DOCUMENT RESUME

ED 232 056 CE 036 587

AUTHOR	Schipper, Lowell M.; Doherty, Michael
TITLE	Decision Making and Information Processing under Various Uncertainty Conditions. Final Report.
INSTITUTION	Bowling Green State Univ., Ohio. Dept. of Psychology.
SPONS AGENCY	Air Force Human Resources Lab., Williams AFB, Ariz.
REPORT NO	AFHRL-TR-83-19
PUB DATE	Aug 83
CONTRACT	F33615-81-K-0014
NOTE	57p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Adult Education; *Aircraft Pilots; Cognitive Processes; *Decision Making; Higher Education; *Military Personnel; Military Service; Psychological Studies; Reliability; *Undergraduate Students
IDENTIFIERS	*Air Force; *Uncertainty

ABSTRACT

Seven experiments were conducted concerning decision making and information processing under conditions of uncertainty. Several different experimental tasks were used; all presented the subject with multiple independent sources of information regarding the likelihood that some event would occur. Study 1 subjects were Air Force pilots; all other subjects were undergraduate college students. The independent variables included the number of inputs, the inputs' reliability, the effects of discrepant inputs, the format of the problem, and the time available to respond. All tasks and subsequent analyses were non-Bayesian, yielding both normative and idiographic information. The subjects' basic instruction was to indicate the event most likely to occur or to estimate the likelihood of a given event. Results indicated no significant differences between F-16 pilots and student pilots in their use of averaging as the predominant strategy chosen. Simpler strategies were adopted when the time allocated for the task was reduced. The consistent use of a strategy was disrupted when the subject's first experience to the task was under time restrictions. When postdecision feedback was unreliable, the consistency of subsequent decision-making patterns was disrupted. The effects of information reliability were equivocal. Unreliable information was sometimes incorporated into the decision-making processes and sometimes ignored. (Author/YLB)

CE

AFHRL-TR-83-19

AIR FORCE HUMAN RESOURCES LABORATORY

DECISION MAKING AND INFORMATION PROCESSING UNDER VARIOUS UNCERTAINTY CONDITIONS

By

Lowell M. Schipper
Michael Doherty

Psychology Department
Bowling Green State University
Bowling Green, Ohio 43304

OPERATIONS TRAINING DIVISION
Williams Air Force Base, Arizona 85224

August 1983

Final Report

Approved for public release; distribution unlimited.

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235

TE036587

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Public Affairs Office has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report has been reviewed and is approved for publication.

ELIZABETH MARTIN
Contract Monitor

MILTON E. WOOD, Technical Director
Operations Training Division

ALFRED A. BOYD, JR., Colonel, USAF
Commander

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS
BEFORE COMPLETING FORM

1. REPORT NUMBER
AFHRL-TR-83-19

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE *(and Subtitle)*
DECISION MAKING AND INFORMATION PROCESSING UNDER VARIOUS UNCERTAINTY CONDITIONS

5. TYPE OF REPORT & PERIOD COVERED
Final

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR *(s)*
Lowell M. Schipper
Michael Doherty

8. CONTRACT OR GRANT NUMBER *(s)*
F33615-81-K-0014

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Psychology Department
Bowling Green State University
Bowling Green, Ohio 43304

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
61102F
2313T328

11. CONTROLLING OFFICE NAME AND ADDRESS
HQ Air Force Human Resources Laboratory (AFSC)
Brooks Air Force Base, Texas 78235

12. REPORT DATE
August 1983

13. NUMBER OF PAGES
54

14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)*
Operations Training Division
Air Force Human Resources Laboratory
Williams Air Force Base, Arizona 85224

15. SECURITY CLASS *(of this report)*
Unclassified

15.*a*. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of this abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*
decision making
information processing
reliability
uncertainty

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

Seven experiments were conducted in the area of decision making and information processing under conditions of uncertainty. Several different experimental tasks were used, all of which presented the subject with multiple independent sources of information regarding the likelihood that some event would occur. The input was a probabilistic estimate, and the problems had no "correct" answers. Study I used Air Force pilots as subjects. All other studies used undergraduate college students. The independent variables included the number of inputs, the reliability of the inputs, the effects of discrepant (i.e., outlying) inputs, the format of the problem, and the time available to respond. All tasks and subsequent analyses were non-Bayesian, yielding both normative and idiographic information. The

DD Form 1473 EDITION OF 1 NOV 65 IS OBSOLETE
1 Jan 73

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Date Entered)*

Item 20 (*Continued*)

subject's instruction varied somewhat with each study but basically was to indicate which event was most likely to occur or to make an estimate of the likelihood of a given event.

The results of these studies indicated no significant differences between F-16 pilots and student pilots in their use of averaging as the predominant strategy chosen. Also simpler strategies (e.g. cue summing) are adopted when the time allocated for the task is reduced, and the consistent use of a strategy is disrupted when the subject's first exposure to the task is under time restrictions. The effects of unreliable feedback were studied in an information seeking diagnostic task. The results indicated that when the post decision feedback was unreliable, the consistency of subsequent decision-making patterns was disrupted. The effects of information reliability were equivocal in that sometimes unreliable information was incorporated into the decision making processes in predictable patterns, and in other instances, the information labelled as unreliable was simply ignored.

SUMMARY

Objective

The objectives were (a) to identify the types of decision-making strategies used by college students and Air Force pilots when processing probabilistic information, (b) to relate the types of strategies to the reliability, probabilistic distribution, and format of the information available for basing a decision, (c) to ascertain the effects of differentially reliable feedback on information selection strategies, (d) to assess the impact of time limitations or strategy choice and accuracy, and (e) to develop new experimental paradigms that model different aspects of decision making and information processing under a wide variety of uncertainty conditions.

Background/Rationale

Little is known about how people process information, form strategies, and make decisions in situations containing unreliable, contradictory, or uncertain information. Many in-flight piloting situations are of this type, particularly in tactical situations. Results from previous decision-making research suggest that fighter pilots tend to adopt strategies similar to college students, but pilots appear to be more consistent in their use of a given strategy. The present effort explores the apparent difference in decision-making consistency between experienced pilots and college students.

Approach

Seven separate experiments were conducted in which some aspect of the information available to the subject for decision making was varied. Experiment 1 investigated the differences between separate groups of Air Force pilots at the beginning, middle, or near completion of Undergraduate Pilot Training (UPT) on a task requiring them to indicate which of two events was more likely to occur based upon independent probability estimates of the events. The data from the UPT pilots were compared with data collected previously from F-16 and F-15 pilots. The remaining experiments were conducted at Bowling Green State University using college students as subjects. Each experimental task was designed to model selected aspects of real-world decision making such as the time available for deciding information reliability, the presentation format, and feedback consistency.

Specifics

Method. Experiment 1 was conducted at Williams AFB using three groups of UPT trainees (28 beginning, 27 intermediate, 21 advanced) on a task requiring the subject to indicate which of two events (A or B) they thought would be the most likely to occur based on several separate (i.e., independent) probability estimates for each event. The number of pieces of information was either 3, 5, or 7. The task was self-paced. The candidate strategies were averaging, adding, largest cue, and most cue. Experiment 2 used the same paradigm but varied the reliability of the probability estimates by designating each as either of high, medium, or low reliability. Additionally, the arrays contained outlier estimates. Nineteen college students participated in this effort. Experiment 3, using a similar paradigm, manipulated the time available to make a decision. The effects of time restrictions were compared to self-paced decision making. Twenty-two college students served as subjects. Experiment 4 adopted a different paradigm by asking the subjects to make an estimate of the overall likelihood of occurence. Four array variables were manipulated: (a) the probabilistic distance between outlying and clustered estimates, (b) the direction of outlying estimates relative to the clusters, (c) the density of the clusters, and (d) the symmetry of the array. The task was self-paced using 20 college students as subjects. Experiment 5 varied the format of an array of estimates in either a histogram, list form, or geometric numeric format. The task was to indicate the average value of the array. The cue sets had either three or five sources of information, the presentation times were either 3, 6, or 9 seconds. Fifteen college students served as subjects. Experiments 6 and 7 used an entirely different experimental task in which the subject was to decide which of two diseases was present. The subject could ask for information that would aid in making a medical diagnosis. The information was arranged in such a fashion as to be either valuable or not in certain combinations with other information. Both experiments used nursing students as subjects (29 in experiment 6 and 13 in experiment 7). In both efforts the reliability of the feedback was varied according to predetermined rules.

Findings and Discussion. The results indicated that all the groups of pilots tended to use the averaging strategy most often and that they did not differ in their consistency of strategy use. Their response patterns were similar to previously collected data on F-15 and F-16 pilots and college students. Results were equivocal with respect to the effects of reliability because it was unclear that the subjects understood the meaning of "reliability." Under certain circumstances the subjects behaved as predicted (e.g., higher weighting given to information of high reliability), and in other circumstances they tended to disregard differences in reliability. The effects of time restrictions were to increase error and to increase the use of simpler strategies (e.g., cue sum rather than cue average). The disruptive effects of time restrictions can be prevented by allowing the subject initially to self-pace. For the kinds of tasks used in this effort, the geometric numeric format results in significantly less subject processing error than histograms or lists. There were strong individual preferences between format types. The data using the diagnostic paradigm indicated that the subjects tend to choose diagnostically worthless information and continue to do so under a variety of feedback conditions.

Conclusions/Recommendations

Based on results from the time limitations experiments, it can be recommended that initial training should be accomplished in a self-paced situation even if the criterion environment will be time limited. Additional research using different experimental tasks needs to be conducted in order to explicate the role of information reliability and consistency. The results from the display format study indicate that display format does make a difference in accuracy and that individuals differ significantly in their ability to use different formats. Therefore, careful consideration should be given to the display in the design phase. The results from the diagnostic experiments indicate that people persist in choosing inappropriate strategies when searching for information. Therefore, training for situations in which a pilot may need to make a "diagnosis" and has several alternative information sources should include the logic of search. Although the series of experiments conducted in this effort succeeded in providing new methods for studying analogues of real world situations, future research should concentrate on still broader paradigms modeled after more specific application areas.

