This is to certify that the

thesis entitled

THE EFFECTS OF INFORMATION UNRELIABILITY
ON JUDGMENT AND DECISION MAKING PROCESSES

presented by

William Michael Rogers

has been accepted towards fulfillment
of the requirements for

MA degree in Psychology

Neal Schmitt
Major professor

Date 5/2/96

THE EFFECTS OF INFORMATION UNRELIABILITY ON JUDGMENT AND DECISION MAKING PROCESSES

By

William Michael Rogers

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF ARTS

Department of Psychology

1996

ABSTRACT

THE EFFECTS OF INFORMATION UNRELIABILITY ON JUDGMENT AND DECISION MAKING PROCESSES

By

William Michael Rogers

The purpose of this study was to investigate the effects of psychometric unreliability on judgment and decision making processes. Literature is reviewed on the effects of unreliability on decision outcomes and the related research on source credibility. Structural and process models of decision-making are discussed in terms of their relative advantages and disadvantages. Brunswik's lens model and the contingency model of decision making (Beach & Mitchell, 1978) are used as conceptual frameworks to discuss potential effects of information unreliability as well as generate relevant predictions for research. A combined research design using elements of both multiple cue probability learning and process tracing methodologies was conducted. Results suggest that unreliability does have effects on judgment accuracy and selection of decision making strategies.

ACKNOWLEDGMENTS

There are several individuals whose contributions to this thesis are noteworthy. First, I would like to thank the faculty at Michigan State University, especially members of my thesis committee. Neal Schmitt, my committee chair, led me expediently through a process, in which, if left to my own devices, I would still be working. His pragmatic advice often guided me out of motivational cul-de-sac's of my own creation. Rick DeShon provided useful recommendations throughout the entire thesis process, especially for designing the pilot studies and analyzing a complex experimental design. Steve Kozlowski was instrumental in keeping me focused on theoretical issues, and provided important cautionary suggestions during the project's development.

I would also like to thank some individuals who were important to conducting the research. Beverly Bockes was very helpful in scheduling computer lab space and assisting with several equipment problems throughout the study. My undergraduate assistants, Michelle Parobek and Amy Sperti, were of vital importance in coding the data. Without their hard work and willingness to tediously rate several pages of dull decision matrices at a moment's notice, this thesis would never have been completed on schedule.

Finally, I would like to thank my family, especially my parents, for believing in me and providing the constant support and encouragement necessary for me to believe in myself.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Statement of Problem

The majority of decisions made outside the confines of a psychological laboratory involve potentially unreliable information. Decisions are often made utilizing indicators which are subject to environmental constraints or measurement imperfections. A stock broker wishing to gain any degree of professional success cannot make buying and selling decisions based on a set of static values. The broker knows that the values he or she is working with compose only a "snapshot" of the overall machinations of the market, and may be dramatically different from week to week. An airline pilot must exercise caution when interpreting the value of an altimeter when the altimeter is subject to vibration by atmospheric turbulence. The pilot should be aware that the position of the altimeter at any given time is due both to actual altitude and physical vibration of the needle in the meter. It is incumbent upon the pilot to make an estimate of actual altitude on which to base further operation of the aircraft. A psychometrician knows he or she must use and interpret tests with low test-retest reliability with care, as the scores on such instruments are difficult to predict from one occasion to the next. Poor reliability in the psychometrician's world not only reduces predictability of test results, but also directly harms psychometric validity, further reducing the utility of a particular test.

The majority of decision making studies do not address the effects of unreliable information, and thus, implicitly make an assumption regarding the decision making context, namely that "what is, was, and will always be". Frisch (1988) notes:

"*...effect of ambiguity on decisions highlights an important limitation of decision-making models, namely that the optimal decision will always be conditional upon the information one has available at the time.*"

All information is implicitly assumed to be temporally constant, and the consideration of information and accuracy of the decision are equal at all time periods. As previously noted, the real world is not this cooperative, forcing scientists to develop methods of addressing the non-constancy of data over time.

The previous discussion and examples reflect potential effects of unreliability in social, physical, and natural systems. Clearly, information unreliability is potentially an important means by which a decision environment can affect the decision maker. This study attempted to identify the effects of information unreliability on judgment and decision making processes.

First, the construct of psychometric reliability is discussed and generalizations / limitations of the research are noted.

Second, differences and relative advantages of structural and process models of judgment/decision making are discussed, and rationale given for use of a combined approach in studying information unreliability. Brunswik's lens model and Beach and Mitchell's (1978) contingency model of decision making are discussed and used as the framework within which the problem of unreliability is discussed.

Third, relevant research from the source credibility literature is reviewed and applied to the context of this study.

Fourth, methodological issues relevant to structural and process components of the study are addressed. A modified MCPL model is presented which allows the assessment of effects of unreliability on judgment. Also, decision making strategies emerging from process tracing research are outlined. Various process tracing methodologies are discussed, and rationale given for the method utilized in this research. In addition, the dependent measures associated with strategy identification are also discussed and operationalized.

Fifth, hypotheses are developed based on the Beach and Mitchell (1978) model, the previous work of York, Doherty, & Kamouri (1987), and findings from the source credibility literature.

Finally, these hypotheses are statistically tested and results presented in tabular format. A discussion section offers potential reasons behind the findings, as well as limitations of the study and directions for future research.

The Reliability Construct

Many conceptualizations and operationalizations of information unreliability do not view reliability in precise accordance with the psychometric definition. Psychometricians view reliability in a very specific sense, conceptually as the proportion of variance that is true variance (Winer, 1971) and operationally as consistency of observations upon repeated measurements of the same thing (York, Doherty, & Kamouri, 1987); that is, test-retest reliability. After true scores are removed from any measurement, the remaining variance is termed *measurement error*, which is assumed to have a mean of zero. The "vibrating altimeter" and testing examples presented earlier can be thought of as examples of the effects of measurement error in mechanical and natural systems, respectively. Reliability, for the purposes of this study, was operationalized in a manner consistent with the psychometric definition. Such a definition was useful for two reasons: 1) It facilitated generalization of the study's findings to situations where psychometric reliability is the contextually proper definition (e.g. personnel decisions involving information from several tests varying in test-retest reliability); 2) It allowed generalization to situations where true values and errors of measurement are appropriate concepts (e.g. the altimeter example). Using this definition tends to prohibit the generalization of results to situations where psychometric reliability is not the appropriate concept, mathematically (e.g. fluctuating true scores, non-random disturbance terms) or conceptually (e.g. the "reliability" of a co-worker or employee).

It is important to note the distinction between judgment and decision making, as well as the proper conceptual frameworks relevant to each. Structural models and process models are discussed next, with a focus on differences in aims of each framework, and the relative advantages and disadvantages of each method in assessing judgment and decision making involving unreliable information.

Process Models vs. Structural Models

This study used conceptual frameworks from both *process* and *structural* model domains. A process model, Beach and Mitchell's (1978) contingency model of decision making, and a structural model, Brunswik's lens model, were used to consider the effects of unreliability on decision making and judgment processes, respectively.

Process models of decision making have been utilized by researchers in order to gain an understanding of the information processes and strategies at work when an individual is making a decision. The use of process tracing methodology in decision making research was initiated by Payne (1976), who modified process tracing procedures developed by researchers in human problem solving (Newell & Simon, 1972). At the time, the primary decision making paradigms were structural models, such as the lens model, which placed primary emphasis on decision outcomes and their relationship with information stimuli. In these models, inferences regarding underlying decision processes are made by examining statistical models of the information-outcome relationship. Many criticisms have been raised regarding the use of structural modeling to assess decision processes, primarily questioning the fitting of a linear model to decision making (Slovic, Fischhoff, & Lichtenstein, 1977) and the fact that many different cognitive processes can be represented by statistically identical structural models (Anderson, 1969).

Process models offer advantages over structural models (such as MCPL) in explaining decision making processes. Process models focus not on the decision itself, but on the steps intervening between information input and decisional output. (Payne, 1976; Svenson, 1979). Data is collected during the decision process in order to make inferences regarding the algorithms or strategies that individuals utilize to make a decision (Ford et al, 1989). Structural models, such as Brunswik's lens model, are more appropriate for assessing judgment policies and utilization of information.

The two conceptual models used in this study are now outlined, with a focus on how each is appropriate for assessing decision making and judgment.

Decision-Making Strategy: A Contingency Model

Ford, Schmitt, Schechtman, Hults, & Doherty (1989) suggest that researchers using process methodologies might facilitate interpretation of results by linking their studies with existing psychological theory. Specifically, they suggest the use of the cost-benefit model of strategy selection presented by Beach & Mitchell (1978). The Beach & Mitchell contingency model of decision making, a process model, posits strategy selection as primarily dependent on task, environment, and personal characteristics, all of which have effects on costs and benefits associated with the decision. The model suggests that increasing cognitive effort due to the task, environment, or person, causes the decision maker to select strategies which minimally increase cognitive demands. In other words, these strategies seek to reduce the cognitive complexity of the decision by involving a minimum of information to reach a decision. These strategies are non-compensatory and nonanalytic in nature, and according to Beach & Mitchell, generally have lower probabilities of generating a correct solution. Analytic strategies (compensatory), while having much higher probabilities of success, also increase cognitive demands to a greater extent. These strategies generally involve the usage of more information and the application of mathematical operations, such as averaging and summations, prior to reaching a decision.

While many process studies are not explicitly framed within the Beach and Mitchell contingency model, findings are generally consistent with the model. The Beach and Mitchell model categorizes factors influencing decision making strategy into three groups: *person* factors, such as ability, knowledge, and motivation; *task* factors, such as ambiguity, unfamiliarity, and complexity; and *environment* factors, such as time constraints, accountability for decision, significance and irreversibility of the decision. As noted by Ford et al., the majority of process tracing work has been done in the *task* factor category, though some work does address *environment* and *person* factors. Process tracing methods have been used to study the effects of task complexity, defined by the number of dimensions, alternatives, or both dimensions and alternatives (Payne, 1976; Payne & Braunstein, 1978; Onken, Hastie, & Revelle, 1985; Johnson & Meyer, 1984), information redundancy (Gilliland & Schmitt, 1993), and a variety of other factors. Findings regarding task complexity have been very robust. Increased task complexity has been shown to result in decreased proportional information search (Payne, 1976; Payne & Braunstein, 1978; Onken et al, 1985), increased search variability (Biggs et al, 1985; Payne & Braunstein, 1978), and decreased mean search time (Payne & Braunstein, 1977). Gilliland & Schmitt (1993), in a study addressing the effects of information redundancy (intercorrelations between dimensions), found decreased search depths in redundant dimensions, suggesting that, in addition to the amount of information present in the task, the structure of information may play a role in perceived complexity of the task.

Gilliland & Schmitt (1993) concluded that importance is not the only factor considered when selecting information, and speculate that many other factors may play a role in determining which information is selected. Among these factors, Gilliland & Schmitt (1993) note the possible role of information unreliability.

Having elaborated a framework within which to study decision making processes and contingencies, the earlier discussion of information unreliability can be reappraised. The Beach and Mitchell contingency model elaborated above explicitly addresses the role of unreliability in perceived costs (demands) associated with a decision. Beach and Mitchell construe unreliability to be part of a task-related factor called *ambiguity*, defined as "*the degree to which the problem is unclear to the decision maker. This includes the ambiguity of the goals, decision alternatives, constraints, etc. as well as the unavailability, unreliability, and imprecision of relevant information.*" (Beach & Mitchell, 1978).

The Multiple Cue Probability Learning Paradigm

Researchers utilizing the psychometric treatment of reliability have assessed its effects using the multiple cue probability learning (MCPL) paradigm. The MCPL paradigm is a specific example of a lens model analysis, which is representative of the *structural modeling* approach to decision analysis (Ford, Schmitt, Schechtman, Hults, & Doherty, 1989). As noted earlier, structural models describe the relationship between decision responses and their antecedent information stimuli. By generating a statistical model of the decision maker and the decision environment, the effects of information input factors such as cue range, cue validity, etc., on the decision output can be studied.

Brunswik's lens model (Brunswik, 1952) is the predominant structural model used in research on human judgment, forming the basis for the multiple cue probability learning (MCPL) research paradigm. The lens model views the world as having two "sides": the organism and the environment. Interfacing these two domains in the judgment context is information, termed *cues*. These cues (C_i) have a specified functional relationship to an environmental criterion (Y_e). In addition, these cues have a functional relationship to the judge's estimate (Y_s) of the environmental criterion. In MCPL experiments, subjects are presented with cue values and are asked to make a prediction of a criterion value. By manipulating relationships on the environmental side of the lens (e.g. $r_{C1,Ye}$; R_e) , effects can be measured by changes in the organism's (judge's) side of the lens. Three major outcome variables are assessed:

Achievement (r_a) is measured by the correlation between judgments (Y_s) and criterion values (Y_e) , and is generally interpreted as how well the judge is predicting the criterion.

Consistency (R_s) is measured by the multiple correlation between an individual's judgments (Y_s) and the cues (C_i), and indicates the predictability of judgments from the environmental cues.

Knowledge (G) is measured as the correlation between the least squares prediction of the criterion (Y_e) and the least squares prediction of the judgment (Y_s) from the cues. It measures the extent to which the subject has correctly detected the properties of the task (Hammond & Summers, 1972).

Assuming a linear use of cues, the above indices are functionally related in the following fashion:

$$r_a = G \times R_e \times R_s$$

where R_e is the relationship between the cues and the environment or the structure of the task as defined by the experimenter.

In addition, a *cue utilization* index (the correlation between cue values and subjects' judgments or the regression weight on a cue in predicting judgment) can be calculated to assess the extent to which subjects are using each cue in formulating their judgment. Figure 1 illustrates the lens model and its associated indices.

