

HUMAN INFORMATION PROCESSING RESEARCH IN ACCOUNTING: THE STATE OF THE ART IN 1982

ROBERT LIBBY

University of Michigan

and

BARRY L. LEWIS

University of Pittsburgh

Abstract

Awareness of the importance of human information processing research to accounting issues has increased dramatically since 1977. As a result, this literature has expanded in volume and addresses a larger spectrum of accounting problems. Further, it incorporates a wider variety of theories and methodologies. This paper draws upon the framework provided by Libby and Lewis (1977) to synthesize and evaluate accounting research conducted since 1977 using the lens model, probabilistic judgment, predecisional behavior, and cognitive style approaches. In addition, the impact of the research on practice and some directions for future research are discussed.

Along with the recognition that decision making is the focal point of the current practice of accounting, an extensive body of research which analyzes decision making in accounting settings has been developed. This research is usually referred to as *human information processing (HIP)* or *behavioral decision making* research. Accountants have shown particular interest in studies which investigate (1) the role of accounting information in user decisions (e.g. in commercial lending) and (2) the complex decisions required in the practice of accounting (e.g. in auditing). The evidence generated by this research serves a dual purpose. First, it may lead to *improvements* in these accounting decisions. Second, it can add to the basic knowledge of human decision processes.

Four years ago, we provided a review of what was then an emerging research program (Libby &

Lewis, 1977; LL-77 hereafter). Since this initial review, both the interest and research output in this area have grown at an increasing rate. In response to this activity, we have compiled a second state-of-the-art paper. As testimony to the growth in interest we note that this second review contains more than twice as many studies as did LL-77.

As further evidence of the interest in human information processing research in accounting, we can point to the impact which such research has already had on accounting practice. Accepting consensus and consistency as measures of the quality of expert judgment, many audit firms have developed decision aids to increase the consistency of judgments. These notions derive directly from research in behavioral decision theory. Research exposing humans as poor intuitive statisti-

* The authors gratefully acknowledge the Paton Accounting Center for financial support and Garry Marchant for his assistance.

cians has accelerated the application of statistical sampling and regression analysis in auditing. Other decision aids have been developed to help overcome common heuristics which may lead to biased evaluations of audit evidence. We shall discuss the impact of research on practice more fully in the concluding section of the paper.

LL-77 was organized on the basis of a framework for classifying basic underlying information processing variables. This system classified variables of interest for three separate components of an information processing model: input, process, and output. Although this listing is not exhaustive, it provides a basis for linking applied issues to more basic components. Such a linking eases the task of understanding the common elements of different research problems and may lead the researcher to useful psychological theories, evidence and methodologies. We used this classification system to organize the existing literature, to identify common issues and to direct future research. To maintain continuity, the same format is used in this review. The classification of information processing variables which appeared in LL-77 is reproduced as Fig. 1.

This review is limited to decision making research in accounting contexts which employs one of the following four research approaches: (1) lens model, (2) probabilistic judgment, (3) predecisional behavior and (4) cognitive style.¹ Three of these approaches were discussed in our earlier paper but predecisional behavior is a new direction in the accounting literature. Lens model research focuses on the interaction of information and the decision maker and their impact on decision quality. Often these studies involve the building of statistical models of human decision behavior. These models are used to infer the relative importance of different pieces of information and to assess various qualities of the decision and the decision maker. Research in probabilistic judgment has focused on describing and attempting to explain human failure to act in accordance with normative models of rational behavior, most notably expected utility theory. Predecisional

behavior research is concerned with the dynamics of problem definition, hypothesis formation and information search behavior. This research uses process-tracing techniques which attempt to provide a more detailed description of cognitive processes. Research on cognitive style is concerned with the impact of personal characteristics of the decision maker on the quality of his decisions and with the impact of information load on decision quality.

