIDIS situation may involve any potential in-flight situation, emergency or otherwise. Each scenario requires that a decision be made among several alternatives presented. The decision influences the occurrence of subsequent scenarios since it selects the transitional probabilities that will operate.

Two considerations determined the scenario sequencing structure used in SETSCENE. First, there is the problem of devising a scenario structure that gives the appearance of being unbounded to the user while in fact having a constrained formal structure. The second consideration concerns the need for this structure to represent the pattern of deteriorating circumstances that often characterizes aircraft mishaps. These misfortunes do not usually occur as a result of one poor decision or one technical malfunction, but rather as a result of several concatenated events opening successive "gates" to an accident. Figure 5 represents a structure designed to keep the progress of the simulated flight "on track," while at each stage allowing digressions into successively less optimal scenarios. (For clarity the figure shows just three branches from any one scenario. In fact there are ten.) This structure is built around "core" scenarios that represent situations at points along a cross-country flight-track. Core scenarios are to some extent independent of each other, for although they must make chronological sense they do not form a tight causal chain. Other scenarios, generally less favorable to the success of the flight (here labeled "side" scenarios for convenience) become more probable as decisions become less optimal. The further down the chain of side scenarios the subject proceeds, the less probable is his return to a core scenario.

2.2.1 SETSCENE 2. SETSCENE permits access to up to ten scenarios from any starting point scenario and up to six decision options per scenario. Along with each scenario a comprehensive range of instrument panel readings is also stored plus rate of change information. These data are accessed by MIDIS 2 as a subject progresses through a "flight." A realistic time limit is incorporated with each set of decision options in SETSCENE. This is because

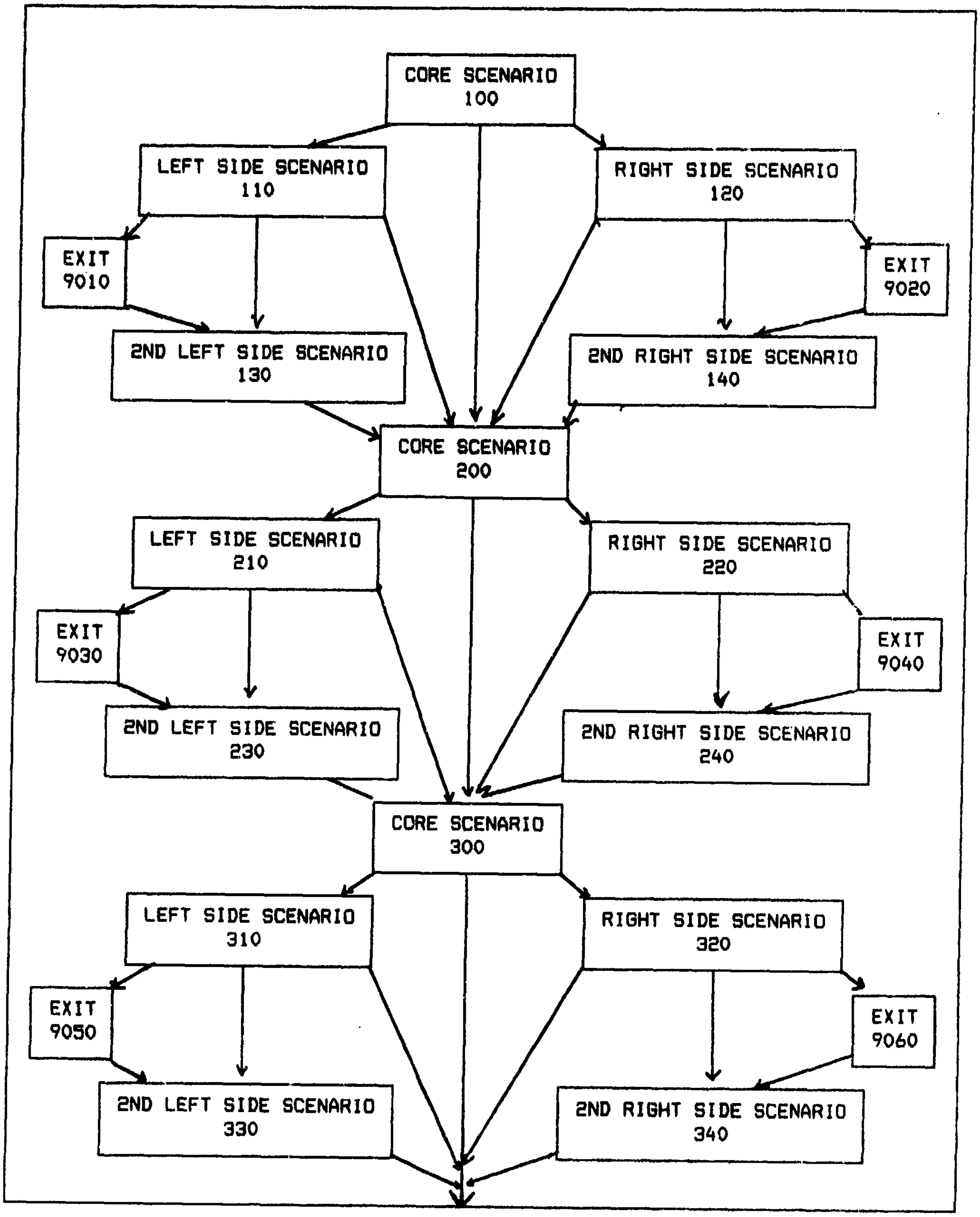


Figure 5. Branching structure illustrating the possible paths of a simulated MIDIS flight.

some problems are very circumscribed in the amount of time that can be allowed for the decision process (engine failure, for example), whereas others are more open-ended (e.g., radio failure). If no decision is made within the time allowed, SETSCENE ensures that MIDIS defaults to the situation most likely to occur should the pilot fail to intervene.

SETSCENE incorporates two further sets of algorithms: one set uses Boolean logic to permit any decision to have a delayed effect upon any subsequent scenario as desired. Another set automatically counts syllables in scenarios and decision options. This was developed to provide accurate counts on both normal text and radiocommunication language, permitting MIDIS to factor out reading speed variance in problem study and decision selection times.

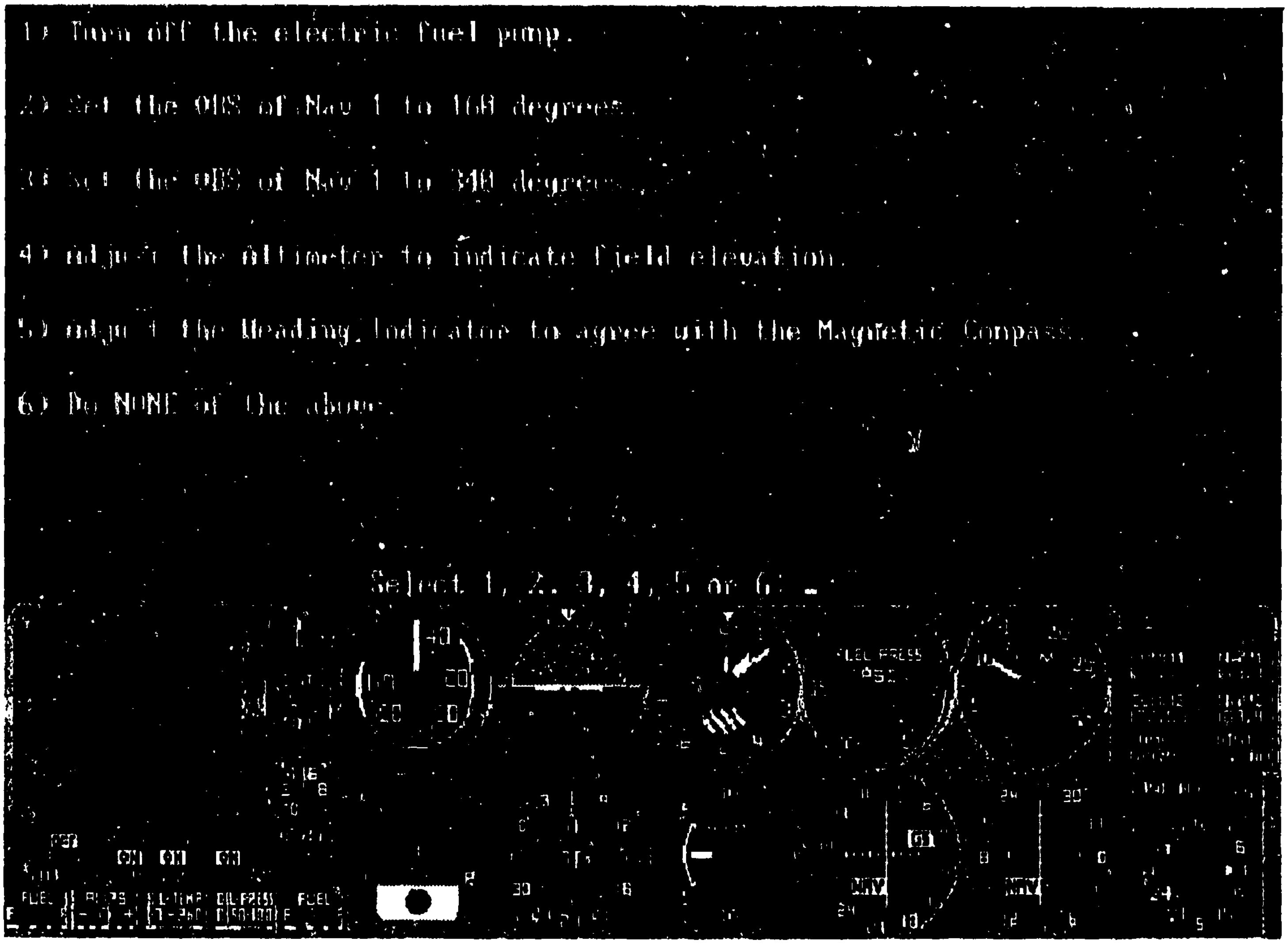
In addition to its "MIDIS driving" functions, SETSCENE also performs a number of important "housekeeping" operations. As discussed earlier, SETSCENE 2 has a structure capable of modeling event sequences with considerable realism and flexibility. How far this potential is realized, however, still depends heavily upon the quality of the "flight information" in the database. An "item bank" of scenarios has been prepared by flight instructors on the project team, and this database is continually being expanded. The program has been designed to assist in keeping track of the scenarios and options in the database. This facilitates the construction of different flights as well as the post-hoc analysis of those flights. Therefore each scenario may be identified according to a set of bibliographic descriptors (such as "cruise," "approach," "weather problem," "system malfunction," etc.), and cross-indexed searches can be carried out on these descriptors.

We also perform a content analysis of the situations themselves in terms of their psychological attributes. As discussed in more detail below, each situation is rated on each of 11 cognitive attributes. These ratings indicate our hypotheses concerning the extent to which reaching an optimum solution on the problem depends upon the strength of the attribute. SETSCENE stores these ratingr, and its search and retrieval capability permits scenarios to be identified or selected on the basis of similar problem structure. This is important to our componential study of decision-making detailed below.

2.2.2 MIDIS 2. MIDIS has a full, high-fidelity instrument panel based on a Beech Sport 180, the type of aircraft used for training at the University of Illinois Institute of Aviation. This display, implemented via the HALO graphics package and 16 color Enhanced Graphics Adaptor, represents a full IFR "blind flying" panel with operating attitude, navigational and engine instruments. MIDIS accesses SETSCENE files to change the readings on the instrument panel throughout the course of the "flight" in synchrony with the prevailing scenario. MIDIS does not attempt to simulate the flight dynamics of an aircraft from control inputs - the province of flight simulators - but it does provide for a flight-relevant concurrent psychomotor task, not used in the present experiment. Figure 6 gives a screen print of a MIDIS 2 display.

Seven performance variables are monitored, most of them unobtrusively. Four of these relate to response selection: decision choice, optimality, decision time (latency), and decision confidence. The last three of these are combined to form a decision quality hierarchy, with accurate, fast, confident decisions at the top and inaccurate, fast and confident decisions at the bottom. Slow responses made with low confidence have intermediate scores, with correct choices obviously receiving higher scores than incorrect ones.

Other variables monitored are problem detection, problem study time, and mean reading speed (text inspection time). A scenario can be defined by either the particular normal or abnormal configuration of the instrument panel alone, or by the instrument panel together with a text description of



1) Turn off the electric fuel pump.

2) Set the OBS of Nav 1 to 160 degrees.

3) Set the OBS of Nav 1 to 340 degrees.

4) Adjust the Altimeter to indicate field elevation.

5) Adjust the Heading Indicator to agree with the Magnetic Compass.

6) Do NONE of the above.

Select 1, 2, 3, 4, 5 or 6:

Figure 6. A representative MIDIS display panel.

particular circumstances. Where text accompanies the panel, the instruments are stable - showing no rate of change. The scenario represents a situation in which the aircraft is in steady flight.

The scenario may represent a problem or it may not. A problem scenario is one in which the circumstances have clear and present implications for the efficiency or safety of the flight, requiring diagnostic and corrective action to be taken.