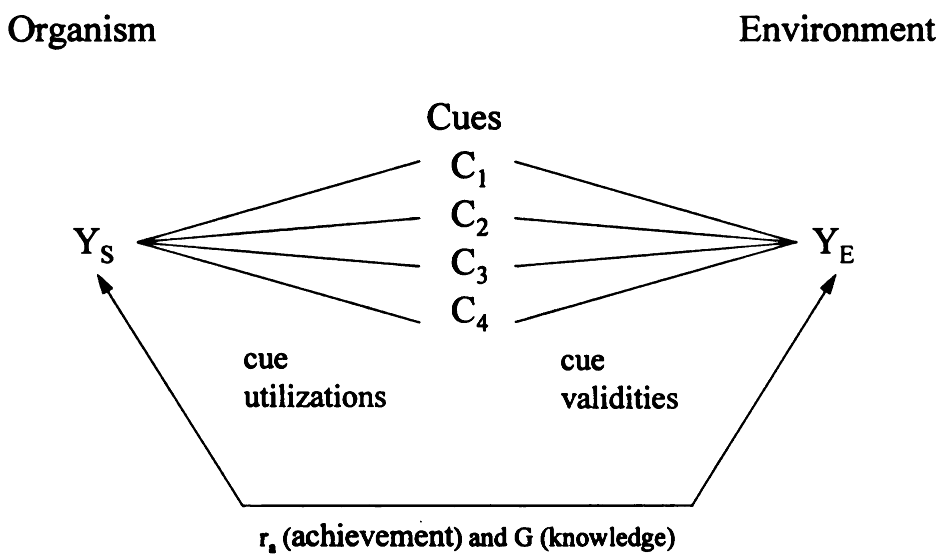

Figure 1. Brunswik's Lens Model

York, Doherty, & Kamouri (1987), in an MCPL study, addressed the influence of cue unreliability on judgment. Subjects were asked to predict plant growth based on five cue levels (for moisture and fertilizer) flashed before them. Reliability was manipulated by adding a random error term of varying standard deviation to each series of cue values at their presentation. York et al found no effects for cue unreliability on achievement, consistency, or cue utilization. The last finding is particularly interesting given the conflicting results found in the source credibility literature discussed below, as well as the fact that the York et al subjects *reported* giving greater weight (subjective weights) to the reliable cue. York et al note the consistency of their results with the work of Anderson's (1981) information integration theory, suggesting that people use an averaging rule in estimation tasks. Anderson's information integration theory posits that the role of averaging in estimation is a basic property of cognitive integration of information.

The MCPL study conducted by York et al (1987) notes theoretical limits to their findings:

"*Clearly, the present study does not provide the data that would be needed to discriminate among possible process models. While the data address directly certain questions about the impact of cue unreliability on judgment, they leave open the question of whether the subjects are taking an average, using the median score, or using some other data reduction strategy.*" (York, Doherty, & Kamouri, 1987)

The MCPL task used in this study builds on the work of York et al:

1) York et al (1987) used two cues, whereas this study used four. A two-cue judgment may have been too easy for subjects, so the incremental complexity added by

the presence of unreliable cues may have been negligible;

2) This study collected MCPL measures in concert with process tracing measures. Interpretation of these measures in light of each other can offer insights into the findings of York et al (1987). Specifically, it allows an interpretation based on both process and outcome components of the judgment and decision making situation.

Having presented the two major conceptual frameworks used in this study, and their relevant literature, the discussion now turns to relevant findings in the source credibility literature.

Literature Review: Source Credibility

Researchers examining the influence of information unreliability on decision making and judgment have chosen a broad range of conceptual and operational definitions, including instructional manipulation (Birnbaum, Wong, & Wong, 1976), self-generation of error (Brehmer, 1970), cue consistency (Borko & Shavelson, 1978), presentation of unreliable information (Kahneman & Tversky, 1973), and cue variability (Knowles, Hammond, Stewart, & Summers, 1971). In addition to the York et al (1987) study discussed above, the other primary body of research bearing on information unreliability and its effects is from the literature on source credibility.

Source credibility has been a prevalent means of framing the concept of information unreliability. Though the concept has not been linked to psychometric unreliability, which is the operationalization of unreliability used in this study, it is possible that the effects of deeming a source non-credible are conceptually similar to perceiving information as psychometrically unreliable. It does involve the not-so-dramatic leap of concluding that unreliable information will cause a source to be deemed non-credible. The results of studies involving source credibility are very robust, namely, that information from credible sources will have greater impact on decisions and estimations than information from a source lower in credibility. (Birnbaum et al, 1976; Rosenbaum & Levin, 1968). Findings have been consistent with the notion that there is a multiplicative relationship between credibility and information, with the credibility of a source functioning analogous to a regression weight for the information provided (Rosenbaum & Levin, 1968; Birnbaum et al, 1976; Beach, Mitchell, Deaton, & Prothero, 1978). The multiplicative relationship is robust, and has been observed in a variety of judgments, such as acceptability and probability (Beach et al, 1978), financial values (Birnbaum et al, 1976), and personal favorability (Rosenbaum & Levin, 1968). Further, the multiplicative effect was inversely related to the credibility of other sources, contradictory to results one would expect in additive or constant-weight averaging models, but consistent with a relative-weight averaging model (Birnbaum et al, 1976). Surber (1981) reported consistent results, showing greater reliability associated with greater perceived effects on performance, and greater reliability of one cue associated with lower perceived effects of a second, less reliable, cue. Levin (1980), in a study examining the effects of information differing only in variability, found differential weighing of information, but only in situations where instructional material linked variability to unreliability. This suggests that the mere perception of unreliability may affect decision making. A few articles are of note which discuss results contradictory to the aforementioned research. Contrary to the results consistent with a relative-weight averaging model, Surber (1984) found that, in situations where subjects judged ability and effort to be inversely related, higher reliability of one type of information *increased* the effect of another type of information. A group judging the relationship to be positive displayed results consistent with the averaging model (also shown in Surber (1981), i.e. higher reliability of one type of information was correlated with a lesser effect of the other type of information. Also, Schum (1975), in a summary of research on witness credibility, notes many empirical studies which conclude that people generally do not degrade the impact of testimony from sources in a manner consistent with the reliability or credibility of the source.

Having outlined the theoretical background and literature relevant to this research, the discussion now turns to methodologies. Specifically, a modified lens model will be presented which incorporates the reliability of information the decision maker has available. Then, process tracing methodology and associated decision making strategy classifications are introduced and discussed.

Reliability vs. Validity: A Caveat

The information presented to subjects in MCPL studies is usually done in a way that prevents the separate study of validity and reliability. Brehmer (1970) notes that uncertainty in MCPL studies is present solely in the relationship between the criterion and the environmental cue values (R_e). York et al (1987) defined and examined two sources of uncertainty by modifying the standard MCPL model. The uncertainty in the relationship between cues and criterion discussed by Brehmer (1970) remains. In addition, random error was added to a cue to generate multiple observations of the cue. See Figure 2 for a graphic description of the modified MCPL model. These multiple observations were presented to the subject. Ostensibly, subjects would infer the true value of the cue from these observations. The observed values for a given cue would conform to the psychometric definition of reliability, as they were a function both of the true value of the cue and a random error component.

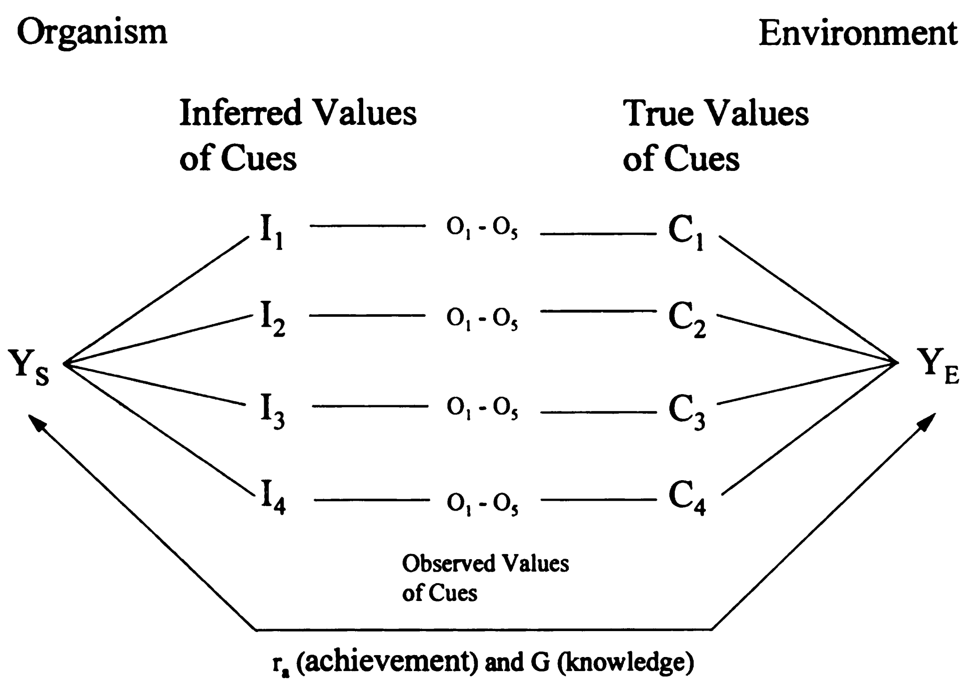

Figure 2. Modified Lens Model (from York, Doherty, & Kamouri, 1987)

Figure 3 illustrates examples of a standard MCPL trial, and two trials (low/high reliability) from the York et al modified method. Three assumptions are made in this reformulation of the MCPL task: 1) subjects can infer the true score for a cue from the multiple observations of the cue; 2) subjects can infer the reliability of a cue from the variability of the cue across and within trials; 3) subjects can infer the validity of a cue from its covariation with the criterion across trials.

Standard MCPL Trial (True Cue Values)	Cue 1 =	15
	Cue 2 =	25
	Cue 3 =	34
	Cue 4 =	8

		Observation 1	2	3	4	5
Modified MCPL Trial (High Reliability)	Cue 1 =	17	14	16	18	19
	Cue 2 =	22	27	25	28	24
	Cue 3 =	33	33	36	35	37
	Cue 4 =	7	6	9	8	9
Modified MCPL Trial (Low Reliability)	Cue 1 =	18	12	8	22	20
	Cue 2 =	20	31	32	30	21
	Cue 3 =	40	28	30	38	29
	Cue 4 =	2	15	4	18	14

Figure 3. Examples of MCPL Data

The modified MCPL model also raises an issue regarding the appropriate conceptualization of task uncertainty (R_e) (i.e. ecological validity). York et al (1987) calculated task uncertainty by regressing Y_e onto the true values for the predictors ($R(Y_e, C_t)$). This is essentially identical to the method used in standard MCPL studies. However, they also note that R_e could have been calculated by regressing Y_e onto the means of the *observed* cue values ($R(Y_e, \overline{C}_i)$). This may be a moot point, as the correlation between the true cue value and the mean of observed cue values is very high, and in fact, for very large numbers of observations, should equal 1.0. York et al (1987) confirmed this by reporting that the two alternatives for calculating R_e were "virtually numerically identical." (This is likely due to the means of observed score sets approaching the true value, which is guaranteed to occur when the observed values are calculated by adding a random error term to a true value.) However, a theoretical distinction can be made between the two measures based on the appropriateness of the index. From the subject's perspective, it would appear that regression on the means of the observed values is more appropriate. For someone who is familiar with the environment when it is undisturbed by error, regression on the true cue values may be a more appropriate index. This study utilized 100 repeated observations of the true value disturbed by an error term, thus making the mathematical difference between the two calculations negligible.

The typical MCPL study focuses on the structural relationships between cues and judgments and cues and an environment. We turn next to process tracing methods, which do not focus on outcomes per se, but on how information is sought when a person must make a decision.

Process Tracing Methodology

Process tracing is a methodology specifically developed to assess the process by which information is gathered and processed in decision making. Data collected during the decision process is used to determine the decision strategies likely used by the decision maker.

The two major process tracing methods are verbal protocols and information boards. Verbal protocols utilize self-reported statements regarding decision making to infer strategy selection. Verbal protocol methods are based on "think aloud" reporting of decision behavior by the decision maker during the decision process. Verbal statements are coded and submitted to further analysis. Information board methods present subjects with information matrices. Information matrices consist of information regarding several attributes' values for several decision alternatives. Individuals make decisions by searching for and accessing data in the matrix, selecting one alternative to reach their decision. A third, far less often utilized process tracing method is eye-movement analysis, in which efforts are made to identify on what information subjects fix their gaze as the decision-making process progresses (Russo & Dosher, 1983). While the measures obtained by this method are more precise than those obtained with verbal protocols and information boards, and are likely less subject to the decision maker's conscious biases, eye-movement analysis usually involves the use of expensive, cumbersome equipment. Due to such equipment needs, analysis is often limited to simplistic displays.

Information board methodology was utilized in this research for two reasons: 1) it represents the most cost-efficient alternative in terms of precision gained for cost and effort expended by the researcher; 2) it has a rich history of use in the decision-making literature, particularly the body of literature of primary relevance to this research.

Prior to an elaboration of process tracing measures used to assess and categorize decision strategies, the presentation of a taxonomy of decision strategies is warranted.

Decision Making Strategies

Decision-making strategies can be categorized into many different taxonomies, though a relatively small number have been used by the majority of researchers in decision-making processes. On a general level, strategies can be categorized based on their *compensatory* or *non-compensatory* nature. Compensatory strategies are strategies in which low values on one dimension can be compensated by high values on another dimension. Two prominent compensatory strategies are the *additive difference model* and the *linear model* (Ford et al, 1989). The additive difference model consists of two sequential phases in the comparison process: 1) alternatives are compared on each dimension, noting the differences between alternatives within each dimension; 2) differences are summed to produce an overall difference index. This index is then used to select one alternative. In the linear model, subjective values are summed across dimensions for each alternative. This summation is then used to select an alternative. Summation within alternatives (of objective differences or subjective values) is the identifying trait of compensatory strategies. Once a summary index is generated for each alternative, all information about dimensions is lost. This allows high values on one dimension to "make up for" low values on another dimension.