TABLE OF CONTENTS

I. Introduction 5

II. Research Effort 7

III. Study 1. A Comparison of Information Processing and Decision Making Strategies with UPT and F-16 Trainee Populations 9

IV. Study 2. Reliability of Information as a Factor in Decision Making 19

V. Study 3. Time Pressure as a Factor in Strategy Use . . .26

VI. Study 4. Information Use as a Function of Cue Distribution Variables 31

VII. Study 5. Presentation Mode and Allocated Processing Time as Factors in Estimating Numerical Averages . . .36

VIII. Study 6. Two Additional Investigations 41

IX. The Bayesian and Pseudodiagnosticity Paradigms 41

X. Study 6. Reliable and Unreliable Feedback 45

XI. Study 7. Reliable and Unreliable Feedback II 47

XII. Discussion 49

References 52

List of Illustrations

Figure	Page
1	10
2	20, 21
3	32
4	37
5	39

List of Tables

Table	Page
1	12
2	14, 15
3	16
4	24
5	28
6	30
7	34
8	42

I. INTRODUCTION

Problem Area

A fundamental attribute of humans is the ability to process multiple sources of stimulation and develop a single response. One of the many ways of conceptualizing this process has involved postulating that the environmental information, the "inputs," are represented in the cognitive system as discrete probabilities or as probability distributions. In psychological research such as probability learning, the investigator presents information to a person in some form other than probabilities, analyzes the data as though the person had encoded the data into probabilistic form, and then processes those probabilities using some algorithm, normative or otherwise.

There is, also, a substantial number of studies in which the experimental data are presented in essentially probabilistic form and subjects are required to respond with probabilities. This technique, the so-called Bayesian aggregation paradigm, was thoroughly explored in a widely cited article by Peterson and Beach (1967). This model assumes that people revise their opinions of the probability (P) of some event (E) given some data (D) by taking into account the probabilities of the data given the possible event. This is represented by conditional probabilities $P(D/E_1)$, $P(D/E_2)$, ... which are read as "The probability of the data given Event 1, the probability of the data given Event 2," etc. Under the assumptions of this model, a pilot who observes two different indicators of a malfunction and who must make a decision must first retrieve the following information from memory:

P (Accident prior to the Anomalous Readings)

P (Indicator Reading 1, given an Accident)

P (Indicator Reading 1, given no Accident)

P (Indicator Reading 2, given an Accident)

P (Indicator Reading 2, given no Accident)

for each indicator. These probabilities are then aggregated into an estimate of

P (Accident, given both Sets of Readings).

Normative statistical theory, specifically Bayesian theory, processes information in this manner; however, the problem is that people do not. The consensus of the scientific community on this issue seems to be that, except for primitive sensory and perceptual processes, people do not process data in the fashion described. Nevertheless, it is the authors' belief that people process data using probabilistic representation -- at least many people do. How they do it is the important question.

The authors' belief substantiated by preference data, is that people do not aggregate multiple sources of data by combining P(D/E), that is, the probability of data given the event. The belief, rather, is that people, contrary to normative models, aggregate, in a statistical sense, the wrong data. People aggregate multiple sources of information by aggregating P(E/D) values; that is, the probabilities of events given the data.

A pilot might take the probability of an accident given a low altitude reading and integrate it somehow with the probability of an accident given an additional oil pressure problem. The "somehow" is the point of this research.

A researcher could provide data to people in some non-probabilistic form and then theorize as if the people converted the information to P values. Conversely, the researcher could begin with simpler investigations and present P values directly. The strategy adopted in these first studies investigates how people aggregate these values when given P(E/D) values themselves (i.e., numbers) from sources which conflict. Unfortunately, there is no normative model that says how people ought to do this.

II. RESEARCH EFFORT

Most important cognitive tasks require that multiple sources of information be processed into single responses. Quantities of information, ranging from a few inputs to enormous amounts of data, are reduced to a yes-no, a button push, a turn of a wheel, or a simple directive. This process of data reduction has been studied in concept formation, problem solving, information integration, and information processing. Work within these experimental paradigms is furthering the understanding of this fundamental process of information reduction.

In 1978, Jones, Schipper, and Holzworth (JSH) introduced a new paradigm for investigating decision-making and information aggregation, which, unlike apparently related paradigms, presents decision-makers directly with estimates of the likelihoods of events of interest. As noted earlier, this is considered more representative of decision making in the real world. This new paradigm provides for both idiographic and group analyses.

Consider the following arrangement. An observer has several independent sources of information concerning the occurrence of some event. The same observer has several independent and different sources of information concerning the occurrence of a second event. The observer's task is, first, to consider those sources of information concerned with the first event, second, to consider those sources of information concerned with the second event, and then to make a decision as to which of the two events is more likely to occur.

Example 1. A hypothetical illustration of this type of task might be a situation in which one source of information says simply that The probability of war in the Middle East is .30. A second source of information says that the probability of war in the Middle East is .40. A third source of information says that the probability of war in the Middle East is .60. A fourth source of information says that probability of war in Central Africa is .20. A fifth source of information says that the probability of war in Central Africa is .70. The observer must decide in which area war is more likely to occur.

Example 2. An indicator device provides information that the probability of a faulty landing system is .30. A second independent indicator says that the probability of a faulty landing system is .50. A third source of information says that the probability of improper left engine spool-up is .20, and a fourth system status readout shows the probability of an improper left engine spool-up to be .60. Which is the more likely event: landing system failure or engine trouble?

Numerous situations involving the same type of format occur in many decision-making situations when information is presented concerning one or more possible forthcoming events.

Subjects in studies 1, 2 and 3, as well as in certain studies accomplished before contract initiation, used the JSH paradigm. This paradigm involves presenting subjects with a large number of arrays, each carrying information from one or more sources about the probability that event A will occur and information from two or more other sources about the probability that event B will occur. Figure 1 gives three examples of what an array would look like to the subject.

The JSH paradigm is distinguished from apparently similar paradigms in important ways. First, the information is presented numerically, rather than as substantive variables or as diagnostic events. Second, the probabilities are already P(E/D) values, but ones which do not agree with one another. Finally, the analysis relies on obtaining many responses from each subject, then comparing the subject's set of responses with various sets of predicted responses, each set having been predicted based on alternative possible process models. In this way, certain possible descriptive models can be ruled out for given subjects. The conclusions about the process model used are only tentative since other, untested process models may make predictions similar to the ones most consistent with a subject's responses. In other words, some models are ruled out, others are supported, but none can be proved. The analytical procedure is outlined in Figure 1 and described fully in the original JSH paper. The theoretical strategies against which each subject's data were compared were chosen to represent the ones that subjects are likely to use.

III. Study 1. A Comparison of Information Processing and Decision Making Strategies with UPT and F-16 Trainee Populations

In 1979 and 1980, an inventory similar to that of JSH was used at Williams AFB to obtain data from pilots beginning training to fly the F-16 and from some of their instructor pilots. This inventory consisted of 194 items with the numbers of sources of information (cues) ranging from three through seven.

	1			2			3	
A		B	A		B	A		B
	.20	*		.20		*	.20	
	.30	*		.30			.30	*
*	.40		*	.40		*	.40	
	.50	*		.50		*	.50	*
	.60			.60	*		.60	
*	.70		*	.70		*	.70	
	.80			.80			.80	*

	1	2	3
Adding	A	A	A
Averaging	A	B	B
Largest Cue	A	A	B
Most Cues	B	A	A

In the first panel, five probabilistic cues are shown: two indicate that the probabilities of Event A are .40 and .70, while three indicate that the probabilities of Event B are .20, .30, and .50. The choice of Event A is consistent with Adding, Averaging, and the Largest Cue strategies but essentially rules out a Most Cues strategy. Given a subset of items like item 1, consistent choice of Event A allows the investigator to rule out the possibility that an observer was using a Most Cues strategy. Consistent choice of Event B would rule out Adding, Averaging, and Largest Cue. The items shown in panels 2 and 3 provide other arrangements and show how evidence can be amassed for, and even more strongly against, particular stragegies.

Panel 2 shows an item in which the choice of B is consistent with Averaging but may also be consistent with other unspecified models. The choice of B, however, clearly rules out Adding, Largest Cue, and Most Cues strategies if such choice behavior is sufficiently consistent. Panel 3 shows an arrangement in which each choice would be consistent with two strategies and inconsistent with two other strategies.

Figure 1. Three sample items and descriptions of predictions from four possible strategies.

In September 1981, data using the same inventory were obtained from 76 Undergraduate Pilot Trainees (UPTs) at Williams AFB. This sample of trainees was made up of, roughly, 37% Beginning UPTs (UPT-B), 36% Intermediate UPTs (UPT-I), and 28% Advanced UPTs (UPT-A), where the three classifications refer to the length of time the trainees had been in UPT training. This part of the effort compared the information processing and decision making characteristics of a sample from the UPT population (with relatively short training in the Air Force) with an F-16 transitioning group (with relatively long training in the Air Force).

Figure 1 summarizes the types of items included in the inventory. A complete description of item selection is given in JSH. The experimental items were problems similar to those shown in Figure 1, arranged randomly, one to a page, bound in a looseleaf, three-ring binder. Four different random arrangements of the sets of problems were used. The pilots received a printed set of instructions and completed the problems at their own pace. Responses were made on mark sense answer sheets. Pilots were administered the problem set individually or in small groups of not more than four members. Communication within groups and between successive respondents seemed to be minimal.

Table 1 presents a summary of analyses of the 37 F-16 pilots using 3, 5, and 7 cues for making decisions. Entries in Table 1 for each pilot represent scores which have minimum values of 0 and maximum values of 100. These entries show the differences between a pilot's responses and those that would have been 100% consistent with the strategies designated at the top of the table. Hence a large value is a virtual guarantee that the named strategy was not what the subject was thinking about.

Table 1. Indices of Strategy use

for 37 F-16 Pilots in Study 1

Table entries are percentages of pilots' choice responses that were inconsistent with the strategy designated.[a] Indices are presented for three, five, and seven cues and for the mean of those treatments ($\bar{x}$).