The organization of the papers in this review is as follows. Within each of the four major research approaches, there are two or more distinct categories related to information processing issues of primary interest. Within these categories, we have attempted to group studies by the accounting issue addressed. Within each study, the methods used, variables being studied, and experimental results are described and the implications of the studies are discussed. No attempt is made to critically analyze the individual papers included in this review. However, we do point to some weaknesses in design or interpretation in the discussions of significant results if this is necessary. For each research approach, summary tables are provided which highlight the studies included in that section. In these tables, the information processing variables are keyed to the classification system in Fig. 1. In the final section of this paper, accountants' use of the four approaches is evaluated and suggestions for further research are made.

LENS MODEL APPROACH

Brunswik's lens model, the use of which is now commonplace in the accounting literature, summarizes the basic principles of "probabilistic functionalism", Brunswik's framework for psychological research. The basic tenets of his framework are (see Brunswik, 1952, 1955 and Postman & Tolman, 1959):

- 1) Behavior is primarily a function of the nature of the environment;

¹ Other areas of behavioral accounting research are excluded from this review. Further, research related to other areas of human information processing (perception, psycholinguistics, etc.) is also excluded.

- 2) The numerous cues evident to the individual are usually imperfect and redundant predictors of environmental states; and
- 3) In response to this unpredictability and redundancy, individuals develop a range of substitutable processes for task achievement.

The significance of the task environment as a determinant of behavior is also expressed in his recommendation for *representative design* of experiments where *patterns* of variables in the environment are left undisturbed and environments as well as actors are sampled. This view is receiving increased attention in modern models of problem solving behavior (e.g. Newell & Simon, 1972).

The usefulness of this approach was recognized in the 12 studies reviewed in LL-77 which used analytical methods based on Brunswik's model. Most of these early studies attempted to describe certain characteristics of decision making in accounting and auditing contexts. In particular,

their primary goals were: (1) to build mathematical models which represent the relative importance of different information cues (often called policy capturing) and (2) to measure the accuracy of judgment and its consistency, consensus and predictability. Most were methodologically similar to studies conducted in other contexts using either regression or analysis of variance (ANOVA) to produce algebraic models of judgment.

Three accounting decision problems received considerable attention in the earlier literature: 1) the determination of materiality, 2) the evaluation of internal control, and 3) the analysis of financial statement data. In addition, preliminary attempts were made to investigate the impact of information set changes on information processing. The results were fairly consistent indicating: 1) the relationship to net income is of primary importance in materiality judgments while exact materiality limits for disclosure depend on the nature of the disclosure issue; 2) separation of duties is of greatest importance in internal control evaluation and differences in audit work schedul-