When the panel appears alone, the subject's visual attention is not split between a reading task and a panel monitoring task. In these conditions the instruments can show a rate of change. This allows us to study an important class of decisions - those involving the detection of changes and the integration of decision cues in real time.

Finally, each subject's mean reading speed is unobtrusively calculated in syllables per second during the reading of the program run instructions. Since SETSCENE analyzes scenarios and options for word and syllable counts, as described above, individual differences in reading speed can then be factored out of the data.

2.2.3 Attribute and option coding. After creating each MIDIS scenario, the flight instructors on the design team proceeded to generate two kinds of codes, which were applied to and characterized the scenario in question. First, each option in a decision scenario was assigned an optimality rating, on a scale from 5-1, in which the correct (best) option was arbitrarily assigned a value of 5. The less optimal options were assigned values ranging from 1-4, depending upon how close they were to being plausible alternatives. Second, the correct option in each scenario was assigned an attribute value code for each of the 11 critical cognitive attributes listed in Table 1. These attributes were selected based upon our content analysis of the flight scenarios in MIDIS, guided by our expert analysis of pilot judgment. A value

Table 1. Scenario Demands of Cognitive Attributes.

1. Flexibility of Closure - the ability to find a given configuration in a distracting perceptual field.

2. Simultaneous Mental Integrative Processes - the ability to keep in mind simultaneously or to combine several premises or rules in order to produce a correct response.

3. Simultaneous Visual Integrative Processes - the ability to sample a select number of items from a complex visual display, and to combine this information in order to produce a correct response.

4. Sequential Memory Span - the ability to recall a number of distinct, sequential items from working memory.

5. Arithmetic Load - the ability to perform basic arithmetic operations with speed and accuracy.

6. Logical Reasoning - the ability to reason from premise to conclusion, or to evaluate the correctness of a conclusion.

7. Visualization of Position - the ability to perceive or maintain orientation with respect to objects in space, and to manipulate this image into other arrangements.

8. Risk Assessment and Risk Utilization - the ability to accurately assess the probability or riskiness of a situation, and to utilize this assessment in effectively carrying out decisions.

9. Confirmation Bias - the tendency to seek confirmatory, rather than the more appropriate disconfirmatory evidence, when testing a given hypothesis.

10. Impulsivity-Reflectivity - a measure of cognitive style differentiating those who tend to be fast and inaccurate (impulsive) or slow and accurate (reflective).

11. Declarative Knowledge - the ability to answer correctly a number of "textbook" questions covering a broad range of general aviation issues. This measure specifically excludes procedural or experience-based issues, focusing only on declarative facts and guidelines.

of zero indicated that the attribute was not relevant to the decision. Values from 1-3 indicated how critical it was for the subject to possess strength in the attribure in question, in order to choose the optimum option. In the case of the confirmation bias, this coding was reversed (i.e., high values indicated how critical it was to avoid the confirmation bias).

2.3 Cognitive Battery Development

As shown schematically in Figure 4, the goal of cognitive psychologists in the project was to develop a set of cognitive tests that would match, as closely as possible, attributes that were identified in the scenarios. Our efforts to identify existing cognitive tests that assessed these attributes, parallel an analogous effort performed by Irizarry and Knapp (1986) in their study of individual differences in Army Intelligence Analysts. Based in part upon their study, and upon our own review of the literature on individual differences and cognitive attributes in decision and judgment, the development of the test battery proceeded as follows.

Our initial goal was to locate any existing standardized tests that provide measures on each of the relevant attributes. In some instances, more than one standardized test exists for a single attribute. In that case, one was selected based upon the criteria of administration time, face validity and reliability. For those attributes for which we were unable to locate a standardized measure, specific tests were developed within our laboratory. Thus, the compiled test battery consists of a one-to-one mapping between cognitive attributes relevant to pilot judgment and cognitive tests specifically designed to measure each individual attribute. Table 2 provides a list of the specific tests comprising the cognitive test battery.

A number of cognitive measures were taken directly from the Educational Testing Service (ETS) kit of Factor-Referenced Cognitive Tests. The specific cognitive factors and tests selected from this kit included measures of flexibility of closure (hidden figures), simultaneous integrative processes (following directions), sequential memory span (visual number span),

Table 2. Cognitive Test Battery.

1. Hidden Figures Test
2. Following Directions Test
3. Cue Sampling - Visual Integration Test
4. Visual Number Span Test
5. Subtraction and Multiplication Test
6. Nonsense Syllogisms Test
7. Surface Development Test (Spatial Visualization)
   Card Rotations Test (Spatial Orientation)
8. Risk Assessment and Utilization
9. Wason's 2-4-6 Rule Discovery Task
10. MFF Test and Impulsivity Self-Report Inventory
11. Aviation Declarative Knowledge Test

arithmetic load (subtraction and multiplication), logical reasoning (nonsense syllogisms), spatial orientation (card rotations), and spatial visualization (surface development). For each of the cognitive factors listed, the ETS kit contained two or more specific tests. We selected one based upon the criteria described above. The remaining portion of this section describes the tests that were developed within our laboratory, or were modified in some way.

2.3.1 Rule discovery task (Item 9). The extent to which subjects adopted a confirmatory bias, or the more optimal disconfirmatory strategy, was measured using an adaptation of Wason's (1960) "2-4-6 rule discovery" task. Previous research by Irizarry and Knapp (1986) suggests that this task is a valid measure of individual differences in hypotheses testing strategies.

For each trial, subjects were presented with a set of 3 numbers (e.g., 2-4-6), and asked to generate an hypothesis about the set membership rule (e.g., numbers increasing by two). Subjects were then asked to generate another set of three numbers to test the accuracy of their hypothesis. This response was then scored as adhering to a confirmatory or disconfirmatory strategy.

Subjects adopting a confirmation strategy would test the hypothesis by generating a series of numbers consistent with the hypothesis, while those adopting a disconfirmation strategy should generate a set of numbers inconsistent with the hypothesis. Five trials in total were given, so that each subject's score was the proportion of the total trials that a disconfirming strategy was used.

2.3.2 Reflectivity-impulsivity (Item 10). This test measures cognitive style differences in information processing. Subjects are typically categorized as "impulsive" if their performance on a task is rapid and inaccurate, and categorized as "reflective" if performance is slow and accurate. The primary index of reflectivity-impulsivity is the Matching Familiar Figures (MFF) test (Kagan, 1966; Kagan, Rosman, Day, Albert, & Phillips, 1964). The test requires subjects to select one exact match to a prototype from a set of exemplars. For purposes of the test battery, the adolescent/adult version of the MFF was used.

While pilot data for this test displayed large variances in response times, little variance in accuracy was observed. Thus, to aid in discriminating reflective and impulsive subjects, four items from the impulsivity scale of the Eysenck personality inventory were added (Eysenck & Eysenck, 1963). Previous research by Dickman (1985) and by Dickman and Meyer (in press) suggests that these items predict reflective-impulsive performance on a speed-accuracy tradeoff function. These items are shown in Table 3.

2.3.3 Risk assessment and utilization (Item 8). These critical characteristics were measured by a test developed within our laboratory (see Appendix A). The test consisted of four parts: proportion estimation, cause of death estimation, probability estimation of aircraft accidents, and utilization of gambles. The first part, proportion estimation, required subjects to estimate percentages or proportions various figures (e.g.,

Table 3. Eysenck Personality Inventory - Impulsivity Scale Items.

1. Do you stop and think things over before doing anything?

2. Do you generally do and say things without stopping to think?

3. Do you like doing things in which you have to act quickly?

4. Are you slow and unhurried in the way you move?

estimate the percentage of the circle that is shaded, or estimate the degree of an angle).

In Part II, in keeping with the tradition of Slovic, Fischoff, and Lichtenstein and colleagues (Lichtenstein, Slovic, Fischoff, Layman, & Coombs, 1978; Slovic, Fischoff, & Lichtenstein, 1976), subjects were asked to estimate the number of people killed each year by such factors as electrocution, automobile accidents, cancer, or tornadoes. Similarly, Part III dealt with estimating frequencies of different types of aircraft accidents, thus providing an aviation context.

In Part IV, risk utilization was assessed by presenting the subject with a series of choices between a risky gamble and a sure bet. In each case, the expected utilities (EU) were equal (in some questions, the EU was negative, and others, positive). For each of the four parts, it was hypothesized that well-calibrated subjects would be fairly accurate in their estimations and utilization, while "risky" or "conservative" behavior should fall at the two extremes.

Separate scores were derived for each of the four parts, and thus could be treated separately as individual measures of different types of risk estimation and utilization, or combined to represent a single measure of "riskiness performance." Pilot data revealed that subjects' performance often was inconsistent across the different parts (for example, those who were conservative in risk estimation were often not conservative in risk utilization). Thus, for purposes of the final data analysis, each of the four parts was treated individually.

2.3.4 Simultaneous visual integration (Item 3). This visual cue sampling test was developed in our laboratory and is a computer-based visual integration task. Each trial consisted of twelve lines presented simultaneously on a CRT screen for a brief exposure duration. The lines were of 3 different lengths: short, medium, and long. Above each line was a random distribution of X's. An example of a typical trial screen is presented in Figure 7. The subject's task was to find the four long lines on the screen and total the X's on only those lines. The position of the lines did not vary from trial to trial, only the number of X's presented each time. Thus, this test determines how accurately an individual is able to integrate only the relevant information from known spatial locations in a "cluttered" display.

2.3.5 Declarative knowledge (Item 11). In order to account for variance attributable to declarative flight knowledge, an aviation general knowledge test was developed by our flight instructors. This test was composed of a number of items selected from the FAA instrument exam. These items were carefully selected to comprise a representative sample of meteorological, navigational and systems questions. This test consisted of 25 multiple choice items and is contained in Appendix B.

Analysis of data from pilot studies for those tests developed and refined within our laboratory resulted in a range of individual differences for each independent measure, as well as for specific items within a measure. The one exception was the "2-4-6" Rule Discovery task as a measure of individual differences in hypothesis testing. Despite several revisions, the test failed

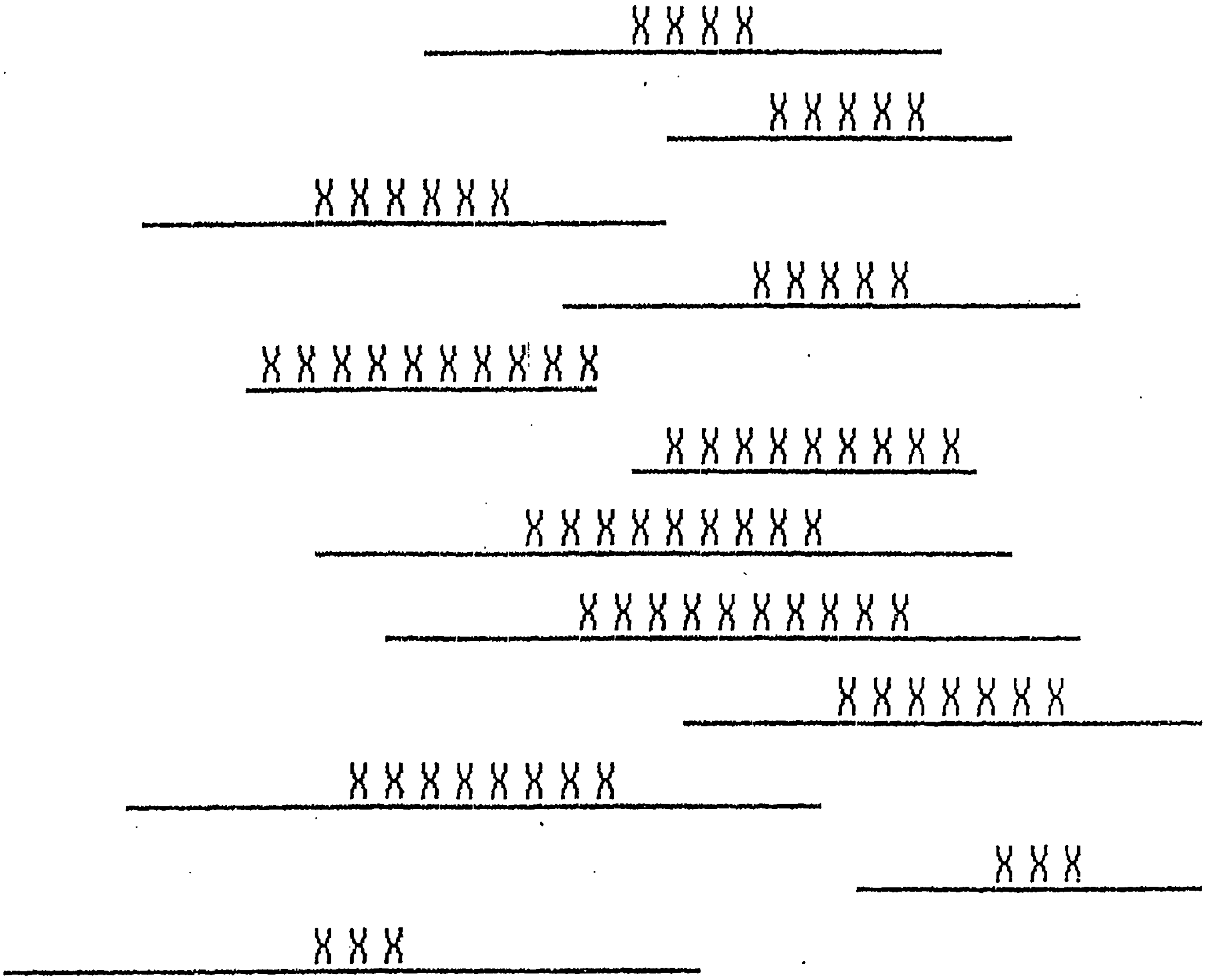


Figure 7. A representative display screen for the visual cue sampling test. For successive trials, the number of X's per line varied, while the position of lines on the screen did not.

to reveal any individual differences. Since it was not clear to us whether all subjects adopted a confirmation strategy or merely did not understand the task at hand, the test was dropped from the final battery.