	Strategies															
	Adding				Averaging				Largest Cue				Most Cues			
Pilot	3	5	7	$\bar{x}$	3	5	7	$\bar{x}$	3	5	7	$\bar{x}$	3	5	7	$\bar{x}$
1	71	62	41	58	6	2	8	5	32	34	51	39	94	83	67	81
2	45	40	33	39	39	28	21	29	19	30	54	34	61	62	54	59
3	58	55	31	48	19	9	18	15	26	32	46	35	81	77	56	71
4	0	0	3	1	77	64	46	62	45	57	49	51	23	21	28	24
5	77	64	46	62	0	0	3	1	32	36	46	38	100	85	72	86
6	71	64	62	66	6	4	13	8	32	36	36	35	94	85	87	89
7	42	51	36	43	35	13	13	20	10	28	51	30	65	72	62	66
8	65	62	46	58	13	2	3	6	26	34	51	37	87	83	72	81
9	55	38	36	43	23	26	18	22	42	57	62	54	77	60	56	64
10	29	28	21	26	48	36	44	43	16	34	41	30	52	49	36	46
11	68	55	38	54	10	9	15	11	42	32	49	41	90	77	64	77
12	77	62	49	63	0	2	5	2	32	38	49	40	100	83	74	86
13	77	64	51	64	0	0	3	1	32	36	46	38	100	85	77	87
14	0	2	3	2	77	62	51	63	45	55	54	51	23	23	23	23
15	32	26	18	25	45	43	31	40	32	45	49	42	55	57	44	49
16	71	55	41	56	6	17	18	14	26	23	46	32	94	77	67	79
17	77	66	54	66	0	2	5	2	32	34	44	37	100	87	79	89
18	32	21	31	28	84	55	54	64	77	66	72	72	16	30	21	23
19	77	54	49	63	0	0	0	0	32	36	49	39	100	85	74	86
20	68	66	54	63	10	2	5	6	29	34	44	36	90	87	79	85
21	74	62	49	62	3	2	10	5	35	38	44	39	97	83	74	85
22	74	62	46	61	3	2	3	3	29	34	46	36	97	83	72	84
23	26	26	13	22	52	51	46	44	32	53	44	43	48	38	38	41
24	48	47	38	41	29	17	21	22	10	49	59	39	71	68	54	64
25	35	47	33	38	42	17	15	25	16	28	49	31	58	68	59	62
26	71	57	49	59	6	6	5	6	26	30	49	35	94	79	74	82
27	74	60	46	60	3	4	3	3	35	36	51	41	97	81	72	83
28	77	62	49	63	0	2	0	1	32	34	49	38	100	83	74	86
29	58	51	41	50	19	11	13	14	26	28	46	33	81	74	77	74
30	65	64	54	61	13	9	15	12	32	28	38	33	87	85	79	84
31	65	60	49	58	13	4	5	7	32	36	49	39	87	81	74	81
32	58	51	36	48	19	13	13	15	39	36	51	42	81	72	62	72
33	68	72	56	62	10	2	8	7	35	38	41	38	90	83	82	85
34	61	53	38	51	16	15	15	15	16	26	49	30	84	74	59	72
35	71	57	46	58	6	11	3	7	32	38	51	40	94	79	72	82
36	74	57	44	58	3	6	10	6	29	34	49	37	97	89	69	82
37	42	38	36	39	35	30	13	26	42	66	51	53	65	55	62	61
Mean	58	50	39	49	21	16	15	17	31	38	49	39	79	71	63	71
HS[b]	11	16	14	14	72	84	84	84	15	0	0	0	3	0	3	3

[a] Note that these values do not reflect the % of time a strategy was used, since one choice was often consistent with more than one strategy. Pilot 5, for example, probably used an averaging strategy all the time, but made a mistake in one of the seven cue items. The residual from the other strategies are not 100% since on some items, other strategies made the same prediction as the averaging strategy. Thus, the pilots who have all very low values for one strategy may be inferred to have used that strategy, or some unspecified one that makes the same predictions. High values permit one to conclude decisively that particular strategy was not used.

[b] Percentage of pilots for whom the designated strategy had the lowest percentage index, for each cue level and for the average.

Of particular interest are (a) the overall use of strategies, (b) the similarities (or, conversely, the differences) among the pilots, and (c) whether strategies changed as the numbers of cues varied. Table 2 presents the same information as Table 1 for each of the UPT groups. The summaries in Tables 1 and 2 can be compared directly.

Statistical analyses show no reliable overall differences among the four pilot groups (F-16, UPT-B, UPT-I, UPT-A), but do show highly reliable differences among strategy uses. The Averaging strategy is used more frequently than are the Adding and Most Cues strategies. The Largest Cue strategy is used more frequently than is the Most Cues strategy. And the Adding strategy is used more frequently than is the Most Cues strategy.

A reliable interaction of strategy by pilot type shows differential use of strategies according to the type of pilot. That is, all pilot types showed reliably different preferences for strategies, tending heavily to rely on an Averaging strategy, but different pilot types tended to use different strategies when Averaging was not used.

A summary of relative magnitudes of residuals from respective strategy types can be obtained by ranking these residuals for the individual pilots. That is, over three, five, and seven cues combined, the four residuals, one for each strategy, are ranked from 1 through 4. These ranks, pilot group by pilot group, can then be correlated to assess the degree of homogeneity within each group. Essentially, the statistic describing this degree of homogeneity is the Kendall Coefficient of Concordance and these coefficients are shown in Table 3, along with the 11 patterns of ranked residuals generated by the pilots.

The data can be interpreted in a fairly straightforward way. When pilots and pilot trainees are asked to aggregate probabilistic infor-

Table 2. Indices of Strategy use for 76 UPT pilots in Study 1

Table entries are percentages of pilots' choice responses that were inconsistent with the strategy designated[a]

	Adding				Averaging				Largest Cue				Most Cues			
Pilot	3	5	7	$\bar{x}$	3	5	7	$\bar{x}$	3	5	7	$\bar{x}$	3	5	7	$\bar{x}$
Beginning UPT																
1	55	26	26	36	35	51	28	38	68	62	51	60	65	34	46	48
2	61	45	31	46	16	23	18	19	23	21	46	30	84	66	56	69
3	65	60	44	56	13	4	21	13	32	40	38	37	87	81	69	79
4	77	64	49	63	0	0	0	0	32	36	49	39	100	85	74	86
5	65	60	39	55	19	4	16	13	39	32	53	41	81	81	61	74
6	77	64	54	65	0	0	10	3	32	36	38	35	100	85	79	88
7	0	2	0	1	77	62	49	63	45	60	51	52	23	23	26	24
8	77	64	46	62	0	0	3	1	32	36	51	40	100	85	72	86
9	74	62	44	60	3	2	10	5	29	38	49	39	97	83	69	83
10	77	66	46	63	6	2	8	5	32	34	41	36	94	87	72	84
11	52	36	38	42	26	36	41	34	6	21	18	15	74	57	64	65
12	52	36	28	39	26	32	31	30	32	47	54	44	74	53	49	59
13	71	55	41	56	6	9	8	8	32	36	51	40	94	77	67	79
14	77	55	36	56	0	9	13	7	32	32	51	38	100	77	62	80
15	77	64	51	64	0	0	3	1	32	36	46	38	100	85	77	87
16	61	53	36	50	16	15	13	15	23	43	51	39	84	74	62	73
17	74	60	41	58	3	4	8	5	29	32	41	34	97	81	67	82
18	71	62	38	57	6	11	15	11	26	34	54	38	94	83	64	80
19	58	40	31	43	26	40	28	31	52	77	72	66	74	45	46	55
20	74	62	44	60	3	2	10	5	29	34	49	37	97	83	69	83
21	77	64	51	64	0	0	3	1	32	36	46	38	100	85	77	87
22	39	53	55	49	39	15	18	24	39	51	82	57	61	70	75	69
23	42	60	51	51	35	34	49	39	3	2	0	2	65	81	77	74
24	61	49	46	52	16	15	18	16	16	21	36	24	84	70	72	75
25	0	0	10	3	77	64	54	65	45	57	62	55	23	21	21	22
26	65	51	33	50	13	17	15	15	26	45	54	42	87	72	59	73
27	61	54	49	55	16	20	10	15	42	41	38	40	84	72	74	77
28	77	64	54	65	0	4	5	3	32	32	44	36	100	85	79	88
$\bar{x}$	61	51	39	51	17	17	18	17	32	38	46	39	83	71	63	72
HS[b]	8	12	14	11	83	77	79	82	8	11	7	7	0	0	0	0
Intermediate UPT																
29	35	40	31	35	42	23	18	28	10	30	46	29	58	62	56	59
30	52	47	31	43	26	21	23	23	26	36	46	36	74	64	56	65
31	71	62	49	61	13	2	5	7	26	38	49	38	87	83	74	81
32	39	26	15	27	39	38	33	37	6	32	41	26	61	47	41	50
33	55	70	59	61	23	15	26	21	16	34	38	29	77	97	85	83
34	77	64	46	62	0	0	3	1	32	36	46	38	100	85	72	86
35	74	68	49	64	3	4	0	2	35	32	49	39	97	89	74	87
36	32	23	21	25	45	40	44	43	13	34	31	26	55	45	46	49
37	3	4	8	5	[illegible]	60	41	58	48	62	49	53	26	26	33	28
38	77	57	54	63	0	6	10	5	32	38	44	38	100	79	79	86
39	32	26	21	26	45	38	33	39	13	36	41	30	55	47	46	49
40	3	4	13	7	74	64	51	63	42	57	54	51	26	21	23	23
41	68	51	41	53	10	13	8	10	29	40	46	38	90	72	67	76
42	16	26	23	22	61	64	56	60	42	53	64	53	39	26	18	28
43	45	45	36	42	32	19	13	21	39	47	51	46	68	66	62	65
44	77	62	49	63	0	2	0	1	32	34	49	38	100	83	74	86
45	23	34	15	24	55	30	33	38	23	36	41	33	45	55	41	47
46	13	36	18	19	65	40	36	47	32	34	33	33	35	53	44	44

Table 2. Concluded

47	19	28	21	23	58	49	33	47	26	38	36	33	42	45	46	44
48	71	60	59	63	6	9	30	15	26	32	55	38	94	81	70	82
49	35	26	18	26	42	38	31	37	10	36	44	30	58	47	44	50
50	65	47	36	49	13	17	13	14	19	28	41	29	87	68	62	72
51	77	70	64	70	0	6	15	7	32	30	33	32	100	91	90	94
52	61	55	26	47	16	17	23	19	42	53	51	49	84	68	51	68
53	16	19	28	21	68	45	46	53	35	47	38	40	32	40	44	39
54	68	66	56	63	10	23	33	22	23	21	31	25	90	83	77	83
55	52	51	44	49	32	17	21	23	32	28	44	35	68	72	64	68
$\bar{x}$	47	43	34	41	32	26	24	27	27	38	44	36	68	62	57	63
HS^b	24	31	37	37	48	59	56	59	28	7	4	4	0	2	4	0