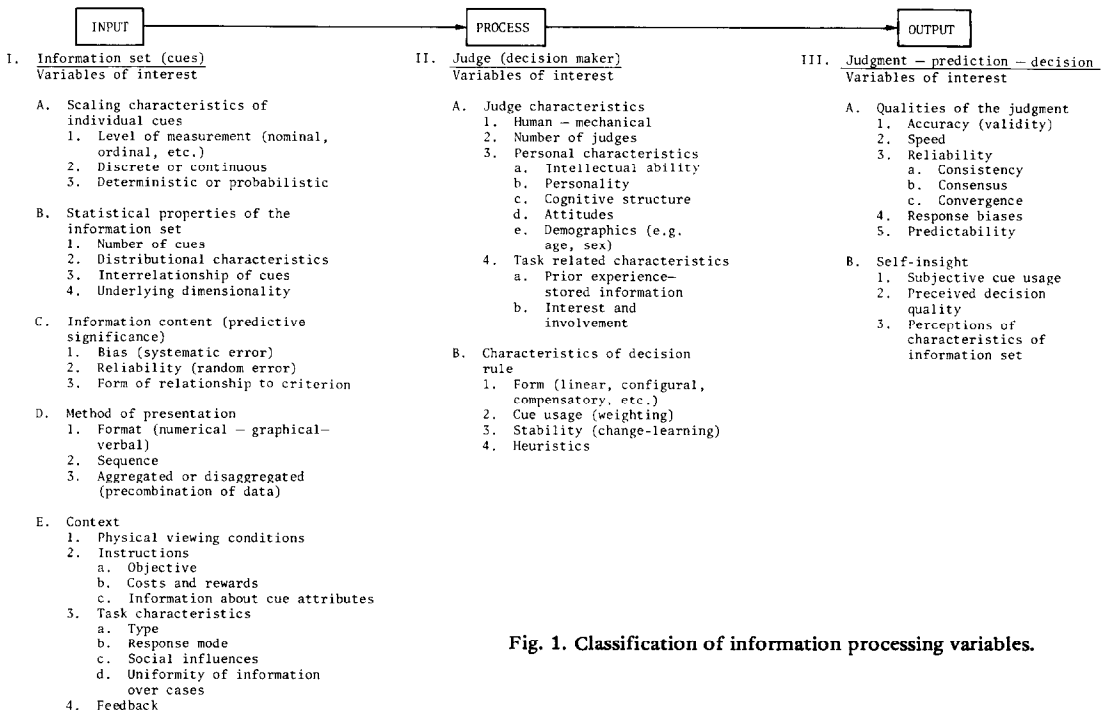


Fig. 1. Classification of information processing variables.

ing are primarily a function of differing utilities for various audit procedures and not of differences in internal control evaluation; and 3) users of financial statements appear to be able to make reasonably accurate judgments based on that data. Subjects in these studies exhibited many of the characteristics uncovered in examinations of other decision makers. In general their judgments were consistent over time and predictable and they exhibited varying degrees of between-judge consensus.

Much recent research has continued in the same vein, providing further description of the characteristics of judgment in the above mentioned decision context and in new contexts. In addition, a number of studies have addressed the more difficult issues of how judges learn information processing rules and the impact of data presentation, feedback, information search and other contextual variables on behavior. A number of novel methodological approaches have also been introduced.

The studies are classified into three categories on the basis of the information processing issues of primary interest. In the first category are *policy capturing* studies which examine the relative importance of different cues in the judgment process and consensus among decision makers. The second group evaluate the *accuracy* of judgments made from accounting data. Studies of the effects of *task characteristics on achievement and learning* are included in the third category. Within each category, results relating to different accounting decision problems are presented separately.

Policy capturing

The main concerns of policy capturing research are between-judge *consensus* and the *relative importance* of individual cues in the judgment process. Also, the *functional form* of the judgment rule and the judges' *self-insight* or awareness of their judgmental processes are often examined. Accountants' interest in these issues is deeply ingrained in accounting practice. Because we lack an objective definition of a "correct" decision in situations such as materiality judgment, consensus judgments of experts are often employed as a substitute criterion. This approach is obvious in our reliance on "general acceptance" as a test for

the validity of auditing and accounting methods. Models of these consensus judgments may also indicate useful decision rules and provide an explicit basis for policy discussions. In other situations such as internal control evaluation, the *degree* of judgmental consensus is often used as a substitute measure of decision quality; lack of consensus indicating that at least some individual's judgments are incorrect. The resulting models and measures of self-insight have additional implications for training. The implications of this research for practice are discussed in more detail in the final section of the paper.

Two policy capturing methodologies dominate the literature. Most often, ANOVA has been employed to construct experimental cases and build the judgment models. Each cue is first partitioned into a few discrete levels and then using each cue as a factor, ANOVA is used to combine the cues into experimental cases. From the judges' responses to the cases, the magnitude of the main effects and interactions are computed to measure cue usage. The strengths and weaknesses of this approach are discussed in Hammond & Stewart (1974) and Libby (1981, Ch. 2).

In situations where cues emanating from the environment cannot be specified and quantified in advance of the study, researchers have employed a second modeling method, multidimensional scaling (MDS), which first identifies the cues or dimensions on which judgments of cases differ and then indicates the perceived position of each case on each cue. Some MDS models also measure the relative weights placed on each cue by different individuals or groups. The analysis is based on measures of the perceived similarity of experimental cases. The method is particularly valuable in exploratory studies as it places fewer restrictions on the experimental design and thus provides a means of studying more realistic decision contexts. In fact, two of the projects discussed below involve analyses of non-experimental real world data.