Non-compensatory models are strategies in which dimensions are used interactively, with high values on one dimension not necessarily compensating for low values on another dimension. Non-compensatory strategies are used to reduce the complexity of the information presented to the decision maker, and involve the use of simplifying rules to eliminate alternatives or dimensions during the decision process. These strategies are used until only one alternative remains or the decision problem has reached a sufficiently low level of complexity to allow the use of one of the compensatory strategies described above. Though many possible non-compensatory strategies have been developed, five have received the majority of attention in process tracing research (Ford et al, 1989; Payne, 1976): *conjunctive*, *disjunctive*, *lexicographic*, and *elimination by aspects* (EBA).

Utilization of a conjunctive rule involves establishing a minimum criterion value for each dimension, and rejecting alternatives which do not exceed the criterion for each dimension (Svenson, 1979). Use of a disjunctive rule constrains the selected alternative to exceed a criterion on at least one dimension and the remaining alternatives to be equal or less than the criterion on different dimensions (Olshavsky, 1979). Lexicographic and EBA rules are different from the previous two in that they require a judgment to be made regarding the importance of each dimension. In using a lexicographic strategy, a decision maker selects the alternative with the highest value on the most important dimension (Payne, 1976). EBA strategy use (Tversky, 1972) involves the establishment of a criterion for the most important dimension. Alternatives not meeting this criterion are eliminated from contention. All remaining alternatives are then evaluated based on a criterion for the second most important dimensions. This process continues until only one alternative remains.

Search Indicators in Process Tracing

In order to assess usage of the decision making strategies discussed above, process tracing investigations using information board methodology have primarily utilized three search indicators: *depth of search*, *pattern of search*, and *latency of search*. Depth of search refers simply to the number of alternative-dimension pairs accessed by the subject. Search pattern is assessed by comparing the *n*th piece of information accessed to the *n*th+1 piece of information accessed (Payne, 1976). Search pattern data can be used to categorize a search pattern as *intra*dimensional or *inter*dimensional, based on sequential accesses within or across dimensions. Latency of search is defined as the elapsed time viewing a single piece of information. Summed over a decision task, this measure can be more accurately thought of as decision latency, as it represents the total time to selection of an alternative.

Search depth and search pattern are primarily used in concert to assess the strategy being applied by the decision maker. Higher search depths are consistent with use of a compensatory strategy. Lower search depths indicate utilization of a non-compensatory strategy. Comparison of search depth indices across alternatives results in the construction of a fourth index: variability of search. Low variability of search (searching same amount of information for each alternative) suggests use of a compensatory strategy. High search variability is thought to indicate use of non-compensatory strategies. Response latency has been shown to be strongly associated with search depth measures. Generally, lower mean search times across all dimensions will result from searching less information. Thus, decreases in mean search time have been associated with use of non-linear, simplifying strategies (Payne & Braunstein, 1978; Olshavsky, 1979). Categorization of search patterns into *intra-* and *inter-* dimensional searches has been shown to suggest specific strategies within both compensatory and non-compensatory strategy sets. Intradimensional search patterns are indicative of additive difference or EBA rule application. Interdimensional search patterns suggest use of the linear model or a conjunctive rule. Lexicographic and disjunctive rule utilization are more difficult to assess, as they result in search patterns very similar to EBA and conjunctive rule application, respectively. These strategies can be differentiated by careful examination of the search order and pattern in the information matrix. This has been done successfully by using expert judges to supplement the process tracing indicators when classifying strategy utilization. (Gilliland, Schmitt, & Wood, 1993; Gilliland, Wood, & Schmitt, 1994). Patterns of search that are representative of each of these strategies are presented in Appendix A. These were prototypes against which the expert judges assessed strategy use for each subject.

Having elaborated the methods to be used for each of the conceptual frameworks, and the relevant performance indices, I turn to my major objective, which is a description and rationale for my hypotheses regarding the role of unreliability in judgment and decision making. My contention is that unreliability will affect both the outcome and strategy associated with judgment and decision making. In accordance with the Beach and Mitchell model, the primary hypothesis underlying this thesis is that information unreliability will be perceived as an increase of cognitive demand in the decision task and that it will result in a diminished probability of selecting analytical, compensatory strategies, and an increased probability of selecting non-analytical, non-compensatory strategies.

Research Design

The proposed research design involved the use of four groups of participants who were presented decision tasks that varied with respect to the reliability of the stimulus information. Reliability was manipulated in a series of MCPL trials in which participants were asked to make judgments and were provided with feedback. Condition 1 (perfect reliability) was presented four cues with no information regarding reliability (one observation). Condition 2 (high reliability) was presented four cues of high reliability. Condition 3 (mixed reliability) was presented four cues with mixed reliability - two of high reliability and two of low reliability. Condition 4 (low reliability) was presented four cues of low reliability. Reliability was manipulated through presentation of 50 MCPL trials during which time several dependent variables were collected. In addition, the impact of reliability was assessed in five process tracing trials.

Hypotheses

The first set of hypotheses addresses the predicted effects of information unreliability on judgment indices in the MCPL task. These hypotheses are based on the findings from the source credibility and MCPL literature. The logic underlying these hypotheses is that the unreliability associated with cues, along with the total unreliability present in the judgment environment will have direct effects on cue utilization indices and detrimental impacts on the MCPL performance indices.

H_1 : *Subjects' cue utilization indices for reliable cues will be higher than cue utilization indices for unreliable cues.*

H_2 : *Subjects will have higher achievement indices in conditions with more reliable cues. (Perfect Reliability > High Reliability > Mixed Reliability > Low Reliability)*

H_3 : *Subjects will have higher knowledge indices in conditions with more reliable cues. (Perfect Reliability > High Reliability > Mixed Reliability > Low Reliability)*

H_4 : *Subjects will have higher consistency indices in conditions with more reliable cues. (Perfect Reliability > High Reliability > Mixed Reliability > Low Reliability)*

A second set of hypotheses address predictions regarding the questions asked at the end of the MCPL task. H_5 and H_6 are based on subjects' perceptions of the MCPL task and its cue characteristics. H_5 is a manipulation check to determine whether subjects are perceiving the manipulation. H_6 is an extension of H_1, in that it predicts subjects will accurately perceive their own cue usage.

H_5 : *Subjects will report cues higher in reliability as subjectively more reliable.*

H_6 : *Subjects will report higher usage of the cues high in reliability relative to cues of low reliability.*

The third set of hypotheses are based on the process tracing findings regarding search indicators and predictions from the Beach & Mitchell model. Conceptually speaking, this set of hypotheses suggests perceptions of greater task complexity will, in turn, lead to increased utilization of non-compensatory strategies. H_7, H_8, and H_9 all involve predictions regarding the indicators of decision strategy. H_{10} and H_{11} are predictions based on strategy classification.

H_7 : *Increases in unreliability will be associated with decreases in search depth.*

(Perfect Reliability > High Reliability > Mixed Reliability > Low Reliability)

H_8 : *Increases in unreliability will be associated with increases in search variability.*

(Perfect Reliability < High Reliability < Mixed Reliability < Low Reliability)

H_9 : *Increases in unreliability will be associated with decreases in mean search times.*

(Perfect Reliability > High Reliability > Mixed Reliability > Low Reliability)

H_{10} : *Non-compensatory strategy utilization will occur more frequently as unreliability increases.*

(Perfect Reliability < High Reliability < Mixed Reliability < Low Reliability)

H_{11} : *Compensatory strategy utilization will occur less frequently as unreliability increases.*

(Perfect Reliability > High Reliability > Mixed Reliability > Low Reliability)

An additional hypothesis addresses issues raised in the source credibility literature. This hypothesis also extends the logic of the third hypothesis set to a lower level: between dimensions rather than between groups. Findings from the source credibility literature suggest information perceived low in credibility will have less impact and be utilized less than information perceived as credible. (Levin, 1980; Birnbaum et al., 1976) Though this study did not measure perceived credibility, the following hypothesis is based on the assumption that perceived reliability will affect information usage in a similar fashion as perceived credibility. This hypothesis is evaluated only in Condition 3 (mixed reliability), where both reliable and unreliable information is present.

H_{12} : *Subjects will have greater search depths for reliable cues than for unreliable cues.*

A fifth set of hypotheses seeks to evaluate the relative weighting relationship (Birnbaum et al., 1976; Surber, 1981) between credible and noncredible information use when both types of information are present. This will be evidenced by a dimension's utilization being determined not only by its own reliability, but by the reliability of other dimensions presented in the matrix. Specifically, it is predicted that a given reliable dimension's impact will be greater when it is presented with unreliable dimensions than if presented with other reliable dimensions. Conversely, a given unreliable dimension's impact will be less when it is presented with reliable dimensions than if presented with other unreliable dimensions. Operationally, these predictions result in the comparisons listed below. Since the source credibility literature cited addresses information utilization, hypotheses will involve both search depth and cue utilization measures.

H_{13} : *Subjects in the mixed reliability condition will have greater search depth in reliable dimensions than subjects in the high reliability condition.*

H_{14} : *Subjects in the mixed reliability condition will have lower search depth in unreliable dimensions than subjects in the low reliability condition.*

H_{15}: *Subjects in the mixed reliability condition will have greater cue utilization in cues of low reliability than subjects in the high reliability condition.*

H_{16}: *Subjects in the mixed reliability condition will have lower cue utilization in cues of high reliability than subjects in the low reliability condition.*

METHOD

Subjects

Subjects were obtained from the university's experimental subject pool, and received nominal course credit for their experimental participation. Subjects were also informed of cash prizes awarded for the top six performers in the two experimental tasks. A sample size of 144 subjects was collected based on a power analysis, assuming a medium effect size, a four group mean comparison, and a desired power of .80. (Cohen, 1988).

Procedure

As previously mentioned, the study is a 4 group design (1 - no information given and reliability assumed perfect, 2 - four cues high in reliability, 3 - two cues high in reliability and two cues low in reliability, 4 - four cues low in reliability). Reliability is manipulated as a between subjects factor. The MCPL dependent variables were collected in sixty trials. Dependent measures were not collected on the first ten trials, as these were designed for the subject to learn the task and cue-criterion relationships. For blocking purposes, 45 of the remaining fifty trials (trials 6 through 50) were used to calculate dependent measures. Thus, dependent measures reflect performance after fifteen unscored "practice" trials. Although a review of MCPL literature provided no concrete value at which subjects' learning asymptotes, studies seem to generally show learning leveling off around twenty (Hammond & Summers, 1972) to forty (Hammond & Summers, 1965) trials. In addition, the fifty scored trials allowed the same information to be used in both the MCPL and process tracing tasks. A portion of the fifty trials was used as the thirty (5 trials x 6 alternatives/trial) alternatives in the process tracing task. The process tracing dependent variables were collected in five trials, allowing the collection of reliability indices on the search indicators previously discussed.

Tasks

The experiment involved two phases. The first phase was an MCPL task used to "teach" the subjects to associate varying reliability with certain cues. The manipulation of reliability occurred in this phase. The second phase was a process tracing task in which subjects selected an alternative based on a decision matrix and, ostensibly, the observations made in the MCPL task. This phase involved the five repeated trial decisions.

Prior to beginning the first task, subjects completed a cognitive ability measure, the Wonderlic Personnel Test. Scores on this measure were used as a covariate in several of the analyses. It is likely that cognitive ability is highly related to many of the dependent measures used in this study. Implicit in the Beach and Mitchell contingency model is the notion that perceived cognitive costs are instrumental in strategy selection. It is not unreasonable to consider individuals of higher cognitive ability as having greater cognitive resources for potential allocation. Thus, to an individual of high cognitive ability, the additional cognitive costs of making a decision in a complex, unreliable environment, may be relatively negligible.

MCPL Task

The MCPL task instructed subjects to estimate the snowfall generated from an incoming storm based on four labeled cues: wind speed (1), barometric pressure (2), humidity (3), and cloud ceiling (4). True cue values and criterion values were generated from a normal distribution with a mean of 20 and standard deviation of 5. Thus, most true values fell in the 10-30 range. All cue validities were .5 and cue intercorrelations were .00 (both determined before the addition of random error). To manipulate reliability, the values of each cue were treated as a true value to which random error was added to create observed values. One hundred observed values were generated for each cue on each trial, and were presented to subjects as a graphic bar on a meter presented on the computer screen. The one hundred observations were presented over approximately an eight second interval, resulting in the screen displaying a "jittery" meter. Observations of cues with high reliability were determined by adding a normally distributed random variable with a mean of 0.0 and a standard deviation of 1.0. Observations of cues with low reliability were determined by adding a normally distributed random variable with a mean of 0.0 and a standard deviation of 6.0. These values for standard deviation were based on a pilot sample of twenty subjects who indicated a perceived difference in variability. The one hundred generated values for each cue were rounded to the nearest integer. In cases where the added error distribution resulted in an observation outside the stated range of a cue, the value was truncated to the meter's limit.

Observed values were displayed simultaneously in meter format for all four cues. The meters displayed the one hundred observed cue levels for 8 seconds. After all observed cue levels were displayed, subjects were required to estimate the amount of snowfall on a 1-40 scale. Subjects could review the presented cues by entering "0" for a judgment. Data was not collected on this behavior as the majority of pilot subjects and observed experimental subjects rarely reviewed cue observations after the first few trials. Immediately after an estimate was made, the computer displayed the actual snowfall for approximately 3 seconds. Fifty total trials were completed and recorded. Each trial had identical cue labels and reliability properties. An example screen image of a single set of observations within a trial is shown in Figure 4.