Advanced UPT

56	74	55	38	56	3	9	10	7	29	36	49	38	97	77	64	79
57	42	38	33	38	35	26	15	25	29	49	49	42	65	60	59	61
58	58	49	38	48	19	15	15	16	39	47	49	45	81	70	59	70
59	71	60	38	56	6	4	15	8	26	32	44	34	94	81	64	80
60	74	60	49	61	3	4	0	2	29	40	49	39	97	81	74	84
61	77	60	49	62	0	4	0	1	32	36	49	39	100	81	74	85
62	23	13	23	20	55	53	31	46	23	51	59	44	45	32	44	40
63	77	64	51	64	0	0	8	3	32	36	41	36	100	85	77	87
64	74	66	54	65	3	6	10	6	35	34	44	38	97	87	79	88
65	58	45	54	52	39	53	51	48	58	47	54	53	61	49	54	55
66	74	64	46	61	3	0	3	2	35	36	46	39	97	85	72	85
67	74	64	44	61	3	0	5	3	29	36	49	38	97	85	69	84
68	77	64	46	62	0	4	8	4	32	40	46	39	100	81	72	84
69	77	64	51	64	0	0	3	1	32	36	46	38	100	85	77	87
70	61	57	44	54	16	6	5	9	23	30	54	36	84	79	69	77
71	65	57	41	54	13	6	8	9	39	30	51	40	87	79	67	78
72	65	49	49	54	13	15	26	18	32	43	38	38	87	70	74	77
73	65	53	38	52	13	11	10	11	32	30	49	37	87	74	64	75
74	29	43	31	34	48	26	18	31	35	57	56	49	52	60	56	56
75	52	36	33	40	26	28	15	23	19	43	54	39	74	57	59	63
76	71	60	46	59	6	4	3	4	32	32	46	37	94	81	72	82
$\bar{x}$	64	53	43	53	14	13	12	13	32	39	49	40	86	73	67	75
HS^b	7	10	5	5	81	90	95	95	12	0	0	0	0	0	0	0

[a,b] See Table 1.

Table 3. Ranked Residuals

Pattern	F-16	Beg.	Int.	Adv.
2-4-3-1	29	20	13	15
4-1-2-3	3	2	8	1
4-2-1-3	2	1		
2-3-4-1		2	1	
3-4-2-1		2	2	3
3-4-1-2		1		1
4-1-3-2			3	
4-2-3-1	1			
4-3-2-1	1			
3-2-1-4	1			
1.5-3-4-1.5				1
	37	28	27	21
w =	.59	.60	.17	.72

This table shows the ranked mean residuals for the Adding strategy (column 1 of the Pattern), Averaging strategy (column 2), Largest Cue strategy (column 3), and Most Cues strategy (column 4) for each of the pilot groups. Thus , the first pattern shows the Averaging strategy (rank 4) to have the lowest residual, the Largest Cue strategy (rank 3) to have the next lowest residual, the Adding strategy (rank 2) to have the next lowest residual, and the Most Cues strategy (rank 1) to have the highest residual.

Only these 11 patterns of residual ranks were used among all 113 pilots.

The lower the residual, the stronger the indication of strategy use.

The entries under each pilot type show the numbers of pilots who gave that particular pattern.

The Coefficient of Concordance (w) gives a measure of the degree of homogeneity within each pilot group.

mation of the form used in this study, most, by far, average the probabilities. However, to reiterate, averaging is not necessarily the "correct," or normative solution. There is no normative solution in the domain of probability theory. If odds were presented and the odds were averaged before converting to P values, very different answers would be obtained than if the odds were first converted to probabilities, then averaged. This is, of course, related to the scale properties of the various methods of encoding uncertainty. There is no particular reason to believe that either objective or subjective uncertainty, when encoded as probabilities, is measured on the interval scale necessary for averaging.

The result, that people generally average, is novel only in the sense that so far as is known, subjects had not previously been placed in this decision situation. It is not surprising, in that averaging behavior has been found by Norman Anderson and his students (1981), among others, to be a nearly ubiquitous form of information aggregation. Perhaps what is most surprising about the data is that against this backdrop of averaging behavior, there are several individuals who consistently used some other form of information aggregation. For example, pilots 4 and 14 in Table 1, and pilots 7, 25, 37, and 40 in Table 2 used a strategy that is almost perfectly predicted by assuming that they simply summed the P(E/D) values. Pilot 24 in Table 2 clearly just went with the event with the largest P(E/D) value. It would be most interesting to know if these highly systematic differences in information processing in this task generalized to other cognitive tasks.

It is also clear from Tables 1, 2, and 3 that the intermediate UPT group was by far the most heterogeneous of the pilot groups. Without replication, it is difficult to interpret such data, since the difference may be essentially a cohort effect.

Tables 1 and 2 show that the JSH methodology was reasonably successful in isolating strategies, though there are clearly subjects whose data are not explained by any of the four hypothesized strategies. Subject 2 in Table 1 is one such case. Other plausible strategies have been considered, and the strategies of still more of the subjects may be identified. A prime possibility is a "Most Cues Over .5" strategy. In general, although the pilots tended to average these event estimates, there are "mavericks" - highly self-consistent, systematic mavericks. But overall these data are basically consistent with the university student data of the original JSH paper.

Study 1 represents that part of the effort which used Air Force pilots and pilot trainees as subjects. All subsequent studies were carried out in the Department of Psychology laboratories at Bowling Green State University, using undergraduate students as experimental subjects.

Two important questions from our earlier work and other research were as follows:

1. How would information be processed for making decisions if all sources of information were not equally reliable? What if one source was more or less reliable than the others? Would decision makers tend to weight information equally, or would they somehow tend to discount certain sources of information and emphasize others?

2. How do decision makers treat information that seems to be at odds with the majority of information they already have? What if one source of information seems to be discrepant in comparison with the rest of the information?

This latter question arose from discussion with students, as well as with F-16 pilots, who seemed to have similar ideas as to how they reacted to an outlying source of information, or what is called an outlier.

IV. Study 2. Reliability of Information as a Factor in Decision Making

Figure 2 shows the sets of cues presented to the observers for processing and describes how the experimental variable reliability was manipulated.

The cue sets exemplified in Figure 2 were presented one at a time to each subject using a self-paced procedure. The visual presentations, now computer generated, were similar to those shown in Figure 1 but now paper and pencil stimulus and response materials were eliminated. A green on grey video display, 30-cm diagonal measurement, was viewed at a distance of approximately 70 cm. Responses were made on a computer keyboard with one key indicating the observer's decision that the event on the left was more likely to occur and a second key indicating a decision in favor of the event on the right. A press of the RETURN key recorded everything for that trial and presented the next cue set.

Several predictions concerning the use of the reliability information were made prior to the experiment (see Figure 2). The underlying rationale for the predictions is that subjects will react to the manipulation of reliability in a rational manner. That is, they will give more weight to high reliability sources and less to low reliability sources. It is by no means obvious that this will occur, since subjects who are approaching the limit of their ability to process and integrate probabilistic information may simply ignore considerations of relative source reliability.

Prediction 1. The proportion of choices for T-2 should be smaller

T1			T2			T3			T4			T5		
A		B	A		B	A		B	A		B	A		B
	.20			.20			.20			.20			.20	
M	.30		H	.30		L	.30		M	.30		M	.30	
	.40	M		.40	M		.40	M		.40	M		.40	M
	.50			.50			.50			.50			.50	
	.60	M		.60	M		.60	M		.60	M		.60	M
M	.70		M	.70		M	.70		H	.70		L	.70	
M	.80		M	.80		M	.80		M	.80		M	.80	

The above items illustrate one item each from the five treatments with outliers at the low end of the probability scale. Each treatment had 10 different comparisons to be judged, each comparison appearing twice, reversing the event, A or B, with which the outliers were associated.

Outlier arrays had either three or four sources designated. Non-outlier arrays always had two sources. The three-source outlier arrays were 30-70-80, as in the example above. The four-source arrays were 20-50-60-70.

In each treatment, the 30-70-80 array was compared with five non-outlier arrays 40-60, 50-60, 50-70, 60-70, and 60-80. The 20-50-60-70 array was compared with 30-50, 40-50, 40-60, 50-60, and 50-70.

The reliability variable was manipulated across treatments as shown in the examples.

T1: all sources were designated M, medium reliability
T2: the outlier was designated H; highly-reliable, all others M
T3: the outlier was designated L, low reliability, all others, M
T4: the second highest source in the outlier array was designated H, all others M
T5: the second highest source in the outlier array was designated L, all others M

The critical comparisons in this study are the proportions of times the subject chooses the event associated with the outlier array with an H or an L value versus the same outlier arrays with all M values, i.e., the proportion of outlier event choices in T2 with that in T1, T3 with T1, T4 with T1 and T5 with T1. All such comparisons of choices made are against the control set of non-outliers described above, an example of which is shown in the figures as the sources relevant to Event B.

Figure 2. The Plan of Study 2

Figure 2-Continued

T6			T7			T8			T9			T10		
A		B	A		B	A		B	A		B	A		B
	.20			.20			.20			.20			.20	
M	.30	M	M	.30	M	M	.30	M	M	.30	M	M	.30	M
M	.40		M	.40		M	.40		H	.40		L	.40	
M	.50	M	M	.50	M	M	.50	M	M	.50	M	M	.50	M
	.60			.60			.60			.60			.60	
	.70			.70			.70			.70			.70	
M	.80		H	.80		L	.80		M	.80		M	.80	

The above items illustrate one item each from the five treatments with outliers at the high end of the probability scale. Except for the actual values associated with the sources of information, all features of the study are the same as for the outliers at the low end of the probability scale.

The arrays of data associated with the possible events are

OUTLIER ARRAYS

20-30-70 vs.
20-40
30-40
30-50
40-50
40-60

30-40-50-80
30-50
40-50
40-60
50-60
50-70

The reliability variable was manipulated across treatments as follows:

T6: all sources designated M
T7: the outlier was designated H
T8: the outlier was designated L
T9: the second lowest source in the outlier array was designated H
T10: the second lowest source in the outlier array was designated L

than for T-1. (For definitions of treatments T-1 through T-10, see Figure 2.) A highly reliable low outlier would be more heavily weighted (subjectively) thereby decreasing the average of its cue set. This lower average, in comparison with the average of the control set which was composed of all medium reliability cues, should be reflected in fewer choices of the T-2 set in comparison with those of the control T-1 set.