3. METHOD

The subject pool consisted of flight instructors from the University of Illinois Institute of Aviation, experienced Instrument/Commercial pilots with diverse backgrounds (e.g., Air National Guard, professional airline and private business flying), and Instrument Rated student pilots from both the Institute and local flight schools.

The main experiment is based upon a sample of thirty eight subjects divided into two cohorts, twenty pilots from the experienced group and eighteen from the "novice" group.

Data collection was conducted in two sessions for each subject. In the first session, lasting approximately two hours, the battery of psychological tests was administered. The second session (in most cases taking place on a subsequent day) involved the MIDIS simulation itself. Subjects were instructed to plan an IFR flight from Mountain View, Missouri, to St. Louis Regional (Alton) in Illinois. Sectional charts, L-charts, Approach Plates, Airport Facility Directories and a Flight Service Station weather briefing were provided. Although no "stick-and-rudder" flying is involved in a MIDIS simulation pilots unfamiliar with the aircraft simulated by MIDIS, the Beech Sport 180, were given a briefing on the performance characteristics of the aircraft for flight planning purposes. Subjects were also provided an opportunity to review a screen-print of the instrument panel. No time limit was imposed for flight planning and different subjects took between 20 minutes and one hour to complete this phase. Pilots were instructed to plan the flight in their customary fashion with due regard to both the safety and efficiency of the trip. Subjects were informed that these two factors would be evaluated by the MIDIS simulator.

Following the flight planning, subjects performed the MIDIS simulation. The simulation was presented on an IBM AT computer system which was enclosed in a sound attenuating, dimly illuminated subject station. Subjects were instructed to treat the simulation like an actual aircraft flight. They were informed that the entire simulation sequence was under computer control and after it was started would automatically sequence. They were asked to remain in the subject booth until the simulated engine noise stopped indicating an end to the simulation session.

The screen presentation displayed first an overview and general description of the MIDIS system during which reading rate was measured five times. This was accomplished by timing the intervals between the subject's successive key press requests to bring up the next display. The general description was followed by a practice flight designed to train subjects in the use of the color-coded keyboard and MIDIS conventions. The practice flight was not time-limited and could be re-entered and repeated until the subject felt comfortable with the system. The practice sequence was "flown" for an average of 15-20 minutes. After the practice flight a sample "feedback" screen was provided to indicate the form of the safety and efficiency evaluation which would terminate the run. After a reminder weather briefing, the flight from Mountain View to St. Louis was started.

4. ANALYSIS

Each possible decision choice was rated for optimality on a five point scale. The decision quality (DQ) algorithm combines the optimality of the option chosen with the confidence rating and response latency in the following manner:

$$DQ = O+[O-ABS(O-C)] + \text{ or } - (zL*W)$$

where O is the optimality rating 1 to 5, C is confidence rating 1 to 5, zL is the z-score value of the latency of the decision, and W is a weighting which varies with optimality rating. The first expression in the algorithm involving optimality and confidence gives a point score from -2 up tc 10. Subtracting the absolute value (ABS) of the difference in optimality and confidence from the optimality rating gives credit for being "well-calibrated," i.e., for rating level 5 decisions at confidence 5, and level 4 decisions at confidence 4, etc. By the same token it penalizes overconfidence. The second expression introduces latency into the overall DQ score. Lz is derived by computing the standardized score of all pooled responses (across subjects AND scenarios). How latency affects DQ score, however, is conditional upon whether the optimality of the decision choice is above or below 3. Above 3, rapid response time increments the point score. Below 3, rapid response time is penalized. The value of Lz is doubled to give a range approximately equal to that of other components of the DQ score. The final value of DQ ranges over approximately 20 points. The DQ metric chosen is significant in that it acknowledges that options chosen in a decision are not categorically right or wrong, but may vary in their degree of correctness.

Data files from MIDIS runs were merged with the psychometric data files and z-scored using LOTUS, and subsequently analyzed using SPSS-PC on the IBM PC/AT.

5. RESULTS

5.1 Factor Analysis of Psychometric Data

The psychometric test battery data for all 38 subjects was factor analyzed to determine the pattern of abilities defined by the tests. Fifteen psychometric measures were considered in this analysis, including two measures of reflectivity/impulsivity (MFF and self-report items), three measures of visualization of position (hidden patterns test and surface development test), four individual sub-tests within the risk assessment and utilization battery, and seven other measures corresponding in a one-to-one fashion with the remaining seven cognitive attributes presented earlier (see Tables 1 and 2).

The factors were initially extracted using a principal-components analysis. The seven factors obtained in this analysis were then subjected to varimax rotation procedure. The data presented in Table 4 are the results of nine iterations of the varimax procedure. Given the small number of subjects relative to predictor variables however, caution is advised in interpreting each of these factors. Therefore, the following discussion will focus on only the first three factors (those with the highest factor-loadings). For sake of

Table 4. Factor Matrix for Psychometric Test Data.

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
Hidden Pattns.	.826	-.114	.056	.058	.202	-.059	.149
Card Rotations	.779	.185	.105	-.062	-.310	.076	-.079
Surface Devlp.	.706	.229	.186	-.063	.191	-.139	.336
Following Dirs.	-.003	.895	-.109	.082	.010	.014	.060
Nonsense Syllog.	.242	.688	.168	-.209	.199	.238	-.085
Risk IV (Gambles)	-.130	.171	-.789	-.065	-.054	-.061	-.011
Risk II (Deaths)	.123	.287	.665	-.105	-.085	-.351	-.010
Risk III (Avrisk)	-.087	-.177	.047	.858	.039	-.064	.066
Math	.010	.251	-.361	.533	.054	.448	.223
FAA Quiz	.252	.321	.467	.529	-.121	.285	-.014
Vis. # Span	.074	.216	-.057	.148	.830	-.084	-.140
Vis. Scanning	.006	.138	-.069	.422	-.602	-.331	-.075
MFF Test	-.044	.119	-.052	-.005	.028	.737	-.123
Risk I (Probest)	.011	.011	-.121	.086	-.003	-.212	.764
Self Report	-.116	-.033	.315	.294	-.236	.369	.573

clarity, Factors 1, 2, and 3 will be referred to hereafter as the spatial factor, the logical reasoning and integration factor, and the risk factor respectively.

Spatial Factor

This factor is comprised of three psychometric measures: the hidden patterns test (flexibility of closure), the card rotations, and surface development tests (both measures of visualization of position). Factor loadings for the three measures were 0.826, 0.779, and 0.706 respectively. Given that all three tests involve a degree of complex spatial reasoning, this strong interrelationship is to be expected.

Logical Reasoning Factor

Two psychometric measures, the nonsense syllogisms test (logical reasoning) and the following directions test (simultaneous mental integrative processes), have loadings with values of 0.895 and 0.688 respectively on this mental reasoning and integration factor. It was mentioned previously that the first factor combined measures involving complex _spatial_ reasoning. Similarly, the second is comprised of those tests requiring a degree of complex _logical_ reasoning.

Risk Factor

Of the four subtests comprising the risk assessment and utilization factor, two loaded highly on this third "risk" factor. The two tests, Risk II (estimation of causes of death) and Risk IV (utilization of gambles) had loadings of 0.789 and 0.665 respectively. Of particular importance is the inverse relationship between these two variables, which may be interpreted as defining a "dangerous world" syndrome. Individuals who perceive fatal risks to be high (estimating the probability of death as great) are conservative in their choice among gambles (most often preferring the "sure bet" over the risky option).

Those three groups represent the strongest factor relationships of the seven presented in Table 4. Therefore, for purposes of further analyses, three grouped factor scores (spatial, logical reasoning and risk) were computed for each subject. The factor score is comprised of the weighted total of those psychometric variables that load on a given factor. These factor scores, along with the individual psychometric scores that did not load strongly on a particular factor, reflect the profile of individual differences in cognitive abilities that were taken into consideration throughout the remainder of the data analysis.

Finally, a few other interesting results of this factor analysis should be noted. As may be observed in Table 4, a fourth factor groups the declarative knowledge FAA quiz with mathematical ability and Risk III (aviation-specific accident estimation). It is reasonable to assume that both the FAA quiz and the domain-specific risk test reflect measures of declarative aviation knowledge. The final measure of risk assessment, Risk I (probability estimation), appears to be quite different from any of the other risk subtests.

Important for its nonsignificance is the finding that our two measures of impulsivity/reflectivity did not load on a common factor. This finding implies that these two measures proposed in the literature as determinants of the impulsivity/reflectivity trait (MFF-Kagan, 1966; self-report items-Dickman & Meyer, in press) were not in fact tapping the same aspect of cognitive style in the population that was tested here. Such interpretations are extremely limited, however, given the low power of the factor analysis.

5.2 MIDIS Data Processing and Analysis

Data from MIDIS, the Test-Battery and the Attribute Ratings were integrated into three field data files. The first field contained subject identity, biographical information and MIDIS performance data. The second field held the eleven attribute ratings for the individual flight scenarios, and the third field contained the psychometric data from the fifteen measures in the test-battery along with the three factors scores described in the previous section. Three additional variables were also computed.

The first of these, Calibration, was a points score consisting of the optimality rating on the selected action alternative plus an increment for the extent to which an individual was "well calibrated," that is, the extent to which his confidence rating accurately reflected his decision performance in terms of optimality. One further calculation yielded the global "decision quality" variable DQ. This was computed from Calibration and standardized (z) scores of Latency, and thus incorporated all three primary dependent measures - decision optimality, confidence and decision latency.

Analysis of these data was carried out using SPSS/PC+ (Norusis, 1986), the microcomputer version of the Statistical Package for the Social Sciences (Nie, Hull, Jenkins, Steinbrenner, & Brent, 1978). Three procedures were used: Pearson Product Moment correlation, Forward Stepwise Multiple Regression, and Discriminant Analysis. All statistical procedures were implemented for both type 1, static scenarios, and for type 2, dynamic scenarios. As noted, all latency measures were corrected for reading rate, on the basis of a covert assessment of this variable made during the presentation of instructions. The three primary dependent variables in the regression procedure were Decision Optimality, Confidence Rating, and Decision Response Latency. Decision Quality, Problem Study Time and (in dynamic scenarios) Problem Detection Time were also explored as dependent measures.

In a preliminary analysis, visual inspection of the scatter plots between performance and abilities suggested that the relation between the predictor and criterion variables was inconsistent between subjects of low and high experience. A different pattern was manifest for the flight students than for the flight instructors for example, and performance of the third group of local pilots outside the Institute of Aviation, also varied as a function of experience. As a consequence, a decision was made to divide the total sample into two groups on the basis of flight experience and to apply the analyses in turn to each group. A cutoff was chosen at 400 hours because this was a figure that: (a) divided the group approximately into equal groups (a sample of 18 "novices" and 20 "experts"), (b) on a log plot of flight hours, formed the mean point on a roughly normal distribution, and (c) provided a grouping that included all students in one group and all flight instructors in the other.

The several aspects of the data, and the multidimensional characteristics of the experiment (i.e., with attributes, abilities, subjects, and problems) allow for a large number of different approaches to data analysis. Figure 8 provides a framework for describing these different approaches.

Shown at the top of the figure are hypothetical scatter plots relating the assessed level of six subjects on two cognitive attributes, to the decision performance vector. The term "performance vector" is used to acknowledge the existence of 5 or 6 different performance measures (e.g., optimality, latency). The six subjects are shown as belonging to the two cohorts labeled "novice" and "expert."