Since 1977, policy capturing studies of internal control evaluation and materiality judgment have continued in earnest. In addition, a variety of new decisions have been examined. These studies are reviewed below.

Internal control. Three projects substantially replicated Ashton's (1974) study which assessed decision consensus, cue usage, decision rule form, and self-insight of auditors' internal control evaluations. In the experiments, the participants evaluated internal control cases indicating whether different internal control features exist. The cases were formed and analyzed using ANOVA. Ashton found that the auditors' evaluations exhibited a high degree of between-judge consensus and consistency over time in their evaluations. They relied most heavily on the separation of duties in forming their judgments and were quite aware of their judgmental process. The main effects ANOVA model accounted for most of the variance in their responses.

Major issues addressed in the new studies were *experience effects* and the generality of results to *alternative cue presentations*. Hamilton & Wright (1977) made minor modifications in Ashton's experiment to investigate the impact of experience levels. The authors constructed cases by omitting two of Ashton's six cues and splitting the two important separation-of-duties cue into three. Seventeen auditors with varying levels of experience participated. The results substantially mirrored Ashton's. (See Table 1.) Of particular interest was the fact that more experienced auditors exhibited greater consensus. No other differences based on experience were in evidence.

Ashton & Kramer (1980) and Ashton & Brown (1980) also replicated Ashton (1974). Ashton & Kramer (1980) compared the judgments of students and auditors in the same task. They hypothesized differences based on age, experience and wealth. Thirty undergraduate student volunteers completed a single replication of Ashton's 1974 payroll internal control instrument (6 cues in a 2^6 1/2 fractional replication design). The students were less predictable (74% versus 86.6%), placed less emphasis on separation of duties (36.9% versus 51.4%) and had less self-insight than the auditors. However, some of the differences may all have been caused by decreased test-retest reliability, which was not directly measured in the current study but is suggested by the lower linear predictability.

Ashton & Brown (1980) modified Ashton's

instrument to include two additional cues, making the task more complex and thus more realistic. In this study, 31 auditors (most with 1–3 years of experience) evaluated 128 cases (1/2 replication of a 2^8 design, plus 32 repeat cases). The two additional cues related to the rotation of duties and the use of background inquiries for new employees. Again, the results were almost identical to Ashton (1974). Separation of duties was by far the most important factor but the new rotation of duties cue was given little weight. They concluded that the added complexity of the task had no effect.

In addition to the above three replications Mock & Turner (1979) attempted to test the generality of findings of lack of consensus in audit work allocations to situations more representative of real world internal control evaluations. Following Joyce (1976), they investigated the effects of changes in internal control and differences in guidance on *sample size judgments* for four audit tests. Within the context of an extremely thorough set of background data, the authors manipulated the size of the change (weak to fair and weak to strong) and the level of detail in the instructions related to internal control. Unlike most studies, each of the 71 seniors and 2 supervisors from the participating "Big 8" firm evaluated only one case. As a result, reliance on individual cues could not be assessed. The degree-of-change variable was significant for all four procedures, including the procedure which was seemingly unrelated to the change (though probably interrelated with the other items in real life). The level of guidance concerning reaction to the change had no effect, suggesting that the participants were already aware of the firm's guidelines. A number of demographic variables were also unrelated to the responses. Consistent with Joyce's (1976) findings, the different auditors made widely varying audit work allocations in the same circumstances. All of these studies support the generality of the basic findings of Ashton (1974) and Joyce (1976). They also provide interesting insights into the impact of experience on consensus and the importance of rotation of duties.