Prediction 2. The proportion of choices for T-3 should be greater than for T-1. A low outlier of low reliability would be less heavily weighted thereby increasing the average of its cue set. This higher average, in comparison with the average of the control set, which, again, was composed of all medium reliability cues, should be reflected in more choices of the T-3 set in comparison with the T-1 set.

Prediction 3. The proportion of choices for T-4 should be greater than for T-1. A non-outlier of high reliability would be more heavily weighted, thereby giving a higher average than that for the control set and more choices of the T-4 set in comparison with the T-1 set.

Prediction 4. The proportion of choices for T-5 should be smaller than for T-1. A non-outlier of low reliability would be less heavily weighted, producing a lower average than that of the control set and fewer choices of the T-5 set compared with the T-1 set.

These four predictions deal with the cue sets containing low outliers. Four predictions relating to cue sets containing high outliers are essentially complementary to those four just listed.

Prediction 5. The proportion of choices for T-7 should be greater than for T-6. A highly reliable high outlier would be more heavily weighted, increasing the average of its cue set in comparison with the average of the control set, which was composed of all medium reliability cues.

Prediction 6. The proportion of choices for T-8 should be smaller than for T-6. A high outlier of low reliability would be less heavily weighted; thus, a lower average.

Prediction 7. The proportion of choices for T-9 should be smaller than for T-6. A non-outlier of high reliability would be more heavily weighted, thereby reducing the average for T-9.

Prediction 8. The proportion of choices for T-10 should be greater than for T-6. A non-outlier of low reliability would be less heavily weighted giving a higher average for T-10.

Complete replication of the 100 experimental trials provided an evaluation of the reliability of each subject's decision strategies. Comparing the first set of 100 trials with the second set of 100 trials (in differing random orders) gave reliability indices within the range from .68 to .99, with a mean of .87 and a standard deviation of .09. Reliability in this context means the proportion of times the same events were selected in the corresponding event-pair configurations in both replications.

Table 4 is a summary of the results of this study. Individual subject data, along with the pooled information, are presented.

From the summary of all subjects shown at the top of Table 4 it can be seen that the direction of difference was confirmed in six of the eight predictions. Of these six, only three of the differences are of sufficient magnitude to be labeled statistically reliable ($p < .05$ for a Type I error). Of the two predictions that were not confirmed, only one of the differences is reliable, that for Prediction 8, and this difference is in the direction opposite to that which was predicted.

Of the total of 8 (per subject) times 19 (subjects) = 152 predic-

Table 4. Overall Summary

LOW OUTLIERS	T-1 BASELINE $\bar{X}$ = 17.53 S.D. = 3.57	T-2 14.05 7.79	T-3 16.32 4.62	T-4 19.26 1.97	T-5 9.21 6.17
HIGH OUTLIERS	T-6 BASELINE $\bar{X}$ = 15.89 S.D.=4.49	T-7 18.58 2.43	T-8 5.05 4.35	T-9 14.58 6.40	T-10 12.53 5.50

INDIVIDUAL PERFORMANCE

	T-1BASE	T-2	T-3	T-4	T-5	T-6BASE	T-7	T-8	T-9	T-10
S-1	20	20	20	20	19	20	20	13	20	19
2	18	16	16	18	7	15	17	3	13	11
3	18	20	10	20	4	18	19	3	20	9
4	8	3	14	12	5	8	10	8	7	6
5	20	20	12	20	1	20	20	2	20	8
6	13	15	20	20	6	13	20	3	15	13
7	20	20	20	20	20	20	20	19	20	20
8	20	20	20	20	11	17	19	7	19	13
9	20	20	12	20	0	20	20	0	20	5
10	19	19	2	20	0	19	18	0	18	0
11	20	20	20	20	16	20	20	3	20	19
12	20	20	17	20	18	20	20	4	20	19
13	20	20	20	20	11	18	19	6	18	11
14	11	4	15	20	5	11	20	3	6	9
15	18	2	20	20	6	15	19	5	6	19
16	12	5	16	16	8	8	15	3	8	12
17	18	0	19	20	10	13	20	4	5	16
18	20	20	20	20	16	20	20	5	20	18
19	18	3	17	20	12	7	17	5	2	11

These tables show the individual subject preferences and the summary for all 19 subjects for the respective treatments described in the text. The maximum number of preferences, choices, for the outlier event was 20.

tions made on an individual subject basis, a majority of predictions in terms of direction of difference (80) were confirmed, with 26 being in the wrong direction and 46 scored as ties. This finding is informative, not so much for the support of the predictions, but in pointing up a shortcoming of the use of these 20 particular event-pair configurations. It had been assumed that use of an averaging strategy by observers would predominate, and it did, but when the averaging strategy was not used, other simplifying strategies raised the baseline of choices to such a high level that a ceiling effect became operative and higher scores became impossible. Thus, the large number of ties. This ceiling effect really precludes any firm conclusions based on these data. Currently, this study is being replicated with a different set of event configurations, selected in light of the data described.

A conceptual analysis of the several meanings of "reliability" and "unreliability" and of the ways in which the construct can be manipulated and measured is needed. In some ways, the perceived reliability of information may well lie at the core of some problems in human inference and information processing. Whether to throw out the data or change one's opinion based on the data ought to be a central issue, both in prediction research and in research in which the data are used as feedback for predictions. This experiment simply did not get at the limited aspect of the larger issue as intended. Other research, related to the whole thrust of this program, does suggest an important hypothesis. It may well be that a crucial aspect of error in the data (i.e., unreliability) is whether the subject perceives it to be error in a measurement error sense (close but not exact) or perceives it to be error in an all-or-none sense. The latter means that the subject perceives the information to be probably exactly correct, but that if it is not exactly correct, it is "garbage." The latter form of perceived error has much more influence on subjects. A similar distinction will be drawn later in the discussion of the distributions of errors in subjects' responses and the impact of those error distributions on the experimenters' interpretation of the data.

V. Study 3. Time Pressure as a Factor in Strategy Use

Another variable considered to be of major importance in assessing decision making behavior is the effect of time pressure, i.e., limited time for information processing.

As compared with the situation in which the decision maker has virtually unlimited time for information processing, what will be the effect of reducing this processing time to some interval below the amount of time normally used? Will strategies change? Will the reliability of the decisions decrease? Will information be discounted (i.e., ignored, or receive reduced consideration) with shorter available processing time?

A new inventory of 102 problems was constructed. This new inventory was similar to the JSH problem set but was made up of special types of items for answering the questions posed. The display and response equipment were the same as in Study 2 with, of course, different software. Subjects were 22 junior and senior biology and psychology students who were paid for their participation. Each subject participated individually in two sessions 1 week apart. In each session, the set of 102 problems was presented in a random order followed by the same set in a different random order. Half of the subjects received the self-paced treatment first, followed a week later by the time pressure session (4 seconds for responding). The other half responded under time pressure first, with the self-paced treatment a week later.

Table 5 shows the mean reliabilities and mean number of non-responses for both groups. As in Study 2, subject reliability, or consistency of subjects' responses, is the proportion of identical responses between first and second presentations of the same stimuli within the same treatment. In the present study though, the proportions are based only on those arrays to which subjects responded on both occasions. In other words, an array which drew a non-response on one presentation but was responded to on the other was not counted.

The numbers of non-responses in the pressure-first treatment is more than 2.5 times the numbers of non-responses when pressure appeared in the second session. This implies that practice in a non-pressure regimen may have familiarized subjects sufficiently with the task to enable them to process information rapidly and to reduce, markedly, the number of non-responses with the restricted response time.

The overall reliability for the self-paced treatment was .73 and for the time pressure treatment, .71. For the group which responded under time pressure first, reliability was .65 under time pressure and .74 when self-paced. For the self-paced-first group, reliability was .73 when self-paced and .77 under time pressure.

What about strategies? All the items in this study were selected such that the arithmetic means were identical for each <u>pair</u> of events to be compared. For example, one event might have cues of .20, .60, and .80 and be paired with an event with cues of .40, .50, .70. Another event pair might have cues of .30, .60, .70, .80 for one event and .40, .50, .70, .80 for the other. Others might be .50, .60, .70 compared with .40, .80; or .20, .30, .70 compared with .40. In other words, an Averaging strategy using all the information available would not discriminate between the likelihoods of the event pairs.

Table 5. Mean Reliabilities and Mean Numbers of Missing Responses

Both Groups (n = 22)	Self Paced	Time Pressure
Mean Reliability	.73	.71
Missing Responses	0	9.30
Time Pressure First (n = 11)		
Mean Reliability	.74	.65
Missing Responses	0	13.50
Self Paced First (n = 11)		
Mean Reliability	.73	.77
Missing Responses	0	5.20

Three strategies were defined in the following ways:

1. In an Outlier strategy, subjects could have discounted an outlier in either or both of the event sets of cues and made a judgment of greater likelihood of occurrence with only the remaining information. This is essentially the complete discounting of any outlier.

2. In a Cluster strategy, subjects showed preferences for cue sets with lower variability; for example, .40, .50, .60 as compared with .20, .50, .80. Only cue sets with three or more cues in each set were used for this analysis.

3. In an Adding strategy, subjects could choose the cue set that had the higher sum. This is perfectly confounded with a Most Cues strategy, when the average is held constant, but since Study 1 showed that a Most Cues strategy was virtually never used, and that an Adding strategy was next most likely to be used (compared to Averaging), it seems more parsimonious to call this an Adding strategy.

Table 6 shows the overall percentages of all responses to those pairs of events amenable to the three strategy analyses that could have been used by the 22 subjects. The table also shows the number of subjects who used these strategies. Strategy use simply implies a response that is consistent with that type of strategy but does not establish that the actual cognitive operations concomitant with that strategy were used.