The two panels at the top of Figure 8 illustrate the differences in the predictive power of the two hypothetical attributes for this set of hypothetical data. Attribute 1 shown in the left panel provides a reasonably

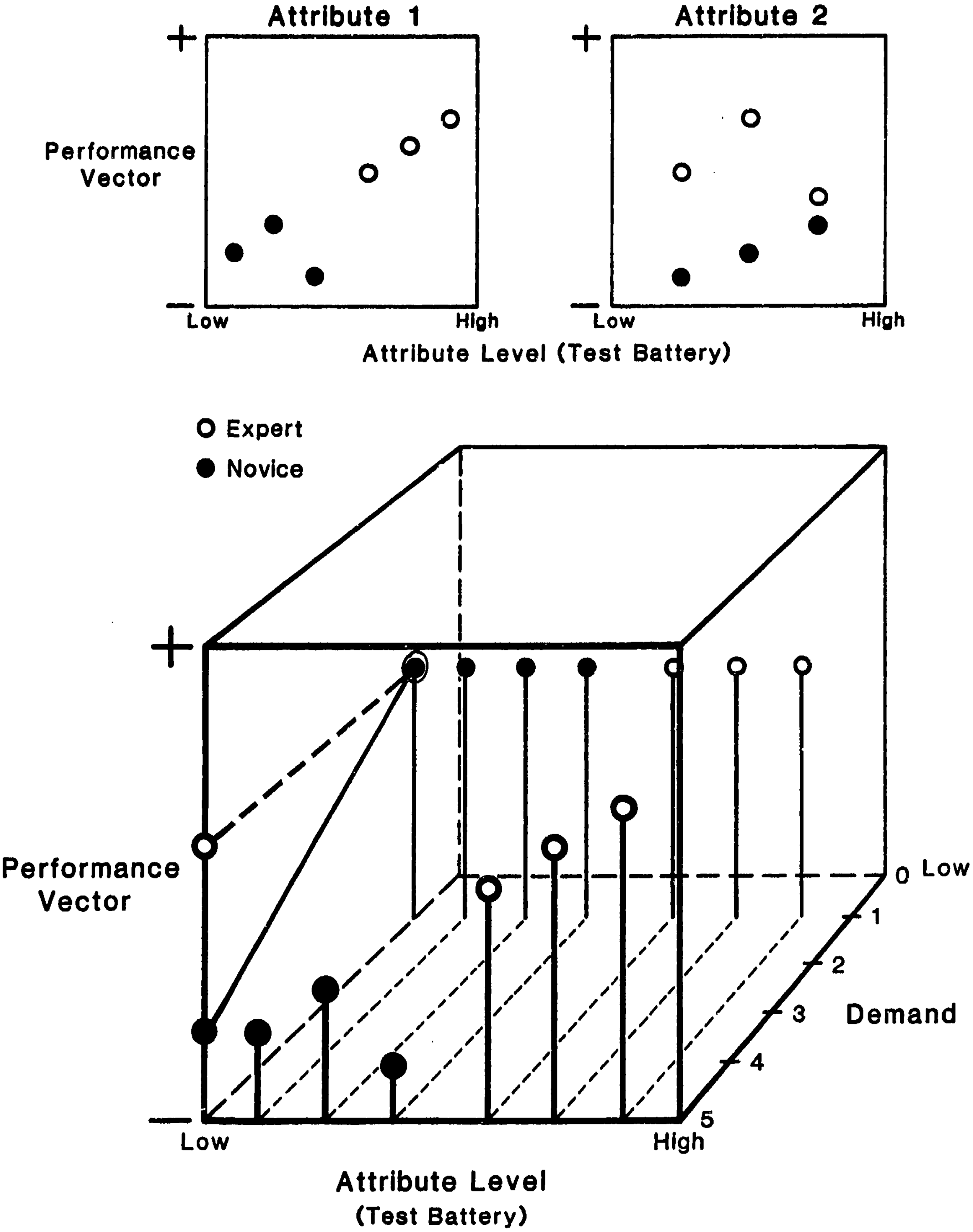


Figure 8. Dimensions underlying hypothetical data for six subjects. The lower panel represents the three dimensional expansion of the data in the upper left.

good prediction of decision quality for the group as a whole. Furthermore it nicely discriminates the performance of novices from those of experts. But when prediction is examined within each of the two cohorts, the attribute predicts expert performance but fails to do so for the novice. Attribute 2 on the other hand does not discriminate the two cohorts, nor provide an effective predictor for the total subject pool. However, in contrast to attribute 1, the second attribute does predict performance of the novice, but not of the expert. Hence in a multiple regression analysis of these data, very different beta weights would be applied to the two attributes when analyzing novice vs. expert performance.

In the top two panels of Figure 8, each performance vector may be considered as the mean value for a subject across all problems encountered. But the development of scenarios in MIDIS allowed different scenarios to be coded differently on a given attribute. This was done to test the hypothesis that only problems which demanded a given attribute would show performance that depends upon the relevant subject's ability. This incorporation of attribute levels is shown in the bottom panel, in which a third dimension is added to define the demand level on attribute 1, for each of two problems of differing demand, encountered by all six subjects. Thus, each subject's data now represents a "slice" in the cube along the perspective depth dimension of the figure. Three features may be highlighted with these hypothetical data. First, for both novice and expert subjects the attribute measured on the test battery is not a predictor of decision performance when the decision problem demands little of the attribute in question (near the back of the cube). This is in contrast to expert performance when the demand level is high (toward the front of the cube). Second, the attribute fails to discriminate expert from novice decision performance on problems when the demand level for that attribute is low, but does so when demand is high. Third, projected along the left side "wall" of the cube is the gradient of decision performance against demand level for the mean performance of the two groups. Here we see that increasing the demand level on the attribute has a strong effect on novice, but not expert performance.

The hypothetical data presented in Figure 8 thus reveal two intrinsically different styles of analysis: correlational analysis, and analysis of differences between groups. Each of these styles in turn may be carried out across all problems, or only on that subset of problems that are rated high on the different attributes, and each of these may also focus on the effects of ability differences or problem demand differences.

5.3 Between Groups Means Analysis

Analysis of the psychometric test battery scores for the two cohorts revealed that there were no significant differences between the two groups on any of the test measures, including the FAA-based test of declarative knowledge. Therefore, all attributes appear to show the pattern of attribute 2 on the top of Figure 8. Table 5 presents a comprehensive list of the differences in decision performance between the novice and expert groups for static and dynamic scenarios. The table highlights those differences that appeared significant with a two tailed t-test in the decision performance measures. Shown on the top row of the table are differences between groups on all problems, pooled across the differences in attribute scores. Shown in the nine rows below are differences observed specifically on problems that were rated high on the attribute listed. Only those performance measures that significantly differentiated the two groups are shown. Two tailed tests were employed because there were no a prior hypotheses about which group would perform "best."

Table 5. Performance Differences Between the Two Cohorts. The underlined value represents the superior group.

	ATTRIBUTE	STATIC				DYNAMIC			
		Performance Measure	Novice-Expert		P	Performance Measure	Novice-Expert		P
A.	ALL ATTRIBUTES	Confidence	4.05	4.40	(.025)	None			
B.	PROBLEMS RATED HIGH ON THE LISTED ATTRIBUTE								
1.	Field Dependence	Confidence	3.81	4.52	(.013)	Latency	47.1	50.7	(.016)
2.	Simul. Integ. Process	Confidence	3.85	4.25	(.029)	Latency	41.8	37.3	(.081)
3.	Sequential Memory	Confidence	3.00	4.60	(.029)	Confidence	44.0	38.3	(.05)
4.	Arithmetic								
5.	Logical Y Reasoning	Confidence	3.99	4.35	(.024)				
6.	Visual of Position	Confidence	3.96	4.38	(.008)				
7.	Vis Cue Sampling	Confidence	3.59	4.16	(.032)	Latency	44.6	47.0	(.02)
8.	Risk Need								
9.	Decl. Know.	Confidence	4.04	4.41	(.021)				

Relative to the total number of measures assessed, the table reveals few differences between groups. For example, the groups did not differ in decision optimality or decision quality for any subset of the problems. The most salient difference is the greater confidence shown by experts in the static decision problems. This variable discriminated the groups on the total pool of problems, as well as on those problems that were high on most of the specific attributes. Only on problems that possessed high demands for arithmetic and risk assessment was confidence equal for the two groups.

Subsequent analysis revealed that a major source of difference in confidence between the two groups may have been related to differences in the subset of problems selected by each. Given the branching nature of the MIDIS program, it was possible for any two people to take quite different "paths" through the flight. In a subsequent analysis a similar comparison of confidence was carried out only on that subset of scenarios to which a majority of subjects responded. This analysis removed data points from more subjects in the expert group. The analysis, now carried out on roughly 80% of the total data set, revealed no significant difference in confidence between the two groups (Novice: 4.00; Expert: 4.17). From this analysis, it was concluded that some experts had a tendency to choose options which led them to follow-on scenarios about which they were more confident than the rest of the sample.

The differences between groups on the dynamic problems were somewhat less consistent. Here experts responded more rapidly on problems that demanded simultaneous integrative processing, while novices made more rapid decisions on those problems with high demands for field independence and visual cue sampling (i.e., the two attributes that defined the perceptual aspects of the task). Novices were also slightly more confident on problems with high sequential memory demand. Once again, for dynamic as well as static scenarios, the two groups did not differ in terms of the optimality of their response.

The relation between confidence and optimality scores for both groups provides an index of "calibration": to what extent is greater confidence rated to choices that are more likely to be optimal. To assess calibration, the mean confidence rating for the two groups was computed separately for problems on which optimality score of 1, 2, and 3 was obtained (i.e., more "incorrect" problems), and for problems on which an optimality score of 4 and 5 was obtained. These data are shown in Figure 9. It is evident from the figure that both groups are somewhat calibrated, in that their confidence grows on choices which are "easier" (i.e., which they are more likely to answer correctly), and does so at the same rate for both groups (F1,35 = 21.77; $p < 0.01$). However, since both confidence and optimality variables were rated on the same 5 point scale, it is easy to see that the change in confidence is not nearly as steep as the change in optimality would dictate. The difference in mean optimality between low and high problems is around 1.5 units. The difference in confidence is only 0.25 units.

The data thus suggest that both experts and novices are reasonably well calibrated for choices which they make optimally (and novices slightly underconfident); but both groups fail to down weight their confidence appropriately as problem difficulty increases, a classic pattern observed in other decision making fields (i.e., Fischoff & MacGregor, 1981; Fischoff, Slovic, & Lichtenstein, 1977; Kahneman, Slovic, & Tversky, 1982).

5.4 Effect of Problem Demand

As we have described above, each problem was coded on the 9-attribute 4 point scale as to the demand for particular attributes. It was anticipated

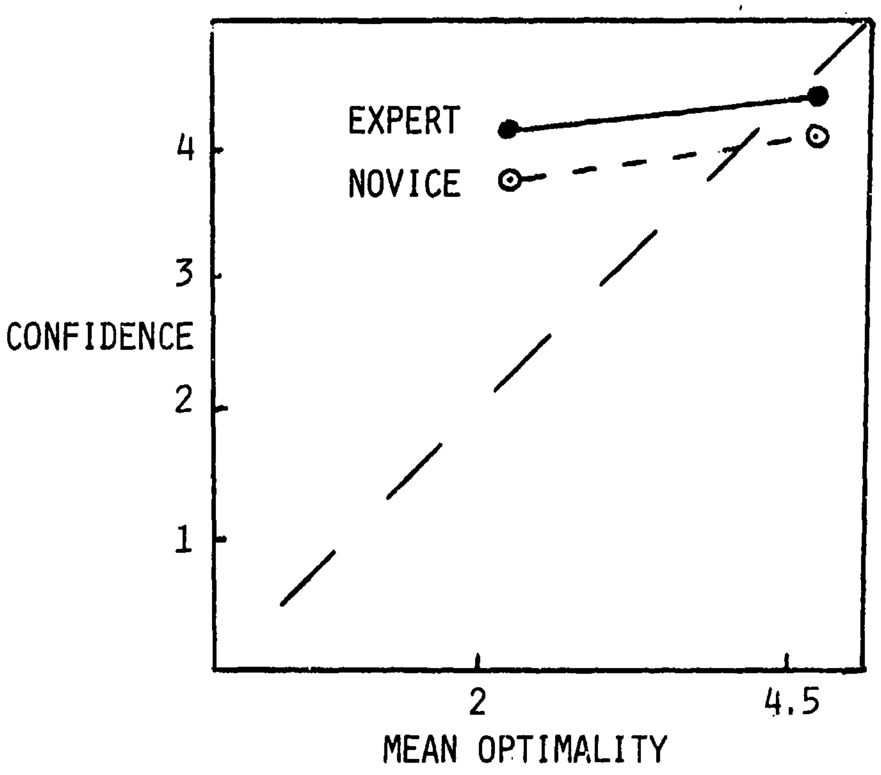


Figure 9. Confidence ratings assigned to "incorrect" (optimality 1, 2, and 3) versus "correct" (optimality 4 and 5) decisions. The figure presents novice and expert performance. The positive diagonal line represents the line of "calibrated" confidence assignment. The abscissa values of 2 and 4.5 represent the mean optimality values of "incorrect" and "correct" judgments respectively.

that to the extent that a problem was rated high on these scales, its level of decision performance would be reduced. To establish if this was the case, a total difficulty score for each problem was computed by summing the ratings across all attributes. This variable was then correlated with the various performance measures, and revealed that, for the "novice" pilot group, difficulty influenced problem study time for the static scenarios ($r = 0.46$; $p < 0.03$) and response latency for the dynamic scenarios ($r = 0.59$; $p < 0.01$). However, none of the performance measures of the expert appeared to be affected by the aggregate problem demand.