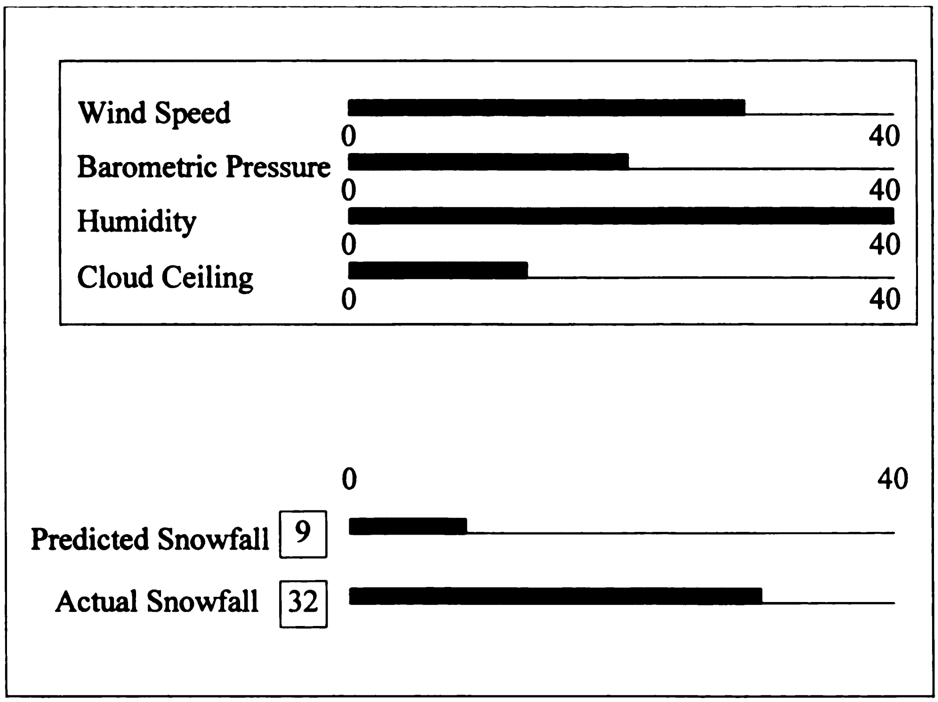

Figure 4. Example Screen Image

At the end of the MCPL task, subjects were given a short computerized self-report survey in which they were asked the extent to which they felt each cue varied within the trials and to what extent they felt they used each cue in their estimations. These items can be found in Appendix B. Appendix C contains the complete set of stimuli for the 50 MCPL trials.

Conditions in MCPL Task

Three groups received varying numbers of cues high and low in reliability. For two groups (0/4 reliable cues) all cues had equal reliabilities and were either adjusted by an error term of 1.0 or 6.0 standard deviation (noted earlier). In the group receiving a cue set with mixed reliabilities (2 high reliability, 2 low reliability), cues were completely crossed to determine which were reliable (resulting in 6 sub-groups within this group). This accounted for all possible combinations of cue label and cue reliability, preventing any confounding of cue label and cue reliability, and allowing separate assessment of both effects. A fourth group received no information regarding unreliability, and was presented with one observation in the MCPL trials. Thus, to these individuals, the computer displayed the same meter reading for each cue for the entire 8 second duration. This method of presentation is identical to that of a traditional MCPL study.

Process Tracing Task

Following the MCPL task, each subject performed a process tracing decision task. Subjects were asked to select a sled dog convoy route from among six cities. The accompanying story instructed subjects to select a route which they felt would have the least snowfall from an approaching storm. Attributes for the six routes had labels identical to those in the MCPL task. Each subject made five sequential decisions in order to facilitate the identification of search strategies and generate reliable search indicators. The information board methodology was used and presented via a computer program. The information board consisted of a matrix of six route alternatives and four sources of information. The information in the matrix was selected from the cue values presented in the MCPL task. Values were selected which resulted in a sufficient snowfall variance across routes to allow one route to be the correct choice.

The computer initially presented the subject with an empty information matrix. The subject was prompted to enter the number of the route they wished to examine. They were then asked to enter the number of the attribute they would like to examine for the previously selected route. The computer then displayed the value for the specified cell in the matrix. A keypress returned them to the matrix and route selection prompt. The entire process was then repeated with subsequent selection of route-attribute pairs. When subjects felt they could make an accurate decision regarding the routes, they were instructed to press "d" on the keyboard (rather than a route or attribute number) and were presented with a prompt to enter the number of the route selected. The entire decision process was repeated in four additional trials (with different cell values, but identical attribute names). Appendix C contains the complete set of stimuli for the five process tracing trials.

Dependent Measures: MCPL Task

Four lens model dependent measures were assessed: achievement (r_a), consistency (R_s), cue utilization, and knowledge (G). In addition, subjects were asked two blocks of questions regarding perceived within-trial cue variability and cue usage, respectively. The first block of four questions involving perceived cue variability asked subjects to estimate the range within which a cue "jumped around". Anchors were constructed in a manner so one response referred to the correct amount of variability for the cue. These items can be found in Appendix B. A second question block asked subjects to allocate 100 points among the four cues in a fashion reflecting their own utilization of each cue. For instance, subjects would allocate 25 points to each cue if they felt each were equally important in their judgments. These questions can also be found in Appendix B.

Dependent Measures: Process Tracing Task

Three dependent measures were assessed: Search depth, search variability, and search latency. The operationalization of these indices is discussed in a previous section. Since there were five process tracing trials, internal consistency reliability analyses for these five-item measures were conducted.

In addition, search pattern information was used to classify subjects' decision-making strategy. Observed search patterns were compared to prototypical matrices for each decision making strategy (i.e. linear, additive difference, lexicographic, EBA, conjunctive, disjunctive) Judgments were made by at least two raters. Interrater reliability regarding strategy utilization was assessed using a random subsample of 75 process tracing trials selected from a sample of all subjects. Two trials were later noted to have zero search depths, and were removed from the reliability assessment. This resulted in 73 total judgments by two raters. The two raters agreed in their categorization of these search strategies in 63% of the cases; kappa was .512. Judgment crossclassifications can be seen in Table 1. The "Uncertain" category reflects a situation in which the judge was not comfortable assigning a strategy class to the observed process tracing matrix. In the majority of cases where both raters categorized a pattern as uncertain, there was no clear systematic search pattern present in the matrix. Neither rater noted the presence of disjunctive or lexicographic decision strategies. It should be noted that prior to training raters to identify strategies using the experimental matrices, raters were shown example search patterns from larger search matrices (12 dimension by 8 alternative). The shift to a 4 dimension by 6 alternative matrix may have blurred the distinction between some search patterns, as well as narrowed the range of strategies actually used by the subjects. This may be responsible for the lack of disjunctive and lexicographic judgments, as well as any disagreements in classification of judgments.

Table 1: Strategy Judgment Classifications

		Rater 1					Row %
		Uncertain	**Add. Diff.**	**EBA**	**Linear**	**Conj.**	
	Uncertain	**10**	**0**	**0**	**0**	**0**	**.137**
	Add. Diff.	**4**	**17**	**6**	**0**	**0**	**.370**
Rater 2	**EBA**	**0**	**7**	**7**	**0**	**0**	**.192**
	Linear	**3**	**1**	**1**	**12**	**0**	**.233**
	Conj.	**5**	**0**	**0**	**0**	**0**	**.068**
Col %		**.301**	**.343**	**.192**	**.164**	**.000**	**1.000**

A large number of disagreements in strategy classification occurs between the two raters' judgments of Additive Difference and Elimination by Aspects strategies. This is likely due to the similarity in prototypical search patterns for these strategies (presented in Appendix A), as well as different judgment policies of the two raters. Rater 1 indicated that EBA strategy was present whenever the search pattern was similar to the EBA prototype, even if the last two cells accessed did not result in a within-dimension comparison (i.e. the last accessed cell was the only one for the dimension). Rater 2 indicated that, in cases where only one cell was accessed for the last dimension searched, a comparison was not being made, and it remained an Additive Difference strategy. This resulted in some of Rater 1's EBA classifications being classified as an Additive Difference strategy by Rater 2. Appendix A shows an example of such a matrix. For the purposes of analysis, Rater 2's judgments were used, as a single access of information within a dimension could not possibly indicate any comparison within the dimension.

RESULTS

MCPL - Performance Indices / Self-Report Measures

Hypothesis 1

Hypothesis 1 predicted that subjects' cue utilization indices for reliable cues would be higher than their cue utilization indices for unreliable cues.

Only cues from experimental conditions 2 (high reliability - 4 cues SD=1), 3 (mixed reliability - 2 cues SD 1, 2 cues SD 6) , and 4 (low reliability - 4 cues SD=6) were used in this analysis, since subjects in condition 1 (perfect reliability, 4 cues SD 0) were given no information about reliability.

Subjects' judgments of snowfall were regressed on the four cue values to obtain beta weights for each cue. Since the cues were uncorrelated and on equal scales, these weights are appropriately used as cue utilization indices. Regressions were run separately for three trial blocks, consisting of trials 6 through 20, 21 through 35, and 36 through 50, respectively. Thus, in total, each subject contributed 12 weights (4 per block). This was done to assess potential differences in perrformance indices over time due to learning or other effects. Each beta weight was then assigned to a cell based on the cue's reliability (1=SD 1, 2=SD 6), its label (1=Wind Speed, 2= Barometric Pressure, 3= Humidity, 4=Cloud Ceiling), and its source block, resulting in a 3 (block) x 4 (label) x 2 (reliability) design.

Due to the nature of this analysis, cue utilization indices were treated as observations. Thus, it should be noted that the reliability factor in this analysis is not identical to the *experimental* condition of unreliability. However, the two factors are related. Cue utilization indices assigned to the low reliability group are from subjects in experimental conditions of mixed (#3) and low (#4) reliability. Cue utilization indices assigned to the high reliability group are from subjects in experimental conditions of mixed (#3) and high (#2) reliability.

The analysis was conducted using a repeated measures ANCOVA with a cognitive ability measure as a covariate. Block (1-3) was used as the repeated within-subjects factor, with Label and Reliability assessed as between-subjects factors. (Note: Using cue utilization indices as the unit of observation precluded treating cue label as a within subjects factor.) The repeated measures ANCOVA statistics are shown in Table 2. Within-subjects F-tests are based on a Wilks' Lambda multivariate test approximation.

Table 2. Cue Utilization by Block, Reliability and Label: ANCOVA Summary

Between-Subjects	**df**	**F**	**p ≤**
Cog. Ability	1	0.78	.38
Reliability	1	0.11	.74
Label	3	1.09	.35
Reliability x Label	3	*3.55*	*.015*
Error	423		
Within-Subjects			
Block	2,422	0.83	.43
Block x Cog. Ability	2,422	0.65	.52
Block x Reliability	2,422	0.48	.62
Block x Label	6,844	*6.35*	*.0001*
Block x Rel. x Label	6,844	0.92	.48

Mean betas adjusted for the cognitive ability covariate and their respective standard deviations, denoted by parentheses, are shown in Table 3.

Table 3. Cue Utilization (Mean Beta and SD) by Block, Reliability, and Label

Label	Reliability	Block 1	Block 2	Block 3	Mean(Label)
Wind Speed	**High Rel.**	.141 (.196)	.148 (.169)	.246 (.248)	.22
	Low Rel.	.219 (.212)	.244 (.208)	.324 (.298)	
Bar. Press.	**High Rel.**	.268 (.256)	.260 (.169)	.256 (.200)	.26
	Low Rel.	.261 (.285)	.226 (.191)	.267 (.196)	
Humidity	**High Rel.**	.255 (.243)	.263 (.243)	.325 (.219)	.25
	Low Rel.	.187 (.242)	.208 (.268)	.234 (.226)	
Cloud. Ceil.	**High Rel.**	.230 (.223)	.280 (.173)	.191 (.202)	.22
	Low Rel.	.262 (.277)	.209 (.227)	.160 (.221)	

Note: Numbers in parentheses are cell standard deviations.

Table 4 shows marginal means for the Reliability x Label interaction. The significant interaction effect is likely due to the different use of Wind Speed and Humidity cues in High and Low Reliability conditions. Though no main effects were significant, the interaction suggests that Wind Speed was used more when it was unreliable than when it was reliable, and the reverse was true for Humidity.

Table 4. Cue Utilization by Reliability and Label

		Label				
		Wind Spd.	*Bar. Press.*	*Humidity.*	*Cld. Ceil.*	*Mean*
	High	.178	.261	.281	.233	*.238*
Reliability	*Low*	.262	.251	.210	.210	*.233*
	Mean	*.220*	*.256*	*.246*	*.221*	

Figure 5 illustrates the Block x Label interaction. The interaction can be interpreted as a tendency for subjects to shift emphasis from Cloud Ceiling to Wind Speed across trial blocks. This is potential evidence of a shifting weighing strategy. Further elaboration and potential implications of this finding can be found in the Discussion section.

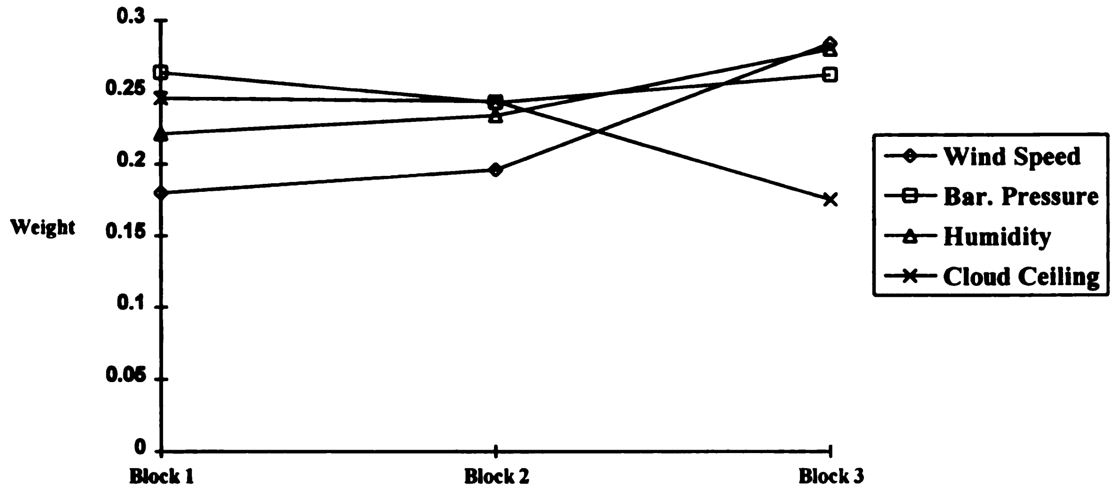

Figure 5. Cue Utilization by Label and Block

Other than the significant effects noted above, results of analyses for Hypothesis 1 suggest that the reliability of a cue did not influence its utilization in the judgment task, disconfirming the hypothesis.