Use of one or more of the three strategies listed in Table 6 does not preclude use of the other strategies since some inventory items could be evaluated according to more than one of the three strategies. Additionally, subjects could use different strategies at different times -- a mix of strategies. Basically, what the table shows is heaviest "use" of an Adding strategy followed by an Outlier strategy when an

Table 6. Strategy Utilization

a. Mean Percentages of Strategy "Utilization"

Strategy	Self Paced	Time Pressure
Outlier	62%	60%
Cluster	50%	48%
Adding	82%	83%

b. Numbers of Subjects "Using" These Strategies a Significant Amount

Strategy	Self Paced	Time Pressure
Outlier	15	17
Cluster	7	13
Adding	21	20

Averaging strategy is non-diagnostic.

The most interesting result of this study is the facilitating effect of the early, self-paced trials on later time-pressured responses. Apparently, even at the level of formation and use of simple information processing strategies, a considerable amount of time may be required to form the integration rule to be used, and the "performance" aspect of that rule application may take a fair amount of practice.

The Discounting and Cluster strategies, which are similar, did seem to be used by subjects, but the predominant strategy seemed to be just to go with the most votes, or with the biggest sum. When one takes away the strategy of choice of most subjects, i.e., Averaging, they seem to dip into a bag of strategy tricks and to use what seems appropriate to the situation at hand, rather than revert to a second most favored strategy which they then stick with over all problems. The technique of taking away the favored strategy, or rather of precluding its use by making it irrelevant, seems to be a potentially useful one to permit the exploration of the set of strategies subjects actually have available. That set may be a large one, in spite of the quite common use of Averaging as a simplifying strategy.

VI. Study 4. Information Use as a Function of Cue Distribution Variables

The fourth study in this sequence looked at an observer's estimates of overall likelihood without making a decision as to which of two events had a greater probability of occurrence. A single linear display of probabilities was shown on each trial, and the subject simply indicated a "likelihood of occurrence" by moving a cursor above the display to a point that indicated this likelihood. Figure 3 shows this type of display with some sample problems.

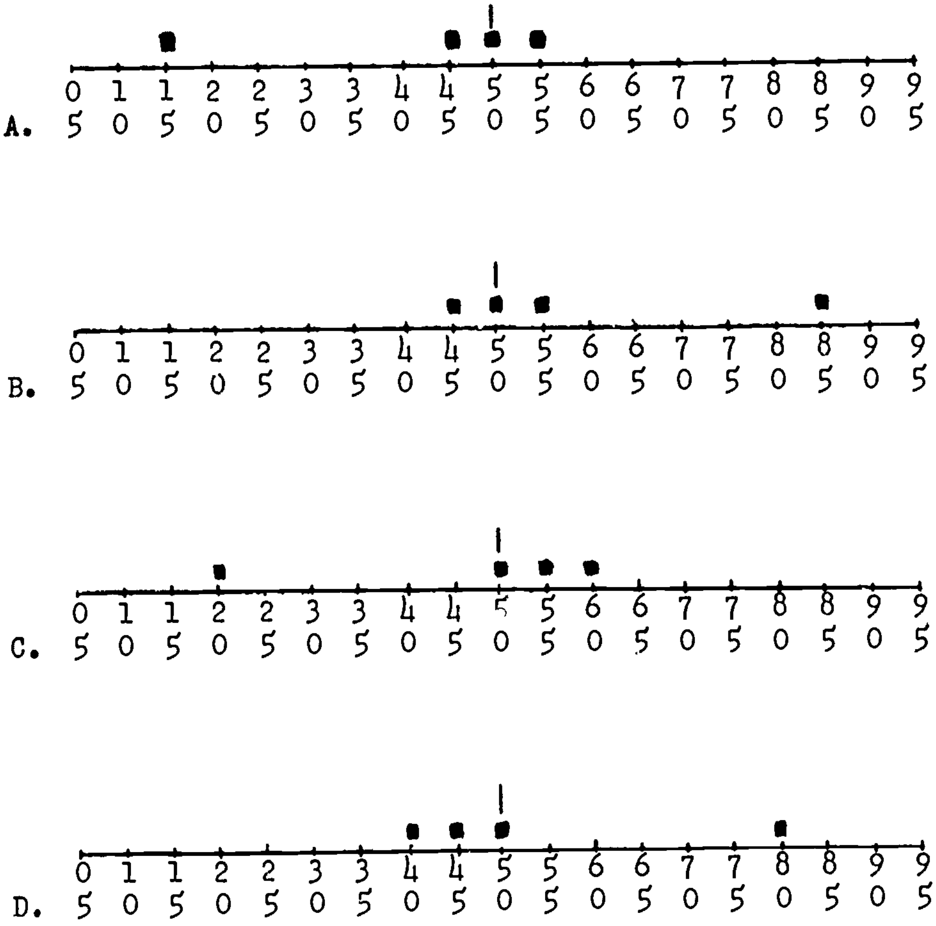

Figure 3. Four sample problems. Problems A and C have low outliers; problems B and D have high outliers. Problems A and B have symmetric clusters; problems C and D have assymetric clusters. All four problems have an outlier distance of .3 and a density of 3.

This task is an abstract analogue of a real-world situation. A person who has several sources of information bearing on the possible occurrence of a single future event is asked to estimate the overall probability of the occurrence of that event. As in the previous studies, the information is presented in the form of probabilities.

Four variables were manipulated systematically to analyze not only their individual effects on this task but their possible combined, or joint, effect. One variable was the probabilistic distance (either .20, .30, or .40) between an outlying bit of information and the rest of the clustered information. The second variable was the direction from the cluster in which the outlying information was to be found (either above or below, i.e., higher than or lower than the cluster on the probability scale). Variable three was the density of the clustered information -- either two, three, or four pieces of information located within the same short range. Variable four, symmetry, placed the clustered information (not the outlier) either directly in the middle of the display or slightly to the left (lower) or right (higher) part of the display.

Subjects were run individually and were self-paced. Each subject received five presentations of each of the 36 stimulus configurations: (3 levels for variable 1) x (2 levels of variable 2) x (3 levels of variable 3) x (2 levels of variable 4) = 36.

Of the 20 subjects, two did not understand the task and were eliminated from any analysis. Of the remaining 18, two more were set aside because of an unusually strong tendency to use only the probability of the highest cue value. Inclusion of the data of these latter two subjects changes no conclusions, nor affects any significance tests. The results for the remaining 16 subjects are summarized in Table 7.

Three of the four variables showed statistically reliable effects.

Table 7. Marginal Means of Directional Deviation Scores for Reduced Sample (n = 16)

Density	Outlier Distance	Outlier Direction	Symmetry
2	.2	Low	Asymmetric
$\bar{X}$ = -.01051	$\bar{X}$ = .00353	$\bar{X}$ = -.00514	$\bar{X}$ = .00136
σ = .02854	σ = .02487	σ = .03673	σ = .03497
3	.3	High	Symmetric
$\bar{X}$ = -.00033	$\bar{X}$ = -.00363	$\bar{X}$ = -.00116	$\bar{X}$ = -.00765
σ = .03607	σ = .03495	σ = .03575	σ = .03703
4	.4		
$\bar{X}$ = .00140	$\bar{X}$ = -.00934		
σ = .04196	σ = .04512		

First, as the outlier distance increased from .20 to .30 to .40 units away from the nearest other datum, the outlier's effect shifted from reducing the estimated likelihood for low outliers and increasing the estimated likelihood for high outliers, respectively, to a discounting effect which increased the estimated likelihood for low outliers and decreased the estimated likelihood for high outliers. Second, as the density of the non-outlying cues changed from two to three to four, the outlier had an increased effect in terms of changing the estimated likelihood of the to-be-predicted event in the direction of the outlier, away from the cluster. In other words, it appeared as though the increased density of the non-outliers, the cluster, became more pronounced, and the observers weighted the outlier more heavily in estimating overall likelihoods. This effect was contrary to what had been expected from the verbal reports of many subjects in earlier studies. Third, symmetric arrangements of displays, i.e., clusters centered about .50, produced average likelihood estimates toward the extreme end of the cluster (away from the outlier) as compared with an arithmetically calculated average. Asymmetric arrangements gave mean likelihood estimates away from the arithmetic average in the direction of the outlier. Direction of the outlier from the clustered information showed no statistically reliable effect on the absolute difference scores, nor did the interactions of any of the set of four independent variables.

In this study, a clear outlier discounting effect emerged, but the outlier must be relatively extreme before the discounting occurs. Since, in Study 3, outliers were defined as being .30 unit away from the nearest datum favoring the same event, the use of the outlier Discounting strategy in Study 3 may have been attenuated. Discounting is probably, based on the data of Study 4, a potent phenomenon, provided that the outlying datum is actually extreme.

The failure to find a directionality effect is somewhat surprising, given the general finding that negative information is more salient than positive information. It is not certain, though, that all subjects properly conceived low probabilities as negative information, or as information that was to be taken as evidence against the occurrence of the event.

The density effect is paradoxical. Outlying data are discounted because these sources are too different from other more coherent sources. This implies that as data in the clusters get tighter, subjects should perceive such data as more reliable. They do not. Any explanation at this juncture would be purely ad hoc, so it will be left as a paradox.

VII. Study 5. Presentation Mode and Allocated Processing Time as Factors in Estimating Numerical Averages

How information is displayed to an observer is the last major variable to be evaluated in this series. In all studies so far, the information presented has been on a geometric numeric (GN) scale. That is, probabilities have been indicated on a scale in which equal distances between equally different probabilities were reproduced geometrically (see Figure 4). Another scale, or list, which shows only the values of the probabilities in list form (LF) without geometrical representation is also presented in Figure 4. A third representation is made up of the familiar histogram bars (HB) where the heights (and areas) of the bars present the information.

The question now is: If the quantitative values of the information are the same in all presentation modes, will responses also be the same? If not, how will they differ?

```
Histogram Bars       9       9
                     8       8
                     7 *     7
                     6 *     6
                     5 *     5
                     4 * *   4
                     3 * *   3
                     2 * * * 2
                     1 * * * 1

Geometric Numeric    9
                     8
                     7 *
                     6
                     5
                     4 *
                     3
                     2 *
                     1

List Form            7
                     4
                     2
```

Figure 4. An example of information in the Histogram Bars (HB), Geometric Numeric (GN), and List Form (LF) displays.

The display formats were presented to subjects with three different times for processing the information and for two different amounts of information.

This study used 15 undergraduate subjects all of whom made judgments (a) of cue sets composed of three or five sources of information, (b) with presentation times of 3, 6, or 9 seconds, and (c) with all three types of displays, GN, LF, HB. The judgment to be made was a simple average of the three or five cue values shown on the video screen. The response was written on paper by the subject during a 4-second intertrial interval.