Subsequently the data were broken down attribute-by-attribute, to examine the sensitivity of performance measures to the demand of each attribute. This demand is indicated by the projected slopes on the left "wall" of the cube at the bottom of Figure 8. The significant ($p \leq 0.05$) correlations between the attribute demand variables and performance for the two cohorts, for both static and dynamic scenarios, are shown in Table 6. The correlations that are underlined are those that run in the unexpected direction of poorer performance (lower optimality or high latency) associated with lower attribute levels. The table suggests some substantial differences in the variables that make problems difficult for the two groups. For example, the expert suffers when static problems require field dependence (the correlation with optimality is negative) while the novice does not, and the expert's decision performance is slowed when static problems demand more simultaneous integrative processes while the novice's performance is not. For dynamic scenarios, the novice becomes less optimal when there is a greater demand for simultaneous integrative processing, while the expert actually becomes more optimal. The table also reveals a relatively large number of correlations that run in unexpected directions, indicating better performance with greater demand.

Table 6. Correlations of Problem Demand with Performance Measures. All correlations reported are $p < 0.05$. (*) indicates $p < 0.01$.

	DQ	OPT	CONFIDENCE	PST	LATENCY
STATIC					
Novice					
VISPOS			.42		
VISCUE					
RISKNEED				.41	
Expert					
FIELD	-.42	-.40			
SIMINT				.53*	.45
ARITH			-.48*		
VISCUE				-.42*	-.45*
DYNAMIC					
Novice					
SIMINT	-.53*	-.53*			
FIELD	.42	.62*			
LOGIC					-.64*
VISPOS					.56*
DECLKNOW			.48		
Expert					
SIMINT	.43	.45			-.39
VISPOS	.42	.48			
FIELD			-.41		
ARITH					-.52*
LOGIC					.51*

This by itself is not altogether too surprising since increasing levels on one particular attribute may have been correlated (across problems) with decreasing levels of other attributes. It is also possible that problems which provide more of a particular kind of attribute may in fact capitalize upon the strength of a cohort. Thus, for example experts seem to benefit more to the extent that visualization of position is required. This hypothesis will be addressed in the discussion.

5.5 Prediction of Overall Decision Quality

The previous sections have focused on the differences between the two cohorts, in their overall performance, and on how their performance is affected by problem demand. In this section we now address how their performance is affected by differences in cognitive ability within each group.

Based upon the output of the factor analysis, the first three factors discussed in section 5.1 (spatial abilities, reasoning and risk assessment) were extracted as relatively stable estimates of those cognitive abilities. Factor scores on these three were calculated for each subject, and these values, along with the FAA test score, and scores on the tests that did not load on the three primary factors were included in a stepwise multiple regression analysis to predict decision performance. The predictor variables then included the Spatial Ability, Reasoning, and Risk Factors, Visual Number Span, Mathematical Ability, Visual Scanning, Probability Estimation and two indices of impulsivity. These were included along with three domain-specific predictor variables: Total Flight Hours and two measures of declarative knowledge, the FAA Instrument Flight Written Test, and Risk III Test - knowledge of aviation accident risks.

Table 7 presents the results of this regression analysis for static (left column) and dynamic (right column) scenarios. Listed down the left margin are the criterion variables. Beneath each criterion variable is listed the order of predictor variables selected, along with the total variance accounted for. This variance (and the associated significance levels) has been corrected downward to guard against the potential capitalization on chance associated with the multiple regression analysis (Tatsuoka, 1976).

The left side of the table which presents the prediction of performance in the static scenarios, reveals that few if any variables are effective predictors. Confidence ratings were predicted by scores on the risk tests and by the total flight hours (high domain specific risk estimates and more hours producing greater confidence), while latency and problem-study time were both

Table 7. Multiple Regression Analysis (N = 38).

Type I (Static)			Type II (Dynamic)		
DQ:	RISKFACT	.07 (NS)	DQ:	VISNMSPN	.088 (.05)
	-			FAA	.144 (.05)
	-			TOTHRS	.224 (.01)
OPT:	FAA	.05 (NS)	OPT:	VISNMSPN	.152 (.01)
	-			FAA	.212 (.01)
	-			TOTHRS	.282 (.01)
CONF:	TOTHOURS	.12 (.05)	CONF:	RISK	.147 (.01)
	PROBEST	.22 (.01)		TOTHRS	.229 (.01)
	AVRISK	.28 (.01)		AVRISK	.287 (.01)
PST:	MFFTEST	.25 (.001)	PDT1:	SPATIAL	.113 (.05)
	REASON	.39 (.001)		AVRISK	.173 (.01)
LATENCY:	MFFTEST	.12 (.05)			
	REASON	.23 (.01)			

predicted in the expected direction by the MFF test of impulsivity-reflectivity. More "impulsive" responders tended to respond faster, and study the problem for a shorter time on the MIDIS task.

Somewhat more variance was accounted for by predictions in the dynamic scenarios shown on the right. Predictors of decision quality and optimality included the working memory test of visual number span and scores on the FAA questionnaire (both in the expected direction of better test scores leading to higher quality decisions). Measures of confidence in the dynamic scenarios, as was true in the static scenarios, were predicted by scores on the risk tests, along with the total flight hours.

Even though the predictions from the analysis depicted in Table 7 are statistically significant, they are in a sense disappointingly small. The proportion of variance accounted for in this analysis ranged from the high teens to the low 30s percent. Hence, as in previous analyses, the multiple regression analysis was repeated on each of the two cohorts separately, to determine whether the pattern of predictors was different. However, because the two cohorts did not entirely proceed through the same set of scenarios, this analysis was carried out on the "restricted set" of scenarios discussed in section 5.3. This refers to the path through the flight that was common to most subjects. Tables 8 and 9 present equivalent regression data for the static and dynamic scenarios respectively, with the novice groups data shown on the left and the expert groups data shown on the right of each table.

Considering first the static scenarios in Table 8, the data reveal an interesting and important contrast between the two groups. For the novice group, the optimality of performance is predicted reasonably well by two measures of declarative knowledge: the FAA test and the "Avrisk" measure of aviation risks. Here we see that those novices who tend to be more conservative (or estimate higher dangers) scored more optimally. In contrast, the optimality of the experts' decision performance is simply not explained. This drop in the predictive power of declarative knowledge from novices to experts has important implications that are discussed below.

Confidence ratings, in the static scenarios, like optimality, are also predicted differently for novices and experts. For novices, greater mathematical ability leads to lower confidence ratings. Given the general trend for overconfidence seen in Figure 9, this finding would suggest that better mathematical ability leads to better "calibration." As with optimality, the battery measures are not at all predictive of expert confidence ratings. Finally, the speed of problem study (PST) and response (Latency) are also predicted differently for the two groups. Spatial ability helps novices to perform more rapidly (a negative correlation with latency), while for the experts a large portion of the variance is accounted for by the

Table 8. Static Scenarios Multiple Regression Analysis. R^2 Values Adjusted for Capitalization on Chance.

	Novice N - 18			Expert N - 20	
OPTIMALITY:	AVIAT RISK	.181 (-)[1]	OPTIMALITY:	---	
	FAA QUIZ	.42			
CONFRAT:	MATH	.21 (-)	CONFRAT:	---	
PST:	SPATIAL	.304 (-)	PST:	MFFT	.427
				VISNUMSPN	.571
LATENCY:	SPATIAL	.307 (-)	LATENCY:	VISNUMSPN	.205

[1] A minus sign indicates a negative correlation between the predictor and criterion variable.

Table 9. Dynamic Scenarios Multiple Regression Analysis. R^2 Values Adjusted for Capitalization on Chance.

	Novice N - 18			Expert N - 20	
OPTIMALITY:	VISNUMSPN	.316	OPTIMALITY:	---	
CONFRAT:	RISK1	.291 (-)	CONFRAT:	---	
	MATH	.512 (-)			
PDT:	SPATIAL	.175 (-)	PDT:	FAA QUIZ	.528 (-)
				RISK FACT	.591 (-)
				MATH	.618 (-)

impulsivity-reflectivity measure of cognitive style (again in the expected direction with those having more impulsive styles responding more rapidly). An interesting finding here is that those experts who have greater working memory capacity, as measured by the visual number span test, take longer to respond.

A similar pattern of greater predictive power for novices than experts reappears in the dynamic scenarios shown in Table 9. For the novices, optimality is predicted by the capacity of their visual working memory.

Confidence ratings are predicted, as with the static scenarios, by mathematical abilities. In addition, confidence is tied to the risk test of probability estimation. Those who estimate probabilities higher tend to be less confident. Finally, for the novices the speed of problem solving was again related to spatial abilities. For the experts, as with the static scenarios, neither optimality nor confidence were well predicted. The speed of expert problem detection was reasonably well predicted by three variables. Faster detections were made by: (a) those who scored lower on the FAA quiz, (b) those who saw the world as "safer" (low scores on the risk faster), and (c) those with higher mathematical ability.

In summary, the pattern of data reveals a few general trends. Prediction is different between the two groups than across groups, which suggests that the pattern of skills predicting pilot judgment may evolve with experience. The usefulness of declarative knowledge declines as experience increases, and is replaced by ability factors that are not apparently assessed in the current battery. Also, confidence is better predicted for novices than for experts, and different variables influence decision speed for the two groups. Spatial abilities facilitated rapid responses for novices, but not for experts. Finally, where variables do predict, they do so in an orderly fashion and generally in the expected direction. For example, subjects with higher estimates of risk tend to perform more optimally. Impulsive experts tended to respond (on MIDIS) more rapidly than did reflective ones (although they were not necessarily more likely to be accurate). Greater mathematical ability produced better calibrated confidence for novices, and novice subjects who had greater working memory capacity were more optimal at diagnosing the dynamic instrument-based problems.

5.6 Regression on High Attribute Problems

Paralleling the procedures employed to discriminate groups in section 5.3 and depicted in Table 5, the correlational analysis between abilities and performance discussed in the previous section was repeated on a restricted subset of problems that were coded high on the attributes. That is, the regression analysis focused on the "front face" of the cube shown at the bottom of Figure 8. In general, the results of this analysis were disappointing. The cognitive abilities did not predict performance substantially better for high attribute scenarios than for the scenarios as a whole. Hence, the details of these results will not be reported.

5.7 Discriminate Analysis

In the previous discussion, all analyses focused on differences between novice and expert groups, defined on the basis of experience. Yet from the outset it was clear that the two groups did not differ substantially in their levels of decision quality. A different approach was taken to try to determine what cognitive attributes discriminated "good" from "poor" decision makers, without reference to the cohort to which they belong. To accomplish this, discriminate analysis was performed on the decision optimality scores obtained from the top and bottom quartile performers. For static scenarios, the two groups were discriminated quite well (Wilks' Lambda = 0.0052; $p < 0.001$) on the basis of all of the variables collectively. However, no particular subset of abilities stood out above others as being the most important discriminator. For dynamic scenarios, the overall discriminant function was less successful in differentiating the two groups on the basis of the cognitive attributes (Wilks' Lambda = 0.078; $p = 0.11$). However, the discriminate function revealed that three variables had significantly higher weightings than others. These were total flight hours, visual number span, and performance on the FAA quiz. Furthermore, the Wilks' Lambda statistic for each of these three variables in isolation produced significance levels of 0.07, 0.04, and 0.05 respectively.

6. DISCUSSION

The results of the present study are complex, and in some places slightly contradictory. Nevertheless, there are a number of general trends that show how problem difficulty measures and individual abilities affect decision performance of low and high experienced pilots on both static and dynamic scenarios. In certain respects the two groups responded alike. In the first place, both groups tended to lose confidence as problems became more demanding (as demand was defined by those problems that were more likely to be responded to incorrectly), although the experts possessed more confidence in general than did the novices. Secondly, neither group "down weighted" their confidence as much as they should, and hence both groups became increasingly overconfident as the problem difficulty increased. This failure to accurately calibrate confidence has been often reported in the literature (e.g., Fischoff & MacGregor, 1981; Fischoff, Slovic, & Lichtenstein, 1977). Finally, both groups failed to show appreciable effects of problem difficulty as this variable was explicitly manipulated in the experiment. That is, problems that had been coded high on scaled attributes were not generally responded to with less accuracy, although rated difficulty did have some effect on the latency with which novices made decisions.

One particularly striking aspect of the results was the absence of any difference in the overall quality of decision performance between the "novice" and "expert" groups. Only in terms of the confidence of their decisions did experts show a "higher" level of performance and this difference was the result in part of differences in the subset of problems that the two groups faced. The absence of group differences might possibly have been attributable to an added factor, favoring the novice group. That is, it is possible that the novices might have achieved some intangible benefits in this cross-sectional study because they were a younger group and therefore, presumably more familiar with the computerized characteristics of the MIDIS program.

Differences between the groups were not extremely large but the correlational analyses did reveal that more extensive qualitative differences exist in terms of how each group was affected by problem difficulty (as shown in Table 6). Likewise each group was affected by differences in ability within a group, as discussed in section 5.5. These latter sources of differences are of particular importance because they demonstrate the relevance of the domain-independent battery items to the domain specific measures of aviation judgment. Hence the main conclusions of this analysis are worth reiterating. Declarative knowledge as assessed by the FAA quiz and the aviation-specific risk test predicted novice but not expert performance, and confidence ratings were predicted by mathematical and probability estimation skills for novices but not for experts. Spatial abilities were good predictors of response speed for novices, but not for experts. Finally, the optimality of decisions for dynamic problems was predicted by the capacity of visual working memory for novices, but not for experts. In fact, response speed was the only aspect of expert performance that was well predicted, and this variable was predicted by a relatively large number of variables.