Hypothesis 2

Hypothesis 2 predicted subjects will attain higher achievement scores in conditions with more reliable cues.

Achievement measures (r_a) were calculated for each subject on each trial block. These scores were normalized using a Fisher's Z transformation and submitted to a 3 (experimental condition) by 3 (block) repeated measures ANCOVA analysis with cognitive ability as a covariate, experimental condition as a between-subjects factor, and Block as a repeated measures factor. (see Table 5) Cell means, standard deviations, and r equivalents are shown in Table 6.

Table 5. Achievement by Experimental Condition: ANCOVA Summary

Between Subjects	**df**	**F**	**p ≤**
Cognitive Ability	1	*8.69*	*.004*
Exp. Condition	3	*2.61*	*.05*
Error	139		
Within Subjects			
Block	2,138	.05	.95
Block x Cog.Ability	2,138	1.75	.18
Block x Exp. Cond.	6,276	.51	.81

Post hoc tests (protected t-tests) of differences among the individual experimental conditions were performed because of the significant effect for experimental condition (see Table 5 for results).

Subjects in conditions with perfect reliability and high reliability had significantly higher achievement scores than subjects in the low reliability condition. Although other mean differences are not significant at the .05 level, the ordering of means is consistent with Hypothesis 2 (i.e., subjects in the perfect reliability condition have the highest achievement, with descending means progressing to the low reliability condition.)

Table 6. Achievement by Experimental Condition

Exp. Condition		Block 1	Block 2	Block 3	(across 3 blocks) Mean z	Mean r
	$Mean_Z$	.963	.708	1.00		
Perfect Rel.	SD_Z	.426	.361	.382	$.890_A$	*.657*
	r	.688	.569	.714		
	$Mean_Z$	.979	.665	.956		
High Rel.	SD_Z	.438	.314	.421	$.867_A$	*.645*
	r	.697	.548	.689		
	$Mean_Z$	.825	.608	.868		
Mixed Rel.	SD_Z	.454	.293	.434	$.767_{AB}$	*.590*
	r	.614	.513	.643		
	$Mean_Z$	.825	.603	.771		
Low Rel.	SD_Z	.394	.279	.355	$.733_B$	*.582*
	r	.628	.512	.606		
	Mean Z	*.898*	*.646*	*.899*		
	Mean r	*.657*	*.536*	*.663*		

Note: Subscripts denote Fisher Z means significantly different at the .05 level or better.

The significant effect associated with the cognitive ability covariate suggests subjects of higher cognitive ability were overall, higher in accuracy of judgment.

Hypothesis 3

Hypothesis 3 predicted that subjects will have higher knowledge indices in conditions with more reliable cues.

Knowledge (G) scores were calculated and prepared in a similar fashion as the aforementioned achievement indices were, in regard to transformation and calculation by blocks. These transformed knowledge indices were submitted to a 3 (experimental condition) by 3 (block) repeated measures ANCOVA with cognitive ability as the covariate (see Table 7). Means and related statistics are shown in Table 8.

Table 7: Knowledge by Experimental Condition: ANCOVA Summary

Between Subjects	**df**	**F**	**p ≤**
Cognitive Ability	1	*8.24*	*.0047*
Exp. Condition	3	1.21	.31
Error	139		
Within Subjects			
Block	2,138	.32	.73
Block x Cog.Ability	2,138	1.24	.29
Block x Exp. Cond.	6,276	.50	.81

There were no between-subjects or within-subjects significant effects, disconfirming Hypothesis 3. The cognitive ability covariate was significantly related to

knowledge indices, with higher knowledge indices for subjects with higher cognitive ability.

Hypothesis 4

Hypothesis 4 predicted that subjects will have higher consistency indices in conditions with more reliable cues.The third and final MCPL index, consistency (R_s) was generated and analyzed in a manner identical to the preceding two indices. The ANCOVA summary table is shown in Table 9. Cell means are shown in Table 10.

Table 8. Knowledge by Experimental Condition

Exp. Condition		Block 1	Block 2	Block 3	(across 3 blocks) Mean z	Mean r
	Mean$_Z$	1.45	1.13	1.57		
Perfect Rel.	*SD$_Z$*	.707	.690	.754	*1.38*	*.791*
	r	.815	.721	.837		
	Mean$_Z$	1.38	1.13	1.34		
High Rel.	*SD$_Z$*	.738	.620	.560	*1.28*	*.769*
	r	.781	.719	.806		
	Mean$_Z$	1.19	1.03	1.29		
Mixed Rel.	*SD$_Z$*	.924	.559	.752	*1.17*	*.720*
	r	.712	.697	.755		
	Mean$_Z$	1.37	1.08	1.26		
Low Rel.	*SD$_Z$*	.769	.576	.659	*1.24*	*.740*
	r	.770	.706	.762		
	Mean Z	*1.35*	*1.09*	*1.36*		
	Mean r	*.769*	*.710*	*.790*		

Table 9. Consistency by Experimental Condition: ANCOVA Summary

Between Subjects	df	F	p ≤
Cognitive Ability	1	*6.98*	*.0092*
Exp. Condition	3	*4.42*	*.0053*
Error	139		
Within Subjects			
Block	2,138	.98	.27
Block x Cog.Ability	2,138	.99	.38
Block x Exp. Cond.	6,276	.92	.48

Table 10. Consistency by Experimental Condition

Exp. Condition		Block 1	Block 2	Block 3	(across 3 blocks) Mean z	Mean r
	$Mean_Z$	1.41	1.16	1.37		
Perfect Rel.	SD_Z	.399	.410	.353	1.31_A	*.830*
	r	.857	.776	.856		
	$Mean_Z$	1.54	1.15	1.42		
High Rel.	SD_Z	.398	.309	.440	1.37_A	*.846*
	r	.888	.797	.852		
	$Mean_Z$	1.31	1.09	1.37		
Mixed Rel.	SD_Z	.390	.265	.401	1.26_{AB}	*.819*
	r	.832	.779	.846		
	$Mean_Z$	1.22	1.03	1.22		
Low Rel.	SD_Z	.363	.310	.363	1.16_B	*.786*
	r	.807	.744	.807		
	Mean Z	*1.37*	*1.11*	*1.35*		
	Mean r	*.846*	*.774*	*.840*		

Note: Subscripts denote Fisher Z values significantly different at .05 level or better

Post hoc tests (protected t-tests) of mean differences among experimental conditions were conducted because of the significant effect observed for experimental condition. Results are highly parallel with those found for achievement scores. Subjects in conditions of perfect and high reliability show significantly higher consistency scores than subjects in a low reliability condition. The cognitive ability covariate again shows a significant effect, with subjects of higher cognitive ability having higher consistency indices.

Recall that achievement is a multiplicative function of knowledge, the environment, and consistency. Since the environment is a constant, the results related to Hypotheses 2, 3, and 4 indicate differences in subjects' achievement in conditions of high and low reliability are due to their increased consistency.

Hypothesis 5

Hypothesis 5 predicted that subjects would report cues higher in reliability as subjectively more reliable.

The data were analyzed in a similar manner as in Hypothesis 1, except the subject's estimate of variability was analyzed instead of a beta weight, and blocking was not a relevant consideration. This resulted in a 2 (Reliability) x 4 (Label) design. The subjective variability estimate was a value from 1 to 6, which is the subject's answer to the four questions in Appendix B. Thus, four values were obtained from each subject. Data were analyzed using an ANOVA, with Reliability and Label as between-subjects factors. The ANOVA summary table is shown in Table 11. Cell means and standard deviations, as well as relevant comparisons, are shown in Table 12. Means significantly different at .05 level are denoted by subscripts.

Table 11. Subjective Variability by Reliability and Label: ANOVA Summary

Between-Subjects	df	F	p ≤
Reliability	1	*66.11*	*.0001*
Label	3	.14	.93
Reliability x Label	3	2.58	.06
Error	420		

Table 12. Subjective Variability by Reliability and Label

		Label				
		Wind Spd.	*Bar. Press.*	*Humidity*	*Cld. Ceil.*	*Mean*
	High	2.85 (.98)	3.06 (1.04)	2.78 (1.04)	2.78 (1.00)	*2.87_A*
Reliability	*Low*	3.66 (.78)	3.41 (1.01)	3.86 (1.09)	3.73 (1.20)	*3.67_B*
	Mean	*3.26*	*3.24*	*3.32*	*3.26*	

Note: Subscripts denote means significantly different at the .05 level or better

Subjects tended to perceive greater variability in cues of lower reliability, suggesting the reliability manipulation was correctly perceived. Two anchors in the subjective variability items referred to the exact amount of variability in the cue. For cues of low reliability, the correct response was 5. For cues of high reliability, the correct response was 3. It is of note that the means for the reliability effect tend to suggest subjects underestimated actual variability, especially for unreliable cues.

Hypothesis 6

Hypothesis 6 predicted that subjects would report higher usage of the cues high in reliability relative to cues of low reliability.

In a manner identical to that presented in Hypothesis 5, subjects' judgments of their own weighting of cues were analyzed in a 2 (Reliability) x 4 (Label) ANOVA. These values were the number of points (0-100) assigned to each cue in the perceived importance items in Appendix B. The summary table is presented in Table 13. Means and standard deviations are shown in Table 14.

Table 13. Subjective Weight by Reliability and Label: ANOVA Summary

Between-Subjects	df	F	p ≤
Reliability	1	.12	.73
Label	3	*10.46*	*.0001*
Reliability x Label	3	.48	.69
Error	420		

Table 14. Subjective Weight by Reliability and Label

		Label				
		Wind Spd.	*Bar. Press.*	*Humidity*	*Cld. Ceil.*	*Mean*
Reliability	*High*	22.19 (7.77)	26.56 (7.21)	27.79 (6.90)	21.73 (8.51)	*24.85*
	Low	24.72 (12.35)	27.34 (8.24)	27.57 (8.30)	22.10 (10.19)	*25.14*
	Mean	*23.46_B*	*26.95_A*	*27.68_A*	*21.91_B*	

Note: Subscripts denote means significantly different at the .05 level or better

Post hoc tests (protected t-tests) were conducted among the four cues due to the significant main effect for label. Means significantly different at the .05 level or better are denoted by subscripts. There was no significant effect for reliability on subjective weights, disconfirming Hypothesis 6. However, subjects did perceive that they placed greater weight on Humidity and Barometric Pressure than on Wind Speed and Cloud Ceiling. These values correspond with the mean beta weights for each label (presented in Table 3), suggesting that subjects were aware of, and able to accurately report their weighting strategy. However, they did not report higher usage (nor did they use) of the cues higher in reliability.

Process Tracing - Search Indicators / Strategy Utilization

Data Cleaning

For the purposes of analyzing Hypothesis 7 through Hypothesis 15, measures were taken to ensure all cases were appropriate to submit to analysis. In hypotheses in which search indicators are evaluated, subjects searching 10 or fewer cells over the five process tracing trials were eliminated from analysis. This criterion was arbitrarily selected to eliminate subjects with an average of 2 or fewer searches per process tracing trial. It was thought that subjects displaying such patterns of search either did not take the task seriously, or did not sufficiently understand the task instructions. This criterion resulted in the elimination of 22 of 144 subjects from analysis.

In addition, 55 subjects were judged not to be using any systematic strategy to make decisions in the process tracing task. These subjects were not removed from the analyses for Hypotheses 7 through 16 , but coded as using a random strategy, denoted by '?' in the frequency tables.

Hypothesis 7

Hypothesis 7 predicted that increases in unreliability would be associated with decreases in search depth.

Mean search depth (within dimension) across the five trials was submitted to an ANCOVA, with experimental condition as the between subjects factor and cognitive ability as a covariate. ANCOVA summary results are shown in Table 15.

Table 15. Search Depth by Experimental Condition: ANCOVA Summary

Between Subjects	df	F	p ≤
Cognitive Ability	1	.58	.45
Exp. Condition	3	1.22	.31
Error	121		

No effects were obtained for experimental condition or the covariate. Hypothesis 7 is disconfirmed. Means are presented in Table 16.

Table 16: Search Depth by Experimental Condition

Experimental Condition	Mean Search Depth	SD
Perfect Reliability	1.79	0.697
High Reliability	1.81	0.790
Mixed Reliability	1.99	0.976
Low Reliability	1.61	0.716

Hypothesis 8

Hypothesis 8 predicted that increases in unreliability would be associated with increases in search variability.

Search variability measures were calculated for each of the five trials and averaged into a single index. Search variability for each trial was calculated as the standard deviation of search depth within each dimension (i.e. the standard deviation of search depth across the four dimensions). This index was submitted to an ANCOVA with cognitive ability as the covariate. The ANCOVA summary table is shown in Table 17.

Table 17. Search Variability by Experimental Condition: ANCOVA Summary

Between Subjects	**df**	**F**	**p ≤**
Cognitive Ability	1	2.17	.14
Exp. Condition	3	.77	.52
Error	121		

No significant effects were obtained for either cognitive ability or experimental condition. Hypothesis 8 is disconfirmed. A table of means for this analysis is shown in Table 18.

Table 18: Search Variability by Experimental Condition

Experimental Condition	**Mean Search Variability**	**SD**
Perfect Reliability	2.34	0.610
High Reliability	2.23	0.783
Mixed Reliability	2.07	0.906
Low Reliability	2.12	0.739

Hypothesis 9

Hypothesis 9 predicted that increases in unreliability would be associated with decreases in mean search time.