Figure 5 gives a summary of the results of this study in several different ways for ease in assessing the effects of the three independent variables and their interactions. In Figure 5 note that the measures are the averages of the absolute values of the differences between the arithmetically correct average and the subjects' estimated averages, i.e., the average absolute error per trial.

The three top graphs show the main effects of the three independent variables. The different formats (HB, GN, LF) all give about the same average error overall, with the HB displays showing the highest processing error (5.91), the GN displays showing the least error (4.92), and the LF in between (5.07). Although the differences are small, they are statistically reliable.

The different display times show superior accuracy for the 9-second processing time, followed by the 6-second time, with the 3-second time allocated to processing the least accurate. These average errors are 3.86, 4.69, and 7.35, respectively. These differences are also reliable.

Finally, as expected, the error for five cues is reliably larger

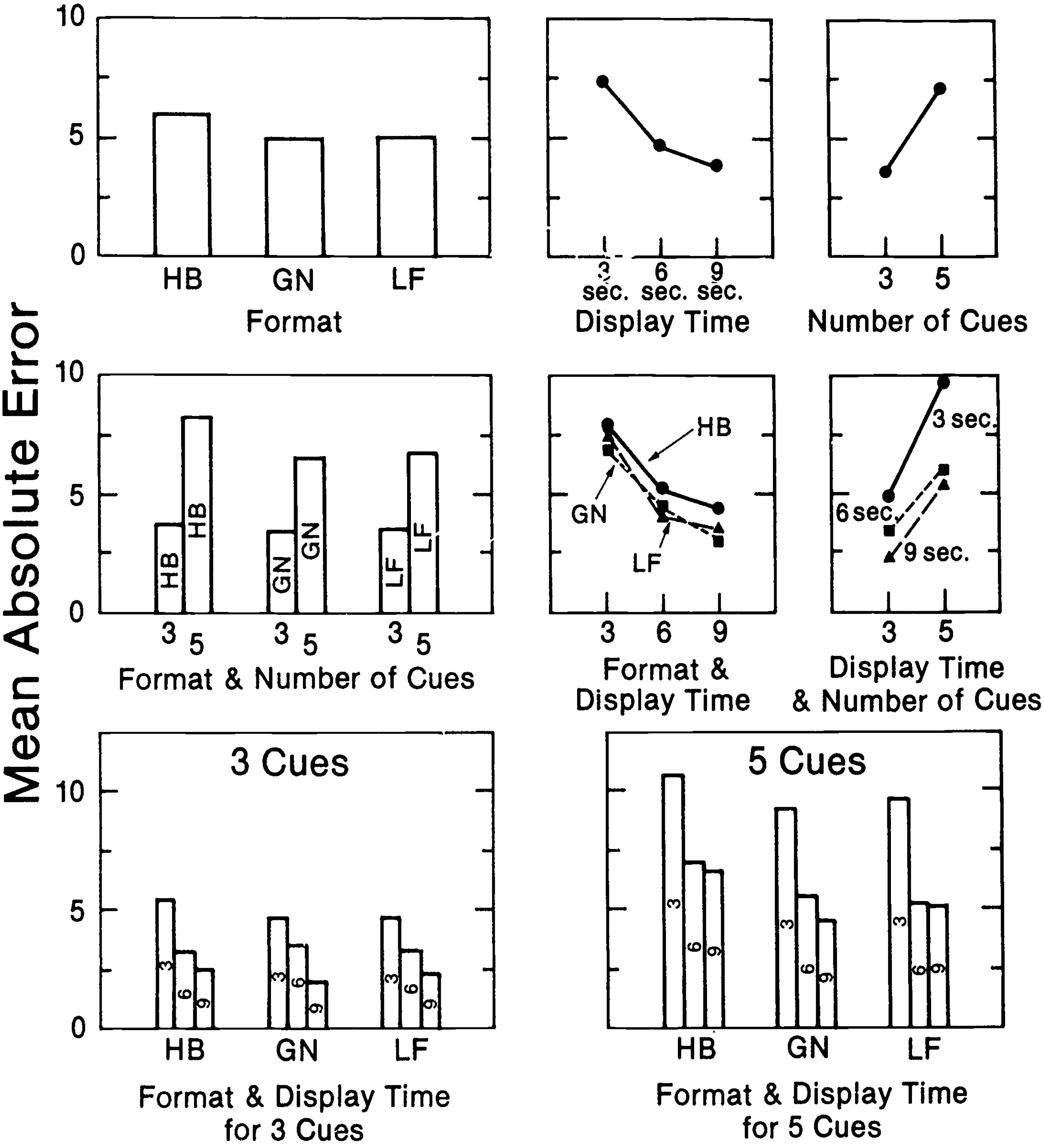

Figure 5. Row 1 shows the significant effects for each of the three main factors. Row 2 shows the three non-statistically significant two-factor interactions. Row 3 shows the two graphs which together allow interpretation of the non-statistically significant three-factor interaction.

than the error for three cues: 7.10 compared with 3.50.

Although the results of this study are quite orderly, there is an additional issue to be considered. Of the 15 subjects in the analyses, four did best with histograms, five with geometric numeric information, and six with the simple numeric list.

The differences for processing time and amount of information serve primarily to show that the results are, in fact, orderly. The relative superiority of the GN over the HB looks, at first glance, like a stimulus response compatibility effect, since both stimuli and responses are in numeric form, except that the GN display is also superior to the LF. It seems that even with a task as simple as averaging three or five digits, the spatial representation of the metric relations carried in the GN display enables better performance. It may be that the geometric displays facilitated some sort of error-checking routine, or made some sort of intuitive (rather than analytical or algorithmic) processing more likely. In either event, different error distributions would be expected in the GN than in the LF displays, a hypothesis that has not yet been assessed. In fact, error distributions may eventually be of considerable interest. Two displays could have the same average error, with one having many small errors but very few exactly correct responses, and with another having very few errors, but large ones. This difference in error distributions is precisely what Brunswik predicted and found when he contrasted intuitive and analytical thought, or perception and reasoning (Hammond, 1966). A difference of this sort could, of course, have profound implications for many activities of concern to the Air Force. The rare but very large judgmental error is almost certainly a far more serious problem than is the frequent but small error.

The apparently (but not actually) anomalous degree of individual differences reflected in the study makes any generalizations from these data tenuous. While the small number of subjects and the within design (with the consequent difficulty of interpretation) preclude unqualified generalizations, the superiority of the GN display is most interesting and very likely is not situation specific. Essentially, the GN display seems to embody the stimulus characteristics that elicit the best features of both intuitive and analytical modes of thought.

VIII. Study 6. Two Additional Investigations

The next two experiments were conducted in addition to the original effort. Both studies further examined the effects of unreliability in a context different from those already described. These experiments used a paradigm developed by Doherty, Mynatt, Tweney, and Schiavo (1979) which showed that people tend to seek and use diagnostically worthless information when diagnostically valuable information is easily available. This experimental paradigm is called pseudodiagnosticity.

IX. The Bayesian and Pseudodiagnosticity Paradigms

Assume someone has one of two -- and only two -- possible diseases. Assume that a probability can be assigned to disease A, P(A), and a complementary probability <u>to</u> disease B, P(B), where P(B) = 1 - P(A). These probabilities are called prior probabilities, or simply, priors.

Assume now, subsequent to the assignment of these priors, two symptoms appear, X and Y, both of which have a known relationship to each of the diseases A and B. All this information is shown in Table 8.

An example of the information described in Table 8 might be the following illustration.

Table 8. Probabilities Utilized in Simplified Paradigm

	Disease A $P(A)$	Disease B $P(B)$
Symptom X	$P(X/A)$ $P(\overline{X}/A)$	$P(X/B)$ $P(\overline{X}/B)$
Symptom Y	$P(Y/A)$ $P(\overline{Y}/A)$	$P(Y/B)$ $P(\overline{Y}/B)$

where

$P(A)$	is the prior probability of disease A
$P(B)$	is the prior probability of disease B
$P(X/A)$	is the probability of symptom X occurring given that disease A is present
$P(\overline{X}/A)$	is the probability that symptom X does not occur given that disease A is present
$P(Y/A)$	is the probability of symptom Y occurring given that disease A is present
$P(\overline{Y}/A)$	is the probability that symptom Y does not occur given that disease A is present
$P(X/B)$	is the probability of symptom X occurring given that disease B is present
$P(\overline{X}/B)$	is the probability that symptom X does not occur given that disease B is present
$P(Y/B)$	is the probability of symptom Y occurring given that disease B is present
$P(\overline{Y}/B)$	is the probability that symptom Y does not occur given that disease B is present

		Disease A P(A) = .50	Disease B P(B) = .50
Row 1	Small Body Rash	.65	.25
	Absence of Rash	.35	.75
Row 2	Elevated Heart Rate	.80	.40
	Absence of Elevated Heart Rate	.20	.60

Now, suppose that a diagnostician can request the information to be found in any two of the cells shown in the illustration for purposes of correctly diagnosing the disease. From which two cells should this information be chosen to enhance the chances of a correct diagnosis?

The correct answer is a straightforward application of Bayes' Theorem. Either the two cells in row 1 or the two cells in row 2 should be selected since corresponding information for both diseases must be obtained. Any other pair of cells provide worthless information.

The appropriate calculations for the selection of row 1 information are

$$P\ (A/Rash) = \frac{P(A)P(Rash/A)}{P(Rash/A)P(A) + P(Rash/B)P(B)}$$

$$= \frac{(.50)\ (.65)}{(.65)(.50) + (.25)(.50)}$$

$$= .72$$

and

$$P(B/Rash) = .28$$

For row 2

$$P(A/Heart) = \frac{P(A)P(Heart/A)}{P(Heart/A)P(A) + P(Heart/B)P(B)}$$

$$= \frac{(.50)\ (.80)}{(.80)(.50) + (.40)(.50)}$$

$$= .67$$

and

$$P(B/Heart) = .33$$

What do people do when confronted with problems like this in laboratory situations? Several published and unpublished studies have shown that undergraduates, graduate students in business administration, and medical residents, among others, ask for the wrong data most of the time. Medical residents tend to want "confirmatory" information and call for the data in columns. University undergraduates generally prefer information contained in a diagonal.