A second issue concerns how the two aspects of differences, between problems and between individuals, relate to each other. The answer, in general, is that they do not, although there is no logical reason why they should. That is, to say that novices as a group are influenced by problems that demand a lot of visual cue sampling does not necessarily suggest that cue sampling ability will predict novice decision performance. Expressed in other terms, the covariance between the two axes at the base of the cube on Figure 4 and the vertical performance axis may be low.

Indeed the only region in which there is some consistency between the dimensions of problem difficulty and individual ability is in the relevance of working memory, as assessed by the abilities test of visual number span, and as coded by the attribute of simultaneous integrative processing. Here novice performance is influenced by this attribute demand, and is predicted by differences among the cohort in this ability.

Thus, overall the data cannot be interpreted to reveal a fully conclusive picture of how pilots differ in their decision making capabilities. It is possible to offer four hypotheses as to why substantial variance in MIDIS performance was not accounted for by either group membership or ability differences.

In the first place, it is possible that the test was sufficiently unrealistic that it did not elicit credible decision behavior. Some evidence that this may not represent a real concern however is provided by the assessments of the subject pilots, many of whom commented on the realism of both the instrument panel and the flight scenarios. Nevertheless it is important to realize that MIDIS does depart from real flight judgments in four important respects: (1) The obvious risk factor of being airborne is missing from MIDIS; (2) MIDIS contains no closed loop perceptual-motor flight control; (3) Much of the information regarding the flight that a pilot would normally discern from environmental cues and views outside the cockpit is here presented in a less compatible textual format; (4) The structured characteristics of MIDIS require that a multiple choice format be offered, in which the most optimal response is presented for recognition. Clearly in actual pilot judgment, the pilot must often <u>recall</u> the correct diagnosis or action.

Secondly, it is certainly likely that two of the critical measures that form the basis for the model testing, optimality rating and problem coding, were perturbed by measurement error that led to a reduction in the power of our statistical tests. This is a result of the fact that both of these variables were subjectively generated by expert opinions. To some extent these noise sources were reduced in the ratings of optimality of each alternative, because this rating was carried out independently by the two flight instructors, and substantial agreement between them was observed (an interrater reliability of +0.75). Measurement error however was probably a greater contributor to the coding of attributes, since this was only performed by one flight instructor. This greater level of error variance was quite possibly responsible for the lack of success in obtaining attribute-specific predictions of the ability tests (see section 5.6).

Thirdly, there was a moderate lack of structure in the MIDIS task. Throughout the entire experimental program we have tried to strike a balance between imposing sufficient experimental control to make the data remain fairly structured, and sufficient freedom to make the MIDIS task unconstrained and realistic from the pilot's standpoint. The former criterion ideally dictates a linear decision path whereby all subjects receive the same scenarios in the same order, independent of what their choices may have been. The latter dictates a highly response-dependent, closed-loop branching structure in which each subject may go through a totally unique sequence of scenarios depending on the particular choices that he or she has made. In hindsight it may be that the current program has been biased too strongly in the latter direction, with a consequence being that similarly coded attribute levels may have been derived (for different subjects) from performance on very different scenarios. This lack of structure meant, for example, that only three dynamic scenarios were encountered by all thirty eight subjects, and led to a situation whereby the novice group encountered more problems of greater difficulty on certain key attributes, specifically visualization of position and simultaneous visual cue sampling. The reduced data analysis which focused on common scenarios addressed this issue to some extent, but with the consequence of eliminating several data points (particularly of the novice group) from the analysis.

A final concern in the present study is the possibility that our results might have been overinterpreted, with an increased likelihood of type I errors. To guard against this possibility, two precautionary steps were taken. The multiple regression analyses, explicitly contained corrections for capitalization on chance, and two tailed rather than one tailed t-tests were used in the between-groups comparison. However, the other tests were not adjusted for the increase in type I error resulting from the multiple comparisons (across dependent variables, attributes, or scenario types). This riskiness was intentional as we viewed the current data as more exploratory than confirmatory, and the experiment was intended to identify hypotheses that should be pursued in future research. As a result we did not want to commit type II errors and ignore effects that might have been present, even at low levels of reliability.

The four problem areas listed above are not trivial, and all are being addressed in ongoing and future research with the MIDIS program. Nevertheless, in spite of the problems, the major trends of the present data encourage continuation of this line of approach, with suitable modifications to both the task and the tests. These trends included the emergence of a reasonable level of MIDIS variance accounted for by the tests of fragile information processing components, such as working memory (visual number span), and spatial abilities, and by tests of more crystallized knowledge such as risk estimation and declarative knowledge. Furthermore, the variance that is accounted for is consistent with other decision making analyses that identify both risk utilization, and the more fragile, resource limited information processing components as important components in decision performance, thus, tying the results back to the initial model presented in Figure 1 (e.g., Einhorn & Hogarth, 1981; Slovic, Fischoff, & Lichtenstein, 1977).

Finally, two intriguing characteristics of the expert cohort suggests an important direction in which the approach should be extended. These refer to (1) the lack of expert variance accounted for in the multiple regression analyses by the declarative knowledge measures of the FAA quiz and the aviation-specific risk factors (see Table 8), and (2) the positive influence on expert judgment performance of the demands imposed by simultaneous integrative processing and visualization of position (see Table 6). The first of these phenomena suggests that important components of the knowledge base underlying expert pilot decision processes have not been captured. It is reasonable to hypothesize that these components may relate to procedural rather than declarative knowledge and be manifest in concepts of scripts (Schank & Abelson, 1977) and mental models (Gentner & Stevens, 1983; Rouse & Morris, 1986) that help to sustain the expert's situation awareness. Decisions or diagnoses may be made by matching mental models, scripts or previous experiences to the environmental circumstances (Stone et al., 1985), rather than by integrating facts of declarative knowledge with sampled environmental cues through computational mechanisms in working memory. Indeed the richer the source of environmental cues, if these are correlated with a pilot's experience, the more information will be available to make such a pattern match unambiguous and the better a pilot's judgment should be. A procedure like this would explain the initially surprising positive relation between the demand of a decision problem for simultaneous integrative processes and positional awareness, and the quality of the expert pilot's decision as shown in Table 6. This hypothesis is intriguing, and its test will depend upon a different methodology for examining pilot expertise (e.g., Schvaneveldt et al., 1985). Such methodology will be pursued in our future work as we try to unravel the mysteries of pilot judgment.

In conclusion, it should be noted that the current study is unique in its efforts to apply theoretical modeling, cognitive theory, and the methods of both experimental and differential psychology to the collection of expert decision making data. While a large amount of work still needs to be done in order to improve and perfect the approach, we feel confident that the current data are leading toward the acquisition of important information in this critical area of pilot judgment.

7. REFERENCES

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. Bower (Ed.), Recent advances in learning and motivation (Vol. 8). New York: Academic Press.

Buch, G., & de Bagheera, I. J. (1985). Judgment training effectiveness and permanency. In R. S. Jensen & J. Adrion (Eds.), Proceedings of the Third Symposium on Aviation Psychology. Columbus, OH: The Ohio State University.

Buch, G., & Diehl, A. (1984). An investigation of the effectiveness of pilot judgment training. Human Factors, 26(5), 557-564.

Dickman, S. (1985). Impulsivity and perception: Individual differences in the processing of the local and global dimensions of stimuli. Journal of Personality and Social Psychology, 48, 133-149.

Dickman, S. J., & Meyer, D. E. (in press). Impulsivity and speed-accuracy tradeoffs in information processing. Journal of Personality and Social Psychology.

Doran, J. E., & Mitchie, D. (1966). Experiments with the Graph Traverser program. Proc. R. Soc. (A) 294, 235-259.

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decisions theory: Processes of judgment and choice. Annual Review of Psychology, 32, 53-88.

Eysenck, S. B., & Eysenck, H. J. (1963). On the dual nature of extraversion. British Journal of Social and Clinical Psychology, 2, 46-55.

Fischoff, B. (1977). Perceived informativeness of facts. Journal of Experimental Psychology: Human Perception and Performance, 3, 349-358.

Fischoff, B., & MacGregor, D. (1981). Subjective confidence in forecasts (Technical Report PTR-1092-81-12). Woodland Hills, CA: Perceptronics.

Fischoff, B., Slovic, P., & Lichtenstein, S (1977). Knowing with certainty: The appropriateness of extreme confidence. Journal of Experimental Psychology: Human Perception and Performance, 3, 552-564.

Fontenella, G. A. (1983). The effect of task characteristics on the availability heuristic for judgments of uncertainty (Report No. 83-1). Office of Naval Research, Rice University.

Gentner, D., & Stevens, A. L. (Eds.). (1983). Mental models. Hillsdale, NJ: Erlbaum.

Gettys, C. F. (1983). Research and theory on predecision processes (Tech. Rep. TR 11-30-83). Norman, OK: University of Oklahoma, Department of Psychology.

Hockey, G. R. (1970). Signal probability and spatial location as possible bases for increased selectivity in noise. Quarterly Journal of Experimental Psychology, 22, 37-42.

Irizarry, V., & Knapp, B. (1986). A preliminary investigation of problem solving and judgmental strategies of expert military intelligence personnel. In Proceedings, Psychology in the Department of Defense. Colorado Springs, CO: United States Air Force Academy.

Jensen, R. J. (1981). Prediction and quickening in prospective flight displays for curved landing and approaches. Human Factors, 23, 333-364.

Jensen, R. S., & Benel, R. A. (1977). Judgment evaluation and instruction in civil pilot training (Final Report FAA-RD-78-24). Springfield, VA: National Technical Information Service.

Johnson, E. M., Cavanagh, R. C., Spooner, R. L., & Samet, M. G. (1973). Utilization of reliability measurements in Bayesian inference: Models and human performance. IEEE Transactions on Reliability, 22, 176-183.

Kagan, J. (1966). Reflection-impulsivity: The generality and dynamics of conceptual tempo. Journal of Abnormal Psychology, 71, 17-24.

Kagan, J., Rosman, B., Day, D., Albert, J., & Phillips, W. (1964). Information processing in the child: Significance of analytic and reflective attitudes. *Psychological Monographs*, *78*, (1, Whole No. 578).

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*, 251-273.

Lester, L. F., Diehl, A. E., & Buch, G. (1985). Private pilot judgment training in flight school settings: A demonstration project. *Proceedings of the Third Symposium on Aviation Psychology*. Columbus, OH: The Ohio State University.

Levison, W. H. (1982). The optimal control model for the human operator: Theory, validation, and application. In M. L. Frazier & R. B. Crombie (Eds.), *Proceedings of the Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics*. Edwards Air Force Base, CA: Air Force Flight Test Center.

Lichtenstein, S., Slovic, P., Fischoff, B., Layman, M., & Coombs, B. (1978). Judged frequency of lethal events. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 551-578.

McRuer, D. T. (1980). Human dynamics in man-machine systems. *Automatica*, *16*, 237-253.

McRuer, D. T., & Jex, H. R. (1967). A review of quasi-linear pilot models. *IEEE Transactions on Human Factors in Electronics*, *8*, 231.

Moray, N. (1986). Monitoring behavior and supervisory control. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), Handbook of perception and human performance: Volume II. Cognitive processes and performance. New York: Wiley.

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. Quarterly Journal of Experimental Psychology, 29, 85-95.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Brent, D. H. (1978). SPSS - Statistical package for the social sciences (2nd Ed.). New York: MacGraw-Hill.

Norusis, M. J. (1986). SPSS/PC+ for the IBM PC/XT/AT. Chicago, IL: SPSS Incorporated.

Pitz, G. F. (1965). Response variability in the estimation of relative frequency. Perceptual and Motor Skills, 21, 867-873.

Pitz, G., & Sachs, N. (1984). Judgment and decision: Theory and application. Annual Review of Psychology, 35, 139-163.

Rasmussen, J. (1981). Models of mental strategies in process plant diagnosis. In J. Rasmussen & W. B. Rouse (Eds.), Human detection and diagnosis of system failures. New York: Plenum Press.

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. Psychological Bulletin, 100(3), 349-363.

Schank, R. C., & Abelson, R. P. (1977). Scripts, plans, goals and understanding. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schum, D. (1975) The weighing of testimony of judicial proceedings from sources having reduced credibility. Human Factors, 17, 172-203.

Schvaneveldt, R. W., Durso, F. T., Goldsmith, T. E., Breen, T. J., & Cooke, N. M. (1985). Measuring the structure of expertise. International Journal of Man-Machine Studies, 23, 699-728.

Slovic, P. (1984). Facts vs. fears: Understanding perceived risk. Presented at a Science and Public Policy Seminar, Washington, DC.

Slovic, P., Fischoff, B., & Lichtenstein, S. (1976). Cognitive processes and societal risk taking. In J. S. Carroll & J. W. Payne (Eds.), Cognition and social behavior. Hillsdale, NJ: Erlbaum.

Slovic, P., Fischoff, B., & Lichtenstein, S. (1977). Behavioral decision theory. Annual Review of Psychology, 28, 1-39.