Time to decision was calculated in the same manner as search variability. The number of seconds elapsing between first search and decision was calculated for each of the five trials. These were averaged to form an overall index. This index was analyzed using an ANCOVA with cognitive ability as the covariate. Results are shown in Table 19.

Table 19. Decision Time by Experimental Condition: ANCOVA Summary

Between Subjects	df	F	p ≤
Cognitive Ability	1	*15.18*	*.0002*
Exp. Condition	3	.77	.51
Error	121		

Significant results were not obtained for experimental condition, disconfirming Hypothesis 9. However, a significant effect was found for cognitive ability. Subjects with higher cognitive ability scores tended to arrive at a decision in less time. A table of means for this analysis is shown in Table 20. Despite the non-significance of results, the pattern of means in conditions where reliability is manipulated is consistent with the hypothesis of decreasing decision times in the presence of increased unreliability.

Table 20: Decision Time by Experimental Condition

Experimental Condition	Mean Decision Time	SD
Perfect Reliability	56.74	18.36
High Reliability	62.26	26.28
Mixed Reliability	56.47	23.28
Low Reliability	54.43	18.60

Hypotheses 10 & 11

Hypotheses 10 and 11 jointly predict that increases in unreliability will be associated with increases in frequency of non-compensatory strategy usage and decreases in frequency of compensatory strategy usage.

Hypotheses involving strategy classification were analyzed using a cross-classification frequency table. Initially, subjects' search matrices were used to make a judgment of the strategy used to reach a decision. Each judgment is indicated in the appropriate cell in Table 21. (ADD=Additive Difference, EBA=Elimination by Aspects, LIN=Linear, ?=uncertain of strategy). The cross-classification statistic for this table is χ^2 (9) = 8.69, $p < .45$, indicating an absence of any relationship between reliability and strategy use. An additional cross-classification statistic was calculated omitting uncertain judgments (the "?"s in the above table).

Table 21: Strategy Classification by Experimental Condition

Strategy	Perfect Rel.	Condition High Rel.	Mixed Rel.	Low Rel.	Row %
ADD	3	8	9	9	23.77
EBA	13	8	7	5	27.05
LIN	1	2	1	1	4.10
?	16	11	13	5	45.08
Column %	27.05	23.77	24.59	24.59	

This resulted in χ^2 (6) = 8.01, p < .23. Data were further classified to generate a frequency table of strategy category (compensatory vs. non-compensatory) by experimental condition. Strategies judged to be linear or additive difference were coded as compensatory. Strategies judged to be EBA (elimination by aspects) were coded as non-compensatory. It should be noted that the non-compensatory class also includes disjunctive, conjunctive, and lexicographic strategies, but none of these strategies were judged to be present in the dataset. Subjects for which a decision strategy could not be determined were not included in this analysis. This table is shown in Table 22.

Table 22: Strategy Category by Experimental Condition

Strategy Category	Perfect Rel.	Condition High Rel.	Mixed Rel.	Low Rel.	Row %
Compensatory	4	10	10	10	50.74
Non-compensatory	13	8	7	5	49.26
Column %	25.37	26.87	25.37	22.39	

For this analysis, χ^2 (3) = 7.17, p < .07. The frequencies noted in the table suggest a steadily decreasing usage of non-compensatory strategies across the experimental conditions, with the least likely usage occurring when reliability is lowest. The pattern for compensatory strategy is not as clear, though compensatory strategy is shown to be least likely in the condition with perfect reliability.

An additional analysis was done collapsing conditions of High, Mixed, and Low reliability and comparing them to the Perfect reliability condition. This resulted in χ^2 (1) = 6.75, p < .01, confirming the difference is primarily between the condition of perfect reliability and the remaining three conditions.

These results disconfirm Hypotheses 10 and 11. In fact, the frequency pattern is opposite that predicted by these hypotheses. Subjects were more likely to use non-compensatory strategies in the perfect reliability context and more likely to use compensatory strategies in the conditions where reliability was manipulated and lower (the conditions of high, mixed, and low reliability).

Hypothesis 12

Hypothesis 12 predicted that subjects would have greater search depths for reliable cues than for unreliable cues.

Analyses proceeded in a similar manner as was true for the cue utilization indices in Hypothesis 1. As was the case for the analysis for Hypothesis 1, only subjects from conditions where reliability was manipulated (the High, Mixed, and Low reliability conditions) were included. In addition, the 22 subjects omitted for previous process tracing analyses (see the beginning of this section for discussion of criteria for omission) were excluded. Mean search depths were calculated for each subject on each dimension.

These values were then assigned a cell based on the dimension (LABEL) and the reliability of the dimension (as it was presented in the MCPL task). An ANOVA was run with dimension label and dimension reliability as between subjects factors. The ANOVA summary table is shown in Table 23.

Table 23. Search Depth by Reliability and Label: ANOVA Summary

Between-Subjects	**df**	**F**	**p ≤**
Reliability	1	.38	.54
Label	3	*6.04*	*.0005*
Reliability x Label	3	.65	.58
Error	348		

Results indicate no effect of reliability on search depth, disconfirming hypothesis 12. Post hoc tests among dimension labels were conducted due to the main effect observed for label. A significant effect is evidenced by subjects tending to search most in Barometric Pressure and Humidity, followed by Wind Speed and Cloud Ceiling. Cell means for this analysis are shown in Table 24. Cell standard deviations are noted in parentheses.

Table 24. Search Depth by Reliability and Label

			Label			
		Wind Spd.	**Bar. Press.**	**Humidity**	**Cld. Ceil.**	**Mean**
Reliability	**High**	1.40	2.56	2.18	1.31	*1.86*
		(1.71)	(1.89)	(2.04)	(1.77)	
	Low	1.67	2.18	1.82	1.30	*1.74*
		(1.80)	(1.98)	(1.54)	(1.85)	
	Mean	1.53_{BC}	2.37_{A}	2.00_{AB}	1.30_{C}	

Note: Subscripts denote means significantly different at the .05 level or better.

Hypotheses 13 & 14

Hypothesis 13 and 14 jointly predict relative weighting relationships. It was predicted that unreliable dimensions presented with reliable dimensions will have lower search depths than when presented with other unreliable dimensions. In addition, reliable dimensions presented with unreliable dimensions will have higher search depths than when presented with other reliable dimensions. Note that the terms "reliable dimension" and "unreliable dimension" are not meant to imply reliability is manipulated in the process tracing task. These labels are meant to denote how the dimension was manipulated as a cue in the MCPL task. A cue low in reliability essentially becomes an "unreliable dimension" in the discussions below.

These hypotheses concerning relative weighting relationships were analyzed using a 2 (reliable dimension vs. unreliable dimension) x 2 (presented with reliable dimension vs. presented with unreliable dimension) ANOVA design. Assignments of dimensions were made using the following logic: 1) all searches on all dimensions in the high reliability condition were coded as reliable dimension (1) presented with reliable dimension (1); 2) all searches on all dimensions in the low reliability condition were coded as unreliable dimension (2) presented with unreliable dimension (2); 3) searches in reliable dimensions in the mixed reliability condition were coded as reliable dimensions (1) presented with unreliable dimensions (2); 4) searches in unreliable dimensions in the mixed reliability condition were coded as unreliable dimensions (2) presented with reliable dimensions (1). These four classification rules result in a 2 (reliability) x 2 (presentation context) table. ANOVA was used to test the effects of these two factors on search depth in a dimension. Results are shown in Table 25.

Table 25. Search Depth by Reliability and Context: ANOVA Summary

Between Subjects	df	F	p ≤
Reliability	1	.36	.55
Presentation Context	1	.39	.53
Rel. * Pres. Context	1	1.84	.18
Error	352		

Cell means and standard deviations (in parentheses) are shown in Table 26. The analysis disconfirms Hypotheses 13 and 14. No significant differences were found for cues of different reliability or presentation context.

Table 26. Search Depth by Reliability and Context

	Presented with:		
Reliability	Reliable	Unreliable	Mean
Reliable	1.81 (1.90)	1.96 (1.94)	*1.86*
Unreliable	2.02 (1.90)	1.61 (1.77)	*1.74*
Mean	*1.88*	*1.72*	

Hypotheses 15 & 16

Hypothesis 15 and 16 predicted a relative weighting relationship identical to Hypotheses 13 and 14, except cue utilization indices were used as a usage/weighting indicator instead of search depth. Assignments to the Reliability and Presentation

Context factors were identical to those used in evaluating Hypotheses 13 and 14. Specifically, it was predicted that unreliable cues presented with reliable cues will have lower cue utilization indices than when presented with other unreliable cues. In addition, reliable cues presented with unreliable cues will have higher cue utilization indices than when presented with other reliable cues.

Hypotheses 15 and 16 were analyzed exactly as Hypotheses 13 and 14, except cue utilization indices were used instead of search depths and the blocking method described in the MCPL results was used. Repeated measures ANOVA results are shown in Table 27.

Table 27. Cue Utilization by Reliability and Context: ANOVA Summary

Between-Subjects	df	F	p ≤
Reliability	1	.07	.79
Context	1	.001	.97
Reliability x Context	1	.98	.32
Error	428		
Within-Subjects			
Block	2,427	2.06	.13
Block x Reliability	2,427	.94	.39
Block x Context	2,427	1.97	.14
Block x Rel. x Context	2,427	.12	.89

Results are nonsignificant, disconfirming Hypotheses 15 and 16. Cell means and standard deviations for this analysis are shown in Table 28.

Table 28. Cue Utilization by Reliability, Context, and Block

Context	Reliability	Block 1	Block 2	Block 3	Mean
with Reliable	**Reliable**	.239	.244	.251	.245
		(.240)	(.201)	(.218)	
	Unreliable	.230	.217	.217	.222
		(.277)	(.210)	(.220)	
with Unreliable	**Reliable**	.191	.225	.263	.226
		(.221)	(.190)	(.231)	
	Unreliable	.233	.224	.260	.239
		(.245)	(.232)	(.255)	

Note: Values in parentheses represent cell standard deviations

DISCUSSION

For the purposes of discussion, results will be divided into three sets, a reclassification of the five sets presented earlier: 1) judgment (MCPL task and its associated dependent measures), 2) decision-making (process tracing task and its associated dependent measures), and 3) relative-weighting hypotheses (involving measures from both MCPL and process tracing tasks). Each section will include a summary of findings for stated hypotheses and potential explanation of these findings and their relationship to relevant published research.

Hypothesis Set 1: Effects of Unreliability on MCPL Measures

Summary of Findings

Hypothesis 1 through Hypothesis 4 predicted that the unreliability present in a judgment context would affect the judgment, and that the unreliability associated with a cue would affect the utilization of that cue in reaching a judgment. Hypothesis 1 was disconfirmed. Subjects did not use cues high in reliability significantly more than they did cues low in reliability. Subjects' achievement scores did increase as overall reliability increased, with the most striking differences between subjects in the perfect reliability condition and the low reliability condition, confirming Hypothesis 2. Hypotheses 3 and 4 addressed the effects of unreliability on knowledge (G) and consistency (R_s) scores, respectively. Hypothesis 3 was disconfirmed and Hypothesis 4 was confirmed. Subjects displayed no significant differences in knowledge indices across the four conditions, but did show significant differences in consistency, with a pattern similar to that of the previously noted achievement scores. These results suggest the differences in achievement scores noted above were due primarily to differences in consistency. Hypothesis 5, essentially a manipulation check, was confirmed, indicating that, at the very least, subjects did perceive a difference in variability between cues differing in reliability. Hypothesis 6 predicted subjects would report greater weight being given to reliable cues. This was disconfirmed. However, subjects did report applying differential weight to certain cues.

Potential Explanation of Findings

Interpreted alone, the significant effect of unreliability on achievement scores is evidence that the presence of unreliability did have detrimental effects on overall accuracy of prediction. The differential effects on knowledge and consistency indices offers some potential explanation for the nature of the detrimental effect. The fact that no effect was observed on knowledge indices suggests that the presence of unreliability did not affect their ability to gain knowledge of the task structure. Given this fact, and that the task itself was completely determined (R^2=1.0), leaves the decreased consistency as the only measurable impact of unreliability (except for achievement, which is directly a function of consistency). When one recalls that decreased consistency measures the multiple R between the four cues and the judgment, it becomes apparent that, in the presence of unreliability, subjects were either using the four cues in an inconsistent manner or were erroneously estimating the true values of the four cues. There was some evidence supporting the former possibility in Graph 1, where it was noted that subjects seemed to have shifted weight from Cloud Ceiling to Wind Speed across the three trial blocks. In support of the latter possibility, it is noteworthy that the effects of unreliability on consistency seemed particularly damaging in Conditions 3 (Mixed Reliability) and 4 (Low Reliability), the only two conditions in which low reliability cues (with a SD of 6) were presented. This suggests that in Conditions 1 (Perfect Reliability) and 2 (High Reliability), the higher consistency may have been due to the easier task of estimating the cues' true values.

In addition to results for stated hypotheses, certain results from the MCPL analyses are worthy of further discussion. Cognitive ability had significant impact on all three of the primary MCPL performance indices, supporting the general reasoning for inclusion of cognitive ability as a covariate. No evidence was found for differential effects of cognitive ability across experimental condition. The relationship between cognitive ability and performance in the MCPL task has several possible explanations. Two notable possibilities emerge from the previously mentioned information integration theory of Anderson (1981), and the information processing paradigm of cognition, generally credited to the work of Atkinson & Shiffrin (1968). Explained from the paradigm of information integration theory, individuals of higher cognitive ability may have simply been better able to accurately estimate the mean value of a meter, and thus, were better able to judge the criterion. Taking an information processing perspective would involve the assumption that individuals of higher cognitive ability had more cognitive resources at their disposal, and were thus overall able to estimate, store, and combine information more effectively, especially in the face of unreliability. The data collected for this study do not allow much conjecture as to the reasons for the effect of cognitive ability, though the lack of differential effect across experimental condition tends to refute an information processing interpretation. If such an interpretation were accurate, one would expect the effect of cognitive ability to be strongest in the low reliability situation, where the cognitive demands are highest.