Do people learn to ask for the correct information? Doherty, Schiavo, Mynatt, and Tweney (1981) using similar problems had subjects select data and make decisions as in the disease problem described. These subjects then were given feedback as to whether their decisions were consistent (correct or incorrect) when compared with the probabilistic model. Additionally, half of the subjects were then given a third cell of the information matrix to guarantee that they would see a properly diagnostic pair of cells. These subjects were again asked to do several more problems.

The results were clear. When subjects selected the wrong data but made the right choice, they continued to ask for the wrong data. When subjects selected the wrong data, made the wrong choice, and were not given the third bit of information, they also continued to ask for the wrong data. Only when subjects selected the wrong data, made the wrong choice, and were given the third unit of information did they shift to the optimal data selection strategy.

In the study just described feedback as to the correctness or incorrectness of the diagnostic decision was always perfectly reliable. That is, if the arithmetically calculated correct probability of choice A was greater than the probability of choice B, the subject was always told that A was the correct choice. Even though this procedure guarantees maximum diagnostic performance, the real world is not structured that way. People often make the best choice possible but find out that they were wrong. How often? In the Bayesian framework, the proportion of times is (1 - P), where P is the probability of an event after the new information has been taken into account.

One of the focal questions of concern in this effort was the impact of unreliability. Specifically, the issue of unreliability in feedback was investigated using the pseudodiagnosticity paradigm. Essentially, this involved providing feedback according to the actual probabilities that a diagnosis would be correct or incorrect. Thus, in one treatment, the subjects were provided feedback that was fairly typical of psychological laboratories and class demonstrations: i.e., the feedback was always of the right sort. If the subjects made the optimal choice of data, they were virtually certain to make the best (most probable) decision, and then the artificial environment would tell them they were right. The other treatment provided feedback much more like that occurring in the real world. That is, the feedback itself had the character of being uncertain, of being predictable only in a probabilistic sense.

X. Study 6. Reliable and Unreliable Feedback

All subjects, 29 students in a nursing program taking a course in statistics, were run individually using the video display and computer keyboard described earlier. Subjects were randomly assigned to one of two experimental treatments. One treatment provided feedback to subjects after their responses according to the rule that says if the probability of the chosen disease given the two symptoms is greater than .50, $P(D/S_1, S_2) > .50$, then the response is called correct 100% of the time. This is feedback F_r.

The second treatment provided feedback after responses according to a rule which says if the probability of the chosen disease given the two symptoms is $P(D/S_1, S_2)$ -- whatever its value -- then the response is called correct $P(D/S_1, S_2)$ proportion of the time. In other words, the feedback for correctness or incorrectness of the choice was randomly determined according to the arithmetically correct proportions of times disease A and disease B occurred. Thus, sometimes "correct" (i.e., most probable) diagnoses would be labeled incorrect, and sometimes incorrect diagnoses would be labeled correct. This is feedback F_u.

In both treatments all subjects saw the same displays and performed the same 40 diagnostic tasks, but 15 subjects received one kind of feedback and 14 the other. The displays were like those already described in the 2 x 2 matrix except that all information was electronically masked at the outset. The computer then randomly selected one of the four cells for the pertinent diagnostic information and presented it on the video display. A second one of the four cells was then selected by the subject and this cell's information was also shown on the display.

With these two probabilities, the subject made the diagnostic response, either disease A or disease B. Immediately thereafter, the display informed the subject of the correctness of the response according to scheduled feedback F_r or F_u and presented the entire 2 x 2 array of diagnostic information for the subject to inspect. A new trial with new information was initiated at the subject's discretion by depressing the RETURN key.

What were the results? The great majority of subjects never adopted the statistically appropriate strategy. Only three subjects in the F_r treatment <u>switched</u> to the appropriate strategy. Three <u>started</u> with the appropriate strategy and never deviated from it. In the

F_u treatment, no one started with the appropriate strategy, three adopted it and used it on at least seven consecutive trials, and one abandoned it after having a rational choice identified as "wrong." The small number of people adopting the appropriate strategy in the F_r treatment deprived this study of an adequate baseline against which to assess the possibly detrimental effects of unreliability in the feedback. There was an overall positive effect of the independent variable, F_r vs. F_u, i.e., in the F_r treatment, 44% of the choices were the appropriate diagnostic type compared with 33% of the choices in the F_u treatment.

Why didn't more people in the F_r treatment adopt or switch to the appropriate strategy? Probably because there was still a fairly high proportion of responses that received feedback as "correct" even when the strategy in arriving at the response was incorrect. As expected, when people are reinforced, they keep doing what they have been doing.

XI. Study 7. Reliable and Unreliable Feedback II

With modified conditional probabilities relating symptoms to diseases, 13 more subjects from a statistics course participated in an experiment methodologically identical to Study 6. Again, subjects in the F_r treatment did not really learn to adopt the appropriate strategy for information acquisition. Nonetheless, it is important to note that the results of these two experiments can be interpreted in a very direct fashion.

The two experiments are strong replications of the Doherty et al. (1979, 1981) findings and show the power of the pseudodiagnosticity effect. Subjects with correct data available at the push of a button are more likely to select and to use incorrect data when making a diagnostic decision. This is even more surprising considering that subjects in these two studies were in a highly selective nursing program, had a substantial introduction to statistical thinking, and worked through 40 experimental trials. Furthermore, on each trial, after making a response and obtaining feedback, the subjects then saw all of the potentially available information. They simply did not bring the necessary cognitive operations to bear on the problem.

The study of the effect of error in the feedback, in a variety of task environments, is potentially extremely important. It is a universal human tendency to want performance feedback. But if performance feedback of various sorts can be shown to be disruptive when that feedback is sufficiently laced with error, then such feedback probably ought to be withheld, no matter what the learner wants, at least early in the learning process. The multiple cue probability learning literature provides clear examples of situations in which feedback disrupts both learning and performance. Given the widespread belief that feedback is always a "good thing," given the clear power of feedback from both reinforcement and informational standpoints, and given what is assumed to be an uncontrovertible fact that feedback in the world is itself strong in uncertainty, investigation of the effects of error in feedback seems critical.

The authors plan to pursue the problem with minor modifications of the procedure used here. The essential modification to be made, interestingly, is a change in the direction of greater representativeness. That is, if a disease -- or any other type of system failure -- does have more than one symptom, then those symptoms should not be independent of one another in the real world. In the jargon of probability theory they should be "conditionally dependent." In these two studies, all P(S/D) values were made to be statistically independent.

Introducing precisely the dependence that probably exists in the world should have the effect of making erroneous strategies of symptom choice less serendipitously informative. This should permit development of a baseline of good performance in an F_r treatment so it can be observed whether F_u has the disruptive effect predicted.

XII. Discussion

First, a novel framework was used in which decision makers compared evidence for the likelihood of one event with the evidence available for the likelihood of another event, then chose which of the two events was more likely to occur. The evidence for each event was in the form of probabilities and was diagnostic (useful for predictive purposes) for the occurrence of that, and only that, event. Then a comparison was made of the use of some simple types of strategies among UPT pilots and F-16 trainees. Few differences were found among these pilot types but strong indications for the predominant use of a type of Averaging strategy -- a result that has been obtained in other multiple cue probability processing situations.

The inference was that most pilots and pilot trainees averaged information systematically with the varying numbers of cues tested. Nevertheless, there was a small number of subjects whose behavior in this primitive, and perhaps fundamental, cognitive task was completely inconsistent with averaging. Some subjects clearly used one or another of the alternative strategies that had been hypothesized. From these data alone, it cannot be determined whether these unusual strategies for processing information would generalize across other tasks since constraints on the pilots' time did not permit using these same pilots in different tasks. If future researh shows these differences to be consistent across information processing tasks, then there may be situations in which it would be desirable to pre-select individuals and to compose groups for homogeneity of cognitive processes. Equally important, there may be situations in which heterogeneity of thinking style would be highly desirable.

In a series of laboratory studies with college students, two variables considered to be important in real-life situations were examined: (a) the reliabilities of the information sources (evidence) used to predict the occurrences of events, and (b) the presence of disparate information sources (outliers). It was found that the reliabilities of the sources do have an effect of the choice of an event inferred to be more likely to occur, but the results are not always consistent with what would be expected. Additional research is investigating this effect further. Outliers were found to have effects both in making decisions and in estimating averages of probabilities, and other variables were identified that are important in considering these effects. These other variables included the overall symmetry of the display of information, the magnitude of difference between the outlier and the cluster, and the size of the clustered information.

This effort examined the effects of allocated time for processing information in two studies. In one study, performance with this reduced time, 4 seconds, was compared with performance with essentially unlimited time. Reliability of respondents' decisions changed, as did certain types of strategy use. In the second study, three allocated processing times showed greater accuracy in estimating average probabilities of occurrence for 9, 6, and 3 seconds, in decreasing order. This same study also examined the effects of three different types of displays: a bar graph of probabilities, a scaled list of probabilities with appropriate interval spacing, and a simple list of probabilities consisting only of numerical information. Differences in accuracy using the three types of displays were small but responses for bar graphs showed less accuracy than did the lists. Among the subjects, however, an almost equal number of people performed best with each of the three types of displays.

Two studies examined the way information is selected for making decisions in a pseudodiagnosticity paradigm. With little or no training, or even with a fair amount or practice, people consistently ask for information that has little or no value from the standpoint of making rationally correct decisions.

References

Anderson, N. H. Foundations of information integration theory. New York: Academic Press, 1981.

Doherty, M. E., Mynatt, C. R., Tweney, R. D., & Schiavo, M. Pseudodiagnosticity, Acta Psychologica, 1979, 43, 111-121.

Doherty, M. E., Schiavo, M., Mynatt, C. R., & Tweney, R. D. The influence of feedback and diagnostic data on pseudodiagnosticity. Bulletin of the Psychonomic Society, 1981, 18, 191-194.

Hammond, K. R. The psychology of Egon Brunswik. New York: Holt, Rinehart & Winston, 1966.

Jones, D. P., Schipper, L. M., & Holzworth, R. J. Effects of amount of information on decision strategies, The Journal of General Psychology, 1978, 98, 281-294.

Peterson, C. R., & Beach, L. R. Man as an intuitive statistician, Psychological Bulletin, 1967, 68, 29-46.