Stone, R. B., Babcock, C. L., & Edmunds, M. S. (1985). Pilot judgment: An operation viewpoint. Aviation, Space, and Environmental Medicine, 149-152.

Tatsuoka, M. M. (1976). Selected Topics in Advanced Statistics: Validation Studies. Champaign, IL: Institute for Personality and Ability Testing.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124-1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. Science, 211, 453-458.

Wallsten, T. S., & Barton, C. (1982). Processing probabilistic multidimensional information for decisions. Journal of Experimental Psychology: Learning, Memory, and Cognition, 8(5), 361-383.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. Quarterly Journal of Experimental Psychology, 12, 129-140.

Wason, P. C., & Johnson-Laird, P. N. (1972). Psychology of reasoning: Structure and content. London: Batsford.

Wickens, C. D. (1984). Engineering psychology and human performance. Columbus, OH: Charles Merrill.

Wright, R. E. (1974). Aging, divided attention, and processing capacity. Journal of Gerontology, 36, 605-614.

ibm>karen>chris>midisfin.rep/07-19-88

APPENDIX A: Risk Assessment and Utilization Test

Name ______________________________

Subject Number ______________________

Risk Assessment and Utilization Test

The following test is designed to measure how accurately you assess the riskiness of a situation, and how you use this information in selecting your preference between alternative courses of action.

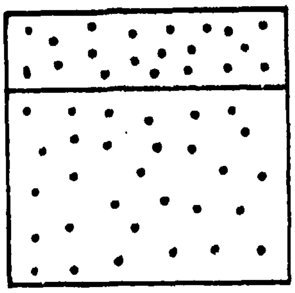
Answer the following questions as quickly and accurately as possible. Do not try to do any actual mathematical calculations. Rather, make your responses based upon your general impression and situation assessment. Please note that there are no "right" or "wrong" answers, and that your test score wil] be maintained in complete anonymity.

You will have 10 minutes in which to complete this test. When you finish each part, move on to the next, until you have completed all 4 parts of the test. Are there any questions?

PART I

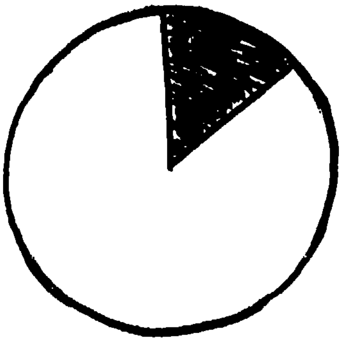
ESTIMATE the following answers after a brief glance at the test picture. Do not try to do any actual counting or mathematical calculations.

1.



What percent of the dots are above the line? ______________

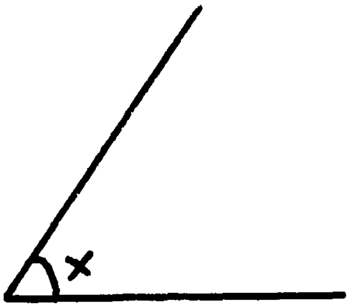
2.



What percent of the circle is shaded? ______________

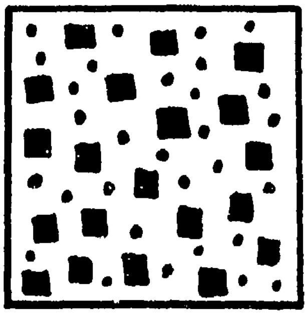
3. 14/250 is equivalent to x/75. x = ______________

4.



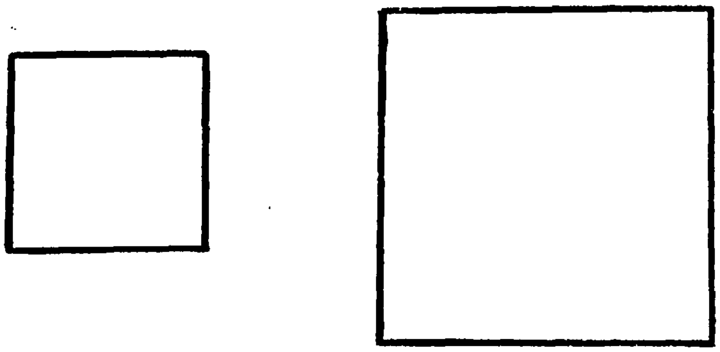
What degree angle is this? ____________

5.



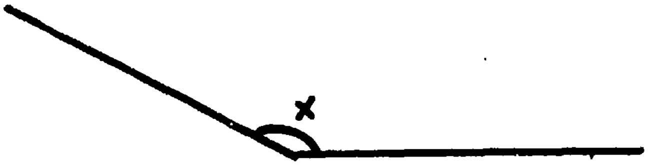
What percent of the figures in the box are squares? ____________

6.



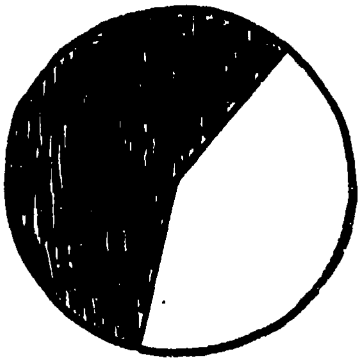
The second square is ____________ times larger than the first.

7.



What degree angle is this? ____________

8.



What percent of the circle is shaded? ____________

PART II

Based upon the frequency of death per 100,000 U.S. residents (It may help to think of 100,000 residents as the population of a town like Champaign-Urbana), how many do you think will die in one year from each of the following causes? Mark an "X" on the response scale to indicate your estimate.

1. motor vehicle accident |____|____|____|____|____|____|____|
.01 .1 1 10 100 1000

2. cancer (any type) |____|____|____|____|____|____|____|
.01 .1 1 10 100 1000

3. drowning |____|____|____|____|____|____|____|
.01 .1 1 10 100 1000

4. electrocution |____|____|____|____|____|____|____|
.01 .1 1 10 100 1000

5. accidental fall |____|____|____|____|____|____|____|
.01 .1 1 10 100 1000

6. firearm accident |____|____|____|____|____|____|____|
.01 .1 1 10 100 1000

7. vehicle-train collision |____|____|____|____|____|____|____|
.01 .1 1 10 100 1000

8. homicide |____|____|____|____|____|____|____|
.01 .1 1 10 100 1000

9. fire |____|____|____|____|____|____|____|
.01 .1 1 10 100 1000

10. tornado |____|____|____|____|____|____|____|
.01 .1 1 10 100 1000

PART III

Respond with your best estimate to the following questions pertaining to aviation-related accidents. To aid you in these estimations, the following statistics may serve as benchmarks: In an average school year, the University of Illinois pilot training fleet logs approximately 10,500 hours. In one year, the total number of flight hours logged by all major commercial airlines is 7.4 million.

1. For all aircraft, how many accidents (both fatal and non-fatal) occur per 100,000 aircraft hours? ______________

2. For all aircraft (general aviation pilots), how many <u>fatal</u> crashes occur per 100,000 aircraft hours? ______________

3. For the normal pilot population, what percentage of accidents may be attributed to continuing flight into deteriorating weather without an IFR flight plan? ______________

4. What percentage of accidents are attributed to fuel exhaustion? ______________

5. Out of all fuel exhaustion accidents, what percentage occurred within 0 to 1 mile of the pilot's destination? ______________

6. During instructional flight, how many accidents (both fatal and non-fatal) occur per 100,000 aircraft hours? ______________

7. For the normal pilot population, what percentage of accidents were attributed to stalls? ______________

8. During instructional flight, how many fatal accidents occur per 100,000 aircraft hours? ______________

9. What percentage of all normal pilot population accidents were collisions (taxiway, runway, or mid-air)? ______________

10. How many total aircraft accidents occur per 100,000 hours for personal and business flights (excluding corporate/executive, instructional and aerial application flights)? ______________

PART IV

In each of the following cases, you will be given 2 alternative courses of action. Mark an "X" on the response scale indicating the degree of preference for 1 option over the other. The neutral point indicates no preference for either option.

1. A) Buy a random card from a standard deck of 52 for $2.00. If the card drawn is a heart, you win $10. If not, you lose your $2.00.

   B) I will give you $1.00.

   Which would you prefer?

   |____________|____________|____________|____________|
   A No Preference B

2. A) Roll a die. If a 1, 2, or 3 is rolled, I will pay you $3.00. If a 4, 5, or 6 is rolled, you pay me $2.00.

   B) Roll a die. If a 3 is rolled, I will pay you $3.00.

   Which would you prefer?

   |____________|____________|____________|____________|
   A No Preference B

3. A) I will pay you $1.50.

   B) Flip a coin. If it's a head, you win $5.00. If it's a tail, you lose $2.00.

   Which would you prefer?

   |____________|____________|____________|____________|
   A No Preference B

4. A) You pay me $2.50.

B) Flip a coin. Heads you win $5.00. Tails you pay me $10.

Which would you prefer?

|__________|__________|__________|__________|

A No Preference B

5. A) Buy a lottery ticket for $10 with a 1/100 chance of winning $1,000.

B) Keep your $10 and buy no ticket at all.

Which would you prefer?

|__________|__________|__________|__________|

A No Preference B

6. A) Draw a card from a standard deck of 52 cards. If a card is a spade, you win $10. If it is a heart, you pay me $5.00. If it is a club, you pay me $1.00. If it is a diamond, you gain or lose nothing.

B) I will give you $1.00.

Which would you prefer?

|__________|__________|__________|__________|

A No Preference B

7. You make a $100 investment in the stock market. Unfortunately, shortly after purchasing the stock, its value dropped substantially. Would you:

A) Sell the stock for a net loss of $40.

B) Wait for a 10% chance of gaining a $500 profit, but a 90% chance of losing the total $100 investment.

Which would you prefer?

|__________|__________|__________|__________|

A No Preference B

8. A) Pay me $20 for a 1/10 chance of winning $100.

B) Pay me $10 with no chance of winning.

Which would you prefer?

|__________|__________|__________|__________|

A No Preference B

9. A) Roll a die. If a 1, 2, or 3 is rolled, you win $10. If a 4, 5, or 6 is rolled, you pay me $10.

B) Flip a coin. If it's a head, you win $1.00. If it's a tail, you pay me $1.00.

Which would you prefer?

|__________|__________|__________|__________|

A No Preference B

10. A) Do not gamble with a die at all.

B) Roll a die. If it's a 5 or 6, you win $10. Anything else, you lose $5.00.

Which would you prefer?

|__________|__________|__________|__________|

A No Preference B

APPENDIX B: Aviation Declarative Knowledge Test

AVIATION RESEARCH LABORATORY MIDIS PROJECT
April 1987
University of Illinois

This test has no official standing and is solely for the research purposes of ARL. Individual scores on the test constitute research data, and as such will be treated as confidential to the individual and the research staff of the MIDIS project.

[1] 7451. What responsibility does the pilot in command of an IFR flight assume upon entering VFR conditions?

(1) Advise ATC when entering VFR conditions.

(2) Report VFR conditions to ARTCC so that an amended clearance may be issued.

(3) Use VFR operating procedures.

(4) To see and avoid other traffic.

[2] 7463. If you are departing from an airport where you cannot obtain an altimeter setting, you should set your altimeter

(1) on 29.92" Hg.

(2) on zero ft.

(3) on the current airport barometric pressure, if known.

(4) to the airport elevation.

[3] 7049. When departing from an airport located outside controlled airspace during IFR conditions, you must file an IFR flight plan and receive a clearance before

(1) takeoff.

(2) entering IFR conditions.

(3) entering controlled airspace.

(4) arriving at the en route portion of the flight.

[4] 7024. Before beginning any flight under IFR, the pilot in command must become familiar with all available information concerning that flight. In addition, the pilot must

(1) list an alternate airport on the flight plan and become familiar with the instrument approaches to that airport.

(2) list an alternate airport on the flight plan and confirm adequate takeoff and landing performance at the destination airport.

(3) be familiar with all instrument approaches at the destination airport.

(4) be familiar with the runway lengths at airports of intended use, and the alternatives available if the flight cannot be completed.

[5] 7039. You check the flight instruments while taxiing and find that the VSI (vertical speed indicator) indicates a descent of 100 ft./min. In this case, you

(1) must return to the parking area and have the instrument corrected by an authorized instrument repairman.

(2) may take off and use 100 ft. descent as the zero indication.

(3) may not take off until the instrument is corrected by either the pilot or a mechanic.

(4) may take off without any correction because this instrument is used very little during instrument flight.

[6] 7070. Which sources of aeronautical information, when used collectively, provide the latest status of airport conditions (e.g., runway closures, runway lighting, snow conditions)?

(1) Airman's Information Manual, Aeronautical Charts, and Distant (D) NOTAMS.

(2) Airport Facility Directory, FDC NOTAMS, and Local (L) NOTAMS.

(3) Airport Facility Directory, Distant (D) NOTAMS, and Local (L) NOTAMS.

(4) Standard Instrument Approach Procedures, FDC NOTAMS, and Airman's Information Manual.

[7] 7077. What are the minimum weather conditions that must be forecast to list an airport as an alternate when the airport has no approved instrument approach procedure?