An additional finding is worthy of note. Despite the lack of significant differences in cue utilization between reliable and unreliable cues (the disconfirmation of Hypothesis 1), there was a significant Reliability by Label interaction for cue utilization. For three of the four cues (Barometric Pressure, Humidity, and Cloud Ceiling), subjects tended to weight the cue more when it was reliable, consistent with the stated hypothesis. However, for Wind Speed, subjects tended to weight the cue more when it was unreliable. While no predictions were made regarding such weighting patterns, there is a plausible reason for the result. Although subjects' perceptions of "face reliability" were not assessed for this study, it is possible that subjects felt the meters were differentially susceptible to environmental disturbance. Of the weather instruments measuring the four variables in this task, it would appear, at least to this author, that an instrument measuring wind speed would be subject to far more environmental disturbance (e.g. gusts, downdrafts, etc.) than any of the other three meters. Such variability could, in fact, be interpreted as true score fluctuations rather than measurement error of the instrument. In such a case, subjects may have used the meter's variability as a cue rather than estimate its value. A highly variable meter for wind speed may have suggested gusts and stormlike conditions, causing subjects to give it greater consideration when they estimated snowfall.

Despite the fact that label effects were not the focus of this study, results suggest that labeling did have significant influence on subjects' judgment policies. This is not surprising, as label characteristics have been shown to have significant, and often unpredictable, results on MCPL performance. The role of pre-task information (termed "feed-forward" information by some MCPL researchers) such as labeling characteristics have been shown to have unpredictable, and in some cases, paradoxical (Koele, 1980) effects on MCPL performance indices.

Many of the results noted above provide for a useful reinterpretation of the previous study conducted by York, Doherty, and Kamouri (1987). The effects of unreliability on cue utilization was similar for both studies: there were none. However, York et al found that subjects did report giving greater weight to the more reliable cue. This study found no such pattern. In fact, as previously noted, the pattern of subjective weights was remarkably similar to the pattern of actual weights, suggesting that subjects were aware of which cues they were using. Though data do not point to a clear reason for this finding, a procedural difference between this study and the York et al study may be relevant. In the interests of having subjects reach a stable judgment policy, York et al did not give subjects outcome feedback during the test trials in their study. This study did give subjects outcome feedback throughout the test trials. This was done primarily to allow subjects to experiment with different weighting strategies as their saw their estimates deviate from the true scores. It was feared that without outcome feedback, subjects would not be able to accurately estimate the effectiveness of their judgment policies.

As noted earlier in this document, it was also thought that the York et al. study may have not revealed any effects of unreliability on MCPL performance due to the overall simplicity of the two-cue task structure. The decreases in MCPL performance in the presence of unreliability noted in this study, which used a four-cue task, lend support to this contention. York et al (1987) also note the consistency of their findings with the work of Anderson's (1981) information integration theory. While the estimation heuristics used by the subjects in the present study are beyond detection given the data collected, findings are not inconsistent with Anderson's theory. Subjects could have indeed been *attempting* to estimate the average of many "bouncing" observations of a meter, but unable to do so accurately.

Hypothesis Set 2: Effects of Unreliability on Process Tracing Measures

Summary of Findings

Hypotheses 7 through 9 involved predictions about the effects of unreliability on dependent measures in the process tracing task. Hypotheses 7, 8, and 9 predicted that subjects would search less information, exhibit a more variable search pattern, and reach a decision in less time, in conditions with lower reliability. All three hypotheses were disconfirmed. However, cognitive ability did have a significant effect on decision time, with subjects of higher cognitive ability tending to reach a decision in less time.

Hypotheses 10 and 11 were designed to assess the effect of unreliability on strategy utilization in the process tracing task. A contingency table analysis was used to simultaneously assess these hypotheses, resulting in a disconfirmation of both. However, additional tests on reduced tables indicated subjects were more likely to use non-compensatory strategies in the perfect reliability condition, and more likely to use compensatory strategies in the remaining three (high, mixed, and low reliability) conditions. These results were significant, and in a direction opposite to that predicted.

Hypothesis 12 essentially repeated the cue utilization analysis of the MCPL task, but used the search depth measure from the process tracing task as a dependent variable. No significant search depth differences between reliable and unreliable dimensions were found, disconfirming the hypothesis. However, a main effect was observed for dimension label, indicating subjects tended to search Barometric Pressure and Humidity more than Wind Speed and Cloud Ceiling.

Potential Explanation of Findings

Reliability did not have a significant impact on any of the search indicators in the process tracing trials. Although the data do not allow much conjecture as to the reason for this lack of effect, it is possible that the small data matrix (at least compared to most process tracing research) may have made it difficult for participants to display (or coders to recognize) some of the complex search strategies. The significant effect on decision time for cognitive ability is likely explained by subjects with higher cognitive ability learning or understanding the task quickly and proceeding with their search strategy.

The effects of unreliability on search strategy selection, essentially the focal point of the process tracing investigation, are worthy of more discussion. In the opinion of this author, there are several potential explanations for the pattern of data, at least two of which are noteworthy. Given the large body of process tracing research which revealed increased non-compensatory strategy usage as task complexity increased (Payne & Braunstein, 1978; Klayman, 1985; Onken, Hastie, & Revelle, 1985), as well as the consistency of such findings with the Beach & Mitchell (1978) contingency theory, it is plausible to conclude that subjects perceived less complexity in the conditions where reliability was manipulated (high, mixed, and low), than in the perfect reliability condition. If one pursues this explanation, one must somehow explain why four cues presented with no variability were deemed more complex than four cues presented with variability. This author believes a second explanation holds more promise.

This explanation is based on a re-examination of the assumed effects of unreliability on task perceptions. Previous process tracing research has examined the effects of a wide array of task attributes; number of dimensions (Capon & Burke, 1980), number of alternatives (Johnson & Meyer, 1984; Johnson, Meyer, & Ghore, 1986), and dimensional redundancy (Gilliland & Schmitt, 1993), to name a few. As noted earlier, research in each of these areas generated results generally consistent with the Beach & Mitchell (1978) contingency model (even if not explicitly theoretically based on it), namely that each of these attributes manifest effects on strategy selection consistent with the notion that the attribute(s) increased increasing task complexity. However, if we, for the moment, reject the notion that unreliability increases or decreases task complexity, the results obtained by this study have a somewhat different interpretation. Ironically, it involves the concept that initiated this research - test-retest reliability, and has its roots in a re-examination of MCPL results.

First, judging from the previously discussed knowledge indices, we can safely conclude that most subjects were reasonably aware of the functional relationship between the four cues and the criterion after the MCPL trials were completed (i.e. that all four cues were important to predicting snowfall). This should not be surprising, as the positive relationship between all cues and the criterion is apparent when viewing the meters, even when disturbed by error. Given this pattern, subjects may have perceived a correlation between the cues, when, in fact, they were completely uncorrelated. If this notion transferred to the process tracing trials, and a subject believed two dimensions in the process tracing trials were highly correlated, then accessing information in both dimensions for one alternative amounts to a test-retest reliability check. In the condition of perfect reliability, no information is given about reliability, and it is assumed to be 1.0. Hence, a non-compensatory strategy such as elimination by aspects could be used without concern about the reliability of the aspects of choice. For example, subjects would have no reason to believe that a Wind Speed value of 15 was not truly 15. In the three conditions where dimensions had reliability attributes, the situation may have been quite different. Elimination by aspects or any other non-compensatory strategy which relies on sequential or hierarchical single-dimension based judgments, is no longer a safe course to select, since any values for an aspect deemed important may not be the true values. If subjects in these conditions did perceive some intercorrelation between dimensions, either of the compensatory strategies (linear or additive difference) could be used to assess a subset of dimensions for all alternatives, or all dimensions for a subset of alternatives, to gain a degree of confidence in a decision. Either would result in a set of decision alternatives for which the same multiple dimensions are assessed.

This explanation is not novel, and suggests subjects may have acted in a manner similar to modern structural equation researchers and industrial psychologists. Such individuals usually include multiple indicators of a predictor or criterion construct in order to reduce the effects of unreliability in any one of the indicators. In a similar fashion, subjects may have been conducting a naive check of test-retest reliability by comparing information from two dimensions thought to be correlated.

I now turn to the main effect observed for Label in the analysis for Hypothesis 12. Subjects tended to access information about Humidity and Barometric Pressure more than Wind Speed or Cloud Ceiling. A comparable effect was not observed for cue utilization in the MCPL trials, so it does not appear to be the result of a transferred judgment policy. The only explanation I can offer for this result is that humidity and barometric pressure are variables more commonly attributed to weather than are wind speed or cloud ceiling. When subjects were asked to seek information rather than being presented with it, there may have been a tendency to look at information from a familiar weather source, such as humidity or barometric pressure.

Hypothesis Set 3: Effects of Unreliability - Relative Weighting Hypotheses

Summary of Findings

Hypotheses 13 through 16 predicted differential usage of information would occur based on both the information's reliability and the reliability of other available information. Hypotheses 13 and 14 predicted that reliable dimensions would be searched more when presented with unreliable dimensions than when presented with other reliable dimensions, and that unreliable dimensions would be searched more when presented with unreliable dimensions than when presented with reliable dimensions. These hypotheses were disconfirmed. Hypotheses 15 and 16 predicted that reliable cues would be used more when presented with unreliable cues than when presented with other reliable cues, and that unreliable cues would be used more when presented with unreliable cues than when presented with reliable cues. These hypotheses were also disconfirmed.

Potential Explanation of Findings

The results obtained do not support the hypotheses based on findings from the source credibility literature. This does not cast debate onto the source credibility literature, but rather casts doubts upon the assumption that unreliability effects judgments and decisions in a manner analagous to source non-credibility. Unreliability offers only a quantitative reason to see a source as non-credible, whereas the source credibility construct likely is composed of many other non-quantitative facets.

Conclusions and Directions for Future Research

In general, the results of the MCPL analysis support the overall contention that unreliable information has effects on judgment. Subjects evidenced sufficient knowledge (G) to suggest awareness of the task structure, but, in situations where reliability was low, showed a lack of consistent usage of their task knowledge in reaching a judgment. The interpretation of this finding highlights an ambiguity in the modified MCPL model. It was thought that subjects, when faced with a "jittery cue", would form an internal estimate of the cue's true value, and possibly use it in reaching a judgment. High achievement scores and consistency indices in conditions of perfect and high reliability suggest subjects were making their internal estimates and judgments with reasonable accuracy. However, the lower consistency (and resulting achievement) scores in conditions of mixed and low reliability offer two possibilities: 1) subjects are making accurate estimates, but changing their cue usage; 2) subjects are attempting to use a set of cues, but with differential success in estimation. The difference in these possibilities is fundamental, as the former involves volitional shifts in policy, while the latter suggests the reason lies solely with difficulties in the perceptual system.

The MCPL portion of this study also provided some useful avenues for future research. The notion of "face reliability" was invoked earlier, in reference to subjects' initial perceptions of the reliability of an instrument or source of information. Future research should examine the potential effects of this concept. The differential effects of perceived vs. actual attributes of information is not a new concept, and has been previously addressed in the decision-making literature by Gilliland & Schmitt (1993), who assessed the differential effects of perceived and actual redundancy of information labels on decision making. Reliability offers a more complex attribute, as there are at least three different definitions: 1) the subject's estimate of reliability based on a cue label ("face" reliability); 2) the subject's estimate of a cue's reliability (perceived reliability/variability); and 3) the cue's actual reliability/variability. A closer look at all three, especially the first, may prove fruitful.

The effects of unreliability on decision making were not as clear. The results of the process tracing analyses were mixed. Despite the disconfirmation of all hypotheses, the significant results interpreted earlier offer insight into the nature of the effects of unreliability on decision making. Specifically, framing unreliability within the Beach and Mitchell contingency model as a source of increased task complexity may have overly simplified its effects. The inconsistency of results with a rather robust finding in the process tracing literature confirms this idea, suggesting that unreliability may have its own unique effects on decision making processes. Future research should examine the effects of reliability on decision making from a variety of theoretic perspectives, with attention to the potential role of pre-decision notions of dimensional characteristics (such as intercorrelation).

The purported role of pre-task perceptions and assumptions was highlighted for both judgment and decision making results. The role of preconceived notions of reliability should be a major focus of future research on the impacts of unreliability.

While the obtained results did not provide evidence which clearly validated or invalidated the use of Brunswik's lens model or the contingency theory of decision making (Beach & Mitchell, 1978) as the conceptual bases for the study, they do offer some insight into the future use of either of these frameworks in future research on the effects of unreliability.

First, the modified lens model proposed by York, Doherty, & Kamouri (1987), appears to be a viable method for studying the effects of unreliability on judgment. Much as this study attempted to expand on limitations in the York et al study, future research using a lens model framework may be able to use the current study as a point of expansion.

Second, and noted earlier, the results are not as favorable for the contingency model of decision making. These results are not interpreted as invalidating the contingency model, but rather, as evidence that the effects of unreliability may be more than simply increases in task complexity.

Limitations

This study made several important assumptions about the nature of information in judgment and decision-making tasks, which may limit the generalizability of the findings. A primary limitation to the study is that the tasks allowed the non-consideration of information. Thus, subjects could make reasonably accurate judgments and decisions if they completely ignored a cue/dimension. This potentially affects generalizability. For instance, it is reasonable to expect a psychometrician or statistician to disregard information from an unreliable measure or test, basing their estimates or decisions on other available measures. However, certain information, like the pilot's altimeter, is vital to high performance, or even survival, in the judgment/decision environment, and ignoring such information would be disastrous. In one instance, the judge or decision maker is afforded the possibility of using surrogate information. In the other, he or she is forced to estimate the unreliable information. Given the lack of information about cue importance or availability of surrogate information, subjects in this study could have simply attempted to estimate all unreliable cues.