Materiality. Two studies by Moriarity & Barron (1976, 1979) attempted to illustrate the use of

TABLE 1. Summary of lens model studies

Study	Type of decision maker	Task	Modeling technique	Variables of interest	Results
Hamilton and Wright (1977)	17 auditors	Rate payroll internal control from 5 cues on a 6-point scale	ANOVA (2 ⁵ factorial)	Decision rule form (IIB1) Cue usage (IIB2) Consensus (IIIA3b) Subjective cue usage (IIB1) Prior experience (IIA4a)	Highly linear Separation of duties most important High ($\bar{r} = 0.66$) High self-insight No effect
Ashton and Kramer (1980)	30 undergraduate auditing students	Rate payroll internal control from 6 cues on a 6-point scale	ANOVA (½ fractional)	Prior experience (IIA4b) Decision rule form (IIB1) Predictability (IIIA5) Cue usage (IIB2) Subjective cue usage (IIB1)	Suggested that differences with Ashton (1974a,b) due to these characteristics of subjects Highly linear Lower than Ashton (1974a,b) (74% versus 86.6%) Separation of duties highest but less important (36.9% versus 51.4%) Less self-insight than auditors (0.77 versus 0.89)
Ashton and Brown (1980)	31 auditors	Rate payroll internal control from 8 cues on a 6-point scale	ANOVA (½ fractional) plus 32 repeats	Decision rule form (IIB1) Predictability (IIIA5) Cue usage (IIB2) Consistency (IIIA3a) Consensus (IIIA3b) Subjective cue usage (IIB1) Sequence of cues (ID2) Number of cues (IB1)	Highly linear High (71.3% of the variance) Separation of duties most important, rotation of duties least important High ($\bar{r} = 0.91$) High ($\bar{r} = 0.67$) High self-insight ($\bar{r} = 0.86$) Different sequences had no effect 2 additional cues had little effect

TABLE 1 contd.

Study	Type of decision maker	Task	Modeling technique	Variables of interest	Results
Mock and Turner (1979)	71 seniors and 2 supervisors	Adjust planned sample size for 4 specific auditing procedures based on improvement in internal controls from extremely thorough case materials	ANOVA	Cue usage (IIB2) Instructions (IE2) Consensus (IIIA3b)	Change in control had significant effect Specificity had no effect Low; sample sizes for "strong" controls varied less than for "fair" controls
Moriarty and Barron (1976)	15 audit partners	Ranking materiality of error in estimate of depreciable life	ANOVA (factorial)	Decision rule form (IIB1) Cue usage (IIB2)	11 subjects classified as additive or nearly additive; 4 subjects classified as configural Net income most important
Moriarty and Barron (1979)	5 audit partners	Estimation of preaudit materiality based on five financial variables	Conjoint analysis	Consensus (IIIA3b) Cue usage (IIB2)	Low Income effect was the strongest
Schultz and Gustavson (1978)	5 actuaries	To assess risk of litigation against CPA firms from 5 cues (characteristics of practice and clients)	ANOVA (factorial) + 4 repeats	Cue usage (IIB2) Consensus (IIIA3b) Subjective cue usage (IIIB1)	All cue weights were significant Low ($\bar{r} = 0.12$) High self-insight
Gibbs and Schroeder (1979)	146 partners and managers from "Big 8" accounting firms	Evaluate the competence of an internal audit staff as required by S.A.S. No. 9	ANOVA (factorial)	Cue usage (IIB2) Consensus (IIIA3b)	Knowledge of company operations and techniques in auditing were the most important cues High

TABLE 1 contd.