(1) The ceiling and visibility at ETA, 2,000 ft. and 3 miles, respectively.

(2) The ceiling and visibility from 2 hours before until 2 hours after ETA, 2,000 ft. and 3 miles, respectively.

(3) The ceiling and visibility from 2 hours before until 2 hours after ETA, 1,000 ft. above the highest obstacle, and 3 miles, respectively.

(4) The ceiling and visibility at ETA must allow descent from MEA, approach, and landing, under basic VFR.

[8] 7166. The absence of a visibility entry in a Terminal Forecast specifically implies that the surface visibility

(1) exceeds basic VFR minimums.

(2) exceeds 10 miles.

(3) exceeds 6 miles.

(4) is at least 15 miles in all directions from the center of the runway complex.

[9] 7192. What is the significance of the "F2" in the remarks portion of this Surface Aviation Weather Report for CLE?

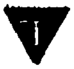
CLE SP 1350 - X E80 BKN 150 OVC 1GF
169/67/67/2105/003/R23LVV11/2 F2

(1) The restriction to visibility is caused by fog and the prevailing visibility is 2 statute miles.

(2) The partial obscuration is caused by fog and the visibility value is variable, 1-1/2 to 2 statute miles.

(3) Fog is obscuring 2/10 of the sky.

(4) The surface based obscuration is caused by fog and is 200 ft. thick.

[10] 7449. What does the symbol  in the minimums section for a particular airport indicate?

(1) Takeoff minimums are 800 ft. and 2 miles.

(2) Takeoff minimums are 1 mile for aircraft having two engines or less and 1/2 mile for those with more than two engines.

(3) Instrument takeoffs are not authorized.

(4) Takeoff minimums are not standard and/or departure procedures are published.

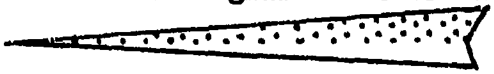
[11] 7249. To which maximum service volume distance from the MFR VORTAC should you expect to receive adequate signal coverage for navigation at the flight planned altitude?

(1) 130 NM.

(2) 100 NM.

(3) 80 NM.

(4) 40 NM.

[12] 7345. What is the significance of the following symbol at Grice intersection?  SEE MAP ATTACHED

(1) It signifies a localizer only approach is available at Harry P. Williams Memorial.

(2) The localizer has an ATC function in addition to course guidance.

(3) GRICE intersection also serves as the FAF for the ILS approach procedure to Harry P. Williams Memorial.

(4) It signifies that the 236° course is a back course approach procedure.

[13] 7373. When departing from an airport not served by a control tower, the issuance of a clearance containing a void time indicates that

(1) ATC will assume the pilot has not departed if no transmission is received before the void time.

(2) the pilot must advise ATC as soon as possible, but no later than 30 minutes, of their intentions if not off by the void time.

(3) ATC will protect the airspace only to the void time.

(4) the pilot must contact FSS and file a flight plan not later than the void time specified in the clearance.

[14] 7376. Which distance is displayed by the DME indicator?

(1) Slant range distance in nautical miles.

(2) Slant range distance in statute miles.

(3) The distance from the aircraft to a point at the same altitude directly above the VORTAC.

(4) Line of sight direct distance from aircraft to VORTaC in statute miles.

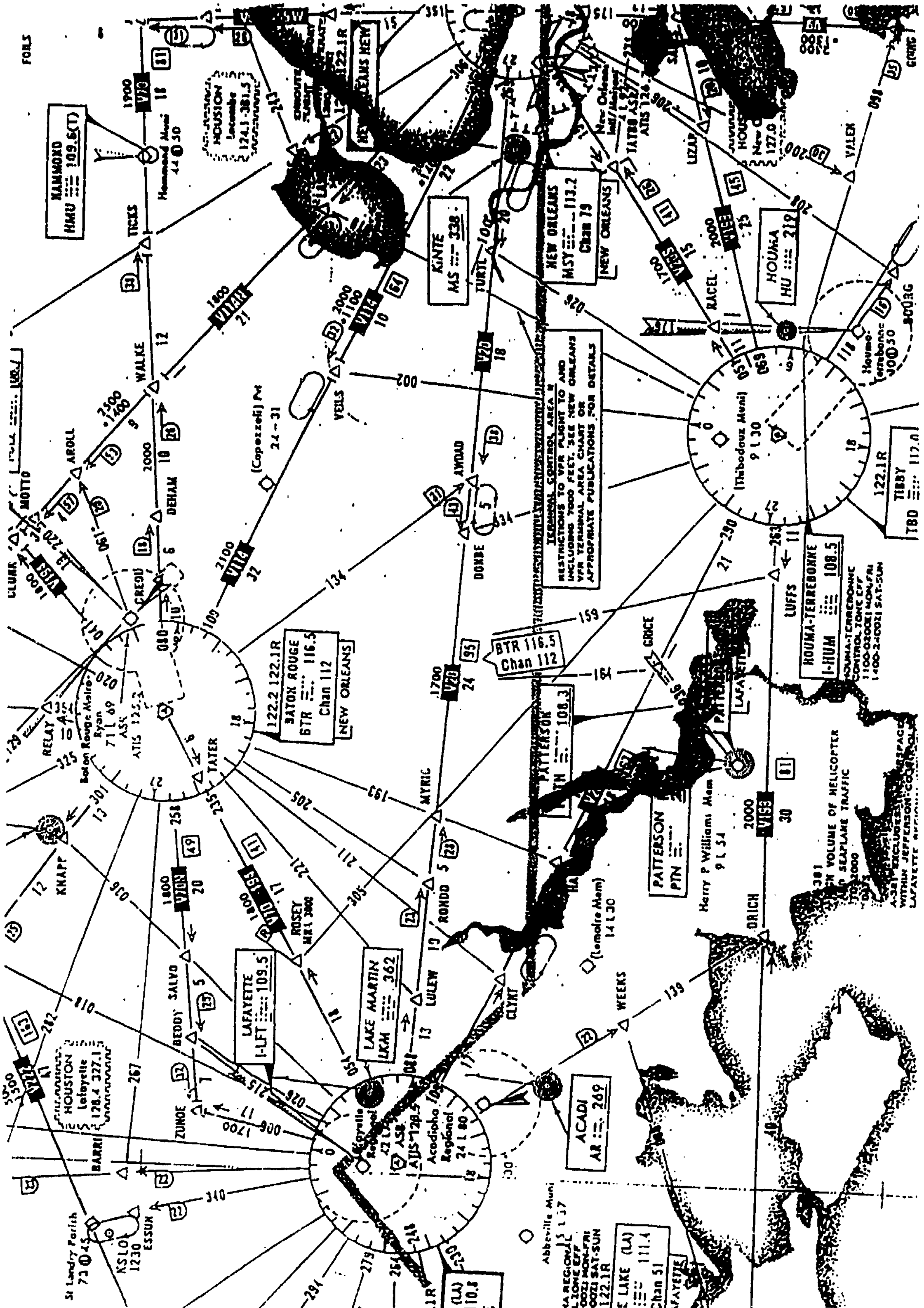
[15] 7388. What service is provided by departure control to an IFR flight when operating from an airport with a terminal radar service area (Stage III)?

(1) Separation from all aircraft operating in the TRSA.

(2) Position and altitude of all traffic within 2 miles of the IFR pilot's line of flight and altitude.

(3) Position of all participating VFR aircraft within the airport traffic area.

(4) Separation from all IFR aircraft and participating VFR aircraft.


HAMMOND
HMU 109.6(T)
Hammond Muni
HOUSTON
Lacombe
124.1 381.5
V70
TICKS
WALKE
AROLL
MOTTO
DEHAM
CREOU
CLURK
V114
VEILS
(Capezzoli) Pvt
KINTE
MS 338
NEW ORLEANS
MSY 113.2
Chan 79
NEW ORLEANS
TERMINAL CONTROL AREA
TURTL
V20
AWDAD
DONDE
RACEL
HOUMA
HU 219
Houma-Terrebonne
(Thibodaux Muni)
TIBBY
TBD 117.0
LUFFS
HOUMA-TERREBONNE
I-HUM 108.5
BTR 116.5
Chan 112
GRICE
PATTERSON
PTN
Harry P Williams Mem
BATON ROUGE
BTR 116.5
Chan 112
NEW ORLEANS
Baton Rouge Metro Ryan
ATIS 125.2
RELAY
TATER
KNAPP
MYRIC
ROSEY
RONDO
LULEW
CLYNT
(Lemoine Mem)
ORICH
WEEKS
SALVO
BEDDY
LAFAYETTE
I-LFT 109.5
LAKE MARTIN
LKM 362
ZUNOE
BARRI
HOUSTON
Lafayette
128.4 327.1
Lafayette Regional
Acadiana Regional
ACADI
AR 269
Abbeville Muni
St Landry Parish
KSLO
ESSUN
WHITE LAKE (LA)
LLA 111.4
Chan 51
LAFAYETTE
V222

[16] 7402. What does the ATC term "Radar Contact" signify?

(1) Your aircraft has been identified and you will receive separation from all aircraft while in contact with this radar facility.

(2) Your aircraft has been identified on the radar display and radar flight following will be provided until radar identification is terminated.

(3) You will be given traffic advisories until advised the service has been terminated or that radar contact has been lost.

(4) ATC is receiving your transponder and will furnish vectors and traffic advisories until you are advised that contact has been lost.

[17] 7408. What is the definition of MEA (Minimum En Route Altitude)?

(1) An altitude which meets obstacle clearance requirements, assures acceptable navigation signals from more than one VORTAC, and assures accurate DME mileage.

(2) The lowest published altitude which meets obstacle clearance requirements, assures acceptable navigational signal coverage, and two-way radio communications.

(3) The lowest published altitude which meets obstacle requirements, assures acceptable navigational signal coverage, two-way radio communications, and provides adequate radar coverage.

(4) An altitude which meets obstacle clearance requirements, assures acceptable navigation signal coverage, two-way radio communications, adequate radar coverage, and accurate DME mileage.

[18] 7414. Reception of signals from an off-airway radio facility may be inadequate to identify the fix at the designated MEA. In this case, which altitude is designated for the fix?

(1) MRA.

(2) MAA.

(3) MCA.

(4) MOCA.

[19] 7441. You enter a holding pattern at a fix, not the same as the approach fix, and receive an EFC time of 1530. At 1520 you experience complete two-way communications failure. Which procedure should you follow to execute the approach to a landing?

(1) Depart the holding fix to arrive at the approach fix as close as possible to the EFC time and complete the approach.

(2) Depart the holding fix at the EFC time, and complete the approach.

(3) Depart the holding fix at the EFC time or earlier if your flight planned ETA is before the EFC.

(4) Depart the holding fix to arrive about 2 minutes ahead of the EFC and then enter a holding pattern at the final fix and adjust pattern to leave the fix inbound at the EFC.

[20] 7041. When making an airborne VOR check, what is the maximum allowable tolerance between the two indicators of a dual VOR system (units indepedent of each other except the antenna)?

(1) 4° between the two indicated bearings to a VOR.

(2) Plus or minus 4° when set to identical radials of a VOR.

(3) 6° between the two indicated radials of a VOR.

(4) 4° when set to identical radials of a VOR.

[21] 7042. What is the oxygen requirement for an unpressurized airplane at 15,000 ft.?

(1) All occupants must use oxygen for the entire time at this altitude.

(2) Crew must start using oxygen at 12,000 ft. and passengers at 15,000 ft.

(3) Crew must use oxygen for the entire time above 14,000 ft. and passengers must be provided supplemental oxygen only above 15,000 ft.

(4) Crew must start using oxygen at 12,500 ft. and passengers must be provided supplemental oxygen at 14,000 ft.

[22] 7071. Where are the compulsory reporting points, if any, on a direct flight not flown on radials or courses of established airways or routes?

(1) Fixes selected to define the route.

(2) The points where the direct course crosses an airway.

(3) There are no compulsory reporting points unless advised by ATC.

(4) At the COP (changeover points).

[23] 7104. Unsaturated air flowing upslope will cool at the rate of approximately

(1) 3°C per 1,000 ft.

(2) 2°C per 1,000 ft.

(3) 2.5°C per 1,000 ft.

(4) 4.4°C per 1,000 ft.

[24] 7125. Which is a characteristic of low level wind shear as it relates to frontal activity?

(1) The amount of wind shear in cold fronts is much greater than found in warm fronts.

(2) With a warm front, the most critical period is before the front passes the airport.

(3) With a cold front, the most critical period is just before the front passes the airport.

(4) With a cold front, the problem ceases to exist after the front passes the airport.

[25] 7197. A Surface Analysis Chart depicts

(1) actual pressure systems, frontal locations, cloud tops, and precipitation at the time shown on the chart.

(2) frontal locations and expected movement, pressure centers, cloud coverage, and obstructions to vision at the time of chart transmission.

(3) actual frontal positions, pressure patterns, temperature, dew point, wind, weather, and obstructions to vision at the valid time of the chart.

(4) actual pressure distribution, frontal systems, cloud heights and coverage, temperature, dew point, and wind at the time shown on the chart.

U S GOVERNMENT PRINTING OFFICE 1988 · 548-[illegible]/[illegible]14