Limitations to generalizability are also present in the judgment task. Findings cannot be confidently generalized beyond the scope of this study . The decision task presented in this study assumed information was linearly related to the optimal decision and that all information was compensatory or additive in nature. Judgments and decisions made in the real world, particularly those most likely to be adversely affected by unreliability, may be highly non-linear or non-compensatory in functional form.

Another important limitation is the nature of the dual tasks presented to the subjects. The marriage of MCPL and process tracing tasks in a single judgment/decision making situation was not a trivial feat, and, in fact, may have created difficulties in comparing these results to other results in either domain. A four cue MCPL task is not problematic (though four cues are near the high end for MCPL tasks) but its corresponding process tracing task was severely limited. The majority of process tracing studies use decision matrices much larger than the 4 dimension by 6 alternative matrix used in this study, tending more toward 8 by 8 and 8 by 12 matrices. Not only does this weaken potential comparisons to other process tracing work, but also caused certain difficulties for judges trying to distinguish the various decision making strategies. Some of these problems could have been alleviated by expanding the MCPL task to include more cues, potentially risking the generalizability of MCPL findings. Future researchers attempting to design an experiment with both MCPL and process tracing components should note this inherent tension between the two research methodologies.

APPENDICES

APPENDIX A: STRATEGY PROTOTYPES

The following represent sample decision matrix patterns which were designed in training of strategy raters as well as used by the raters while making strategy classifications. Numbers in cells refer to progressive search through the matrix. Search begins at the numeral "1" and terminates with the highest integer value.

LINEAR STRATEGY

	Alternatives					
	A	B	C	D	E	F
Dimension 1	1	4	7	10	13	16
Dimension 2	2	5	8	11	14	17
Dimension 3	3	6	9	12	15	18
Dimension 4						

ADDITIVE DIFFERENCE STRATEGY

	Alternatives					
	A	B	C	D	E	F
Dimension 1						
Dimension 2	1	2	3	4	5	6
Dimension 3	7	8	9	10	11	12
Dimension 4						

CONJUNCTIVE STRATEGY

	Alternatives					
	A	B	C	D	E	F
Dimension 1	1	4	8	10	11	14
Dimension 2	2	5	9		12	15
Dimension 3	3	6			13	16
Dimension 4		7				17

DISJUNCTIVE STRATEGY

	Alternatives					
	A	B	C	D	E	F
Dimension 1	10	11	1	12	13	14
Dimension 2	5	6	2	7	8	9
Dimension 3			3			
Dimension 4	15	16	4	17	19	18

LEXICOGRAPHIC STRATEGY

	Alternatives					
	A	B	C	D	E	F
Dimension 1						
Dimension 2	1	2	3	4	5	6
Dimension 3						
Dimension 4						

ELIMINATION BY ASPECTS STRATEGY

	Alternatives					
	A	B	C	D	E	F
Dimension 1	1	2	3	4	5	6
Dimension 2	7	8		9	10	
Dimension 3		11		12	13	
Dimension 4		14			15	

EXAMPLE OF EBA-AD CROSSCLASSIFIED PATTERN

	Alternatives					
	A	B	C	D	E	F
Dimension 1	1	2	3	4	5	6
Dimension 2	7	8	9	10	11	12
Dimension 3		13				
Dimension 4						

APPENDIX B: MCPL ITEMS

a. Subjective variability assessment items:

1. "To what extent do you feel the meter for WIND SPEED jumped around while you were
watching it?"

1. Not at all. The meter did not move.
2. it jumped around in about a 5 point range
3. it jumped around in about a 10 point range
4. it jumped around in about a 15 point range
5. it jumped around in about a 20 point range
6. it jumped around in about a 25 point range

2. "To what extent do you feel the meter for BAROMETRIC PRESSURE jumped around while you were watching it?"

1. Not at all. The meter did not move.
2. it jumped around in about a 5 point range
3. it jumped around in about a 10 point range
4. it jumped around in about a 15 point range
5. it jumped around in about a 20 point range
6. it jumped around in about a 25 point range

3. "To what extent do you feel the meter for HUMIDITY jumped around while you were watching it?"

1. Not at all. The meter did not move.
2. it jumped around in about a 5 point range
3. it jumped around in about a 10 point range
4. it jumped around in about a 15 point range
5. it jumped around in about a 20 point range
6. it jumped around in about a 25 point range

4. "To what extent do you feel the meter for CLOUD CEILING jumped around while you were
watching it?"

1. Not at all. The meter did not move.
2. it jumped around in about a 5 point range
3. it jumped around in about a 10 point range
4. it jumped around in about a 15 point range
5. it jumped around in about a 20 point range
6. it jumped around in about a 25 point range

b. subjective importance judgment item:

Please allocate the total 100 points in proportion to how important you felt each meter was to predicting snowfall:

Example: This person felt all four were equally important to predicting snowfall:

Wind Speed	25
Barometric Pressure	25
Humidity	25
Cloud Ceiling	25

Example: This person felt Wind Speed was most important, followed in order by Cloud Ceiling, Humidity, and Barometric Pressure:

Wind Speed	50
Barometric Pressure	5
Humidity	15
Cloud Ceiling	30

APPENDIX C: MCPL/PROCESS TRACING STIMULI

a. MCPL Stimuli

Trial	*Wind Speed*	*Bar. Press.*	*Humidity*	*Cloud. Ceil.*	*Criterion(Snowfall)*
1	20	12	33	20	20
2	21	19	17	36	27
3	23	29	14	15	23
4	28	9	33	11	20
5	24	14	17	22	19
6	27	16	2	7	7
7	23	26	28	20	27
8	22	29	19	30	30
9	11	17	17	21	14
10	22	24	20	24	24
11	12	21	18	15	14
12	33	16	28	13	26
13	12	31	20	20	23
14	25	19	30	28	29
15	5	19	23	13	11
16	5	3	21	20	6
17	27	13	21	21	23
18	13	28	12	29	19
19	23	15	9	20	14
20	26	22	17	19	22
21	9	21	16	19	10
22	26	26	13	26	28
23	20	10	12	29	17
24	17	27	11	18	15
25	28	18	19	34	28

(continued next page)

26	20	13	19	20	14
27	20	13	24	17	19
28	7	23	27	21	17
29	29	27	31	10	29
30	27	33	26	20	31
31	31	20	14	16	22
32	27	16	14	18	16
33	12	28	21	22	22
34	16	20	18	32	22
35	13	15	24	29	21
36	22	8	23	6	7
37	22	18	30	23	29
38	15	26	14	12	14
39	23	30	19	22	27
40	28	11	22	31	26
41	23	26	26	14	26
42	31	18	12	23	21
43	23	30	19	17	25
44	15	19	13	13	8
45	18	10	7	15	4
46	12	19	29	32	28
47	23	31	24	13	24
48	17	15	21	19	18
49	13	22	34	13	21
50	11	25	16	10	13

b. Process Tracing Stimuli

	Route 1	Route 2	Route 3	Route 4	Route 5	Route 6
Trial 1						
WS	9	13	12	22	33	24
BP	21	15	21	29	16	14
HM	16	24	18	19	28	17
CC	19	29	15	30	13	22
Trial 2						
WS	25	20	23	20	15	11
BP	19	10	30	12	19	17
HM	30	12	19	33	13	17
CC	28	29	17	20	13	21
Trial 3						
WS	27	28	11	5	29	23
BP	16	11	25	19	27	29
HM	14	22	16	23	31	14
CC	18	31	10	13	10	15
Trial 4						
WS	22	26	27	23	17	17
BP	8	22	33	26	15	27
HM	23	17	26	26	21	11
CC	6	19	20	14	19	18
Trial 5						
WS	23	7	31	22	23	18
BP	31	23	20	18	15	10
HM	24	27	14	30	9	7
CC	13	21	16	23	20	15

LIST OF REFERENCES

LIST OF REFERENCES

Anderson, N.H. (1969). Comment on "An analysis-of-variance model for the assessment of configural cue utilization in clinical judgment." *Psychological Bulletin*, **72**, 63-65.

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

Atkinson, R.C., & Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In K.W. Spence & J.T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory.* (Vol. 2, pp. 89-195). New York: Academic Press.

Beach, L.R., Mitchell, T.R., Deaton, M.D., & Prothero, J. (1978). Information relevance, content, and source credibility in the revision of opinions. *Organizational Behavior and Human Performance*, **21**, 1-16.

Beach, L.T., & Mitchell, T.R. (1978). A contingency model for selection of decision strategies. *Academy of Management Review*, **3**, 439-449.

Biggs, S.F., Bedard, J.C., Gaber, B.G., & Linsmeier, T.J. (1985). The effects of task size and similarity on the decision behavior of bank loan officers. *Management Science*, **31**, 970-987.

Birnbaum, M.H., Wong, R., & Wong, L.K. (1976). Combining information from sources that vary in credibility. *Memory and Cognition*, **4**, 330-336.

Borko, H., & Shavelson, R.J. (1978). Teacher's sensitivity to the reliability of information in making causal attributions in an achievement situation. *Journal of Educational Psychology*, **70**, 271-279.

Brehmer, B. (1970). Inference behavior in a situation where the cues are not reliably perceived. *Organizational Behavior and Human Performance*, **5**, 330-347.

Brunswik, E. (1952). *The conceptual framework of psychology.* Chicago: The University of Chicago Press.

Capon, N., & Burke, M. (1980). Individual, product class, and task-related factors in consumer information processing. *Journal of Consumer Research*, **7**, 314-326.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Ford, J.K., Schmitt, N., Schechtman, S.L., Hults, B.M., & Doherty, M.L. (1989). Process tracing methods: Contributions, problems, and neglected research questions. *Organizational Behavior and Human Decision Processes*, **43**, 75-117.

Frisch, D. (1988). Ambiguity and rationality. *Journal of Behavioral Decision Making*, **1**, 149-157.

Gilliland, S.W., & Schmitt, N. (1993). Information redundancy and decision behavior: A process tracing investigation. *Organizational Behavior and Human Decision Processes*, **54**, 157-180.

Gilliland, S.W., Schmitt, N., & Wood, L. (1993). Cost-benefit determinants of decision process and accuracy. *Organizational Behavior and Human Decision Processes*, **56**, 308-330.

Gilliland, S.W., Wood, L., & Schmitt, N. (1994). The effects of alternative labels on decision behavior: The case of corporate site selection decisions. *Organizational Behavior and Human Decision Processes*, **58**, 406-427.

Hammond, K.R., & Summers, D.A. (1965). Cognitive dependence on linear and nonlinear cues. *Psychological Review*, **72**, 215-224.

Hammond, K.R., & Summers, D.A. (1972). Cognitive control. *Psychological Review*, **79**, 58-67.

Johnson, E.J., & Meyer, R.J. (1984). Compensatory choice models of noncompensatory processes: The effect of varying context. *Journal of Consumer Research*, **11**, 528-541.

Johnson, E.J., Meyer, R.J., & Ghore, S. (1986). *When choice models fail: Compensatory representations in efficient sets.* Unpublished manuscript, Carnegie-Mellon University.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, **80**, 237-251.

Klayman, J. (1985). Children's decision strategies and their adaption to task characteristics. *Organizational Behavior and Human Decision Processes,* **35**, 179-201.

Knowles, B.A., Hammond, K.R., Stewart, T.R., & Summers, D.A. (1971). Positive and negative redundancy in multiple cue probability tasks. *Journal of Experimental Psychology*, **90**, 157-159.

Levin, I.P., Ims, J.R., Vilman, J.A. (1980). Information variability and reliability effects in evaluating student performance. *Journal of Educational Psychology*, **72**, 355-361.

Newell, A., & Simon, H.A., (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.

Olshavsky, R.W. (1979). Task complexity and contingent processing in decision making: A replication and extension. *Organizational Behavior and Human Performance*, **24**, 300-316.

Onken, J., Hastie, R., & Revelle, W. (1985). Individual differences in the use of simplification strategies in a complex decision-making task. *Journal of Experimental Psychology: Human Perception and Performance*, **11**, 14-27.

Payne, J.W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance*, **16**, 366-387.

Payne, J.W., & Braunstein, M.L. (1978). Risky choice: An examination of information acquisition behavior. *Memory and Cognition*, **6**, 554-561.

Payne, J.W., Bettman, J.R., & Johnson, E.J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 534-552.

Rosenbaum, M.E., & Levin, I.P. (1968). Impression formation as a function of source credibility and order of presentation of contradictory information. *Journal of Personality and Social Psychology*, **10**, 167-174.

Russo, J.E., & Dosher, B.A. (1983). Strategies for multi-attribute binary choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **9**, 676-696.

Schum, D.A. (1975). The weighing of testimony in judicial proceedings from sources having reduced credibility. *Human Factors*, **17**, 172-182.

Slovic, P. (1969). Analyzing the expert judge: A descriptive study of a stockbroker's decision processes. *Journal of Applied Psychology*, **53**, 255-263.

Slovic, P., Fischhoff, B., & Lichtestein, S. (1977). Behavioral decision theory. *Annual*

Review of Psychology, **28**, 1-39.

Surber, C.F. (1981). Effects of information unreliability in predicting task performance using ability and effort. *Journal of Personality and Social Psychology*, **40**, 977-989.

Surber, C.F. (1984). Inferences of ability and effort: Evidence for two different processes. *Journal of Personality and Social Psychology*, **46**, 249-268.

Svenson, O. (1979). Process descriptions of decision making. *Organizational Behavior and Human Performance*, **23**, 86-112.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, **79**, 281-299.

Winer, B.J. (1971). *Statistical Principles in experimental design.* New York: McGraw-Hill.

York, K.M., Doherty, M.E., & Kamouri, J. (1987). The influence of cue unreliability on judgment in a multiple cue probability learning task. *Organizational Behavior and Human Decision Processes*, **39**, 303-317.