Study	Type of decision maker	Task	Modeling techniques	Variables of interest	Results
Danos and Imhoff (1982)	40 audit partners and managers	Evaluate reasonableness of forecasts on 10-point scale based on 5 cues	ANOVA (½ fractional)	Predictability (IIIA5) Cue usage (IIB2)	High (80% of variance) Track record most important and background change had effect
Libby (1979a)	30 "Big 8" audit partners and 28 "money center" commercial lenders	Perception of messages communicated by different audit reports	MDS (INDSCAL) experimental data	Experience (IIA4a) Perception of characteristics of information set (IIB3) Cue usage (IIB2)	Perceptions of auditors and lenders appeared to be the same Two dimensions were identified and described as "need for additional information" and "amount of audit judgment required"
Libby (1979b)	34 commercial loan officers from 4 money center banks	To evaluate a \$ 2 million term loan request from a medium-sized company with varying degrees of uncertainty disclosure	ANOVA (factorial)	Cue usage (IIB2) Heuristics (IIB4)	Uncertainty disclosure, supplemental report was significant, whereas type of audit report had no effect Subjects appeared to estimate the most likely outcome of the uncertainty and treat it as certain
Rockness and Nikolai (1977)	Members of APB	Make accounting policy choices	MDS (ALSCAL) actual data	Personal characteristics (IIA3) Prior experience (IIA4a) Consensus (IIIA3b)	Three-dimensional solution suggested few systematic patterns except on a pragmatic—conceptual dimension
Brown (1981)	27 respondents representing sponsoring organizations of FASB, large CPA firms, and large industrial companies	Assessment of 51 policy issues to determine policy preference of respondents to FASB discussion memoranda	MDS (ALSCAL) actual data	Prior experience (IIA4a) Consensus (IIIA3b)	Strong separation between preparer and attestor respondents; only one cluster was evident; there were major changes in coalitions from issue to issue; FASB not dominated by sponsors

TABLE 1 contd.

Study	Type of decision maker	Task	Modeling techniques	Variables of interest	Results
Holt and Carroll (1980)	24 federal bank examiners	Classify loans into 5 categories from 12 financial and non-financial cues	Discriminant analysis, 20 simulated cases	Predictability (IIIA5) Cue usage (IIB2)	High (75.8%, two category group model) 4 of 6 variables identified by step-wise procedure were non-financial
Casey (1980c)	46 loan officers	Classify 6 ratio profiles for a 3-year period into fail or not-fail (not told priors)	Discriminant analysis, 30 real cases	Decision rule form (IIB1) Accuracy (IIIA1) Consensus (IIIA3b) Composite judge (IIA2)	Highly linear 56.7% versus Libby (1975) 74% 80% versus Libby (1975) 80% 60% versus Libby (1975) 82%
Zimmer (1980)	40 loan officers and part-time accounting students	Classify 5 ratio profiles for 3-year period into fail or not-fail (told priors)	Discriminant analysis, 42 real cases	Decision rule form (IIB1) Accuracy (IIIA1) Predictability (IIIA5) Composite judge (IIA2) Consensus (IIIA3b)	Highly linear 77% versus Libby 74% 90% versus Libby 88% 86% versus Libby 82% 72% versus Libby 80%
Abdel-khalik and El-Sheshai (1980)	28 commercial lending officers	To purchase up to a maximum of 8 cues with which to discriminate between firms that failed or did not fail	Discriminant analysis, 32 real cases	Human-mechanical (IIA1) Accuracy (IIIA1) Cue usage (IIB2)	A change in processing strategies increased accuracy by 5%; a change in selection strategies increased accuracy by 23.1% On average 3.5 cues purchased in first round and 1.5 cues in second; most frequently purchased were earnings trend, current ratio, and cash flow to total debt; cues added in 2nd round made no difference in accuracy High 84%

TABLE 1 contd.

Study	Type of decision maker	Task	Modeling techniques	Variables of interest	Results
Ebert and Kruse (1978)	5 security analysts	To estimate the returns of 35 securities on the basis of 22 information cues plus 15 repeat cases	Regression	Model of man (IIA1) Composite judge (IIA2) Consistency (IIIA3a)	Average model of man outperformed average man More accurate than 4 of the 5 analysts High
Swieringa <i>et al.</i> (1979)	228 introductory accounting students	Set product prices from 3 cues	Regression, 60 simulated cases	Information about cue attributes (IE2c) Uniformity over cases (IE3d) Stability (IIB3) Predictability (IIIA5)	No effect for amount of information about cue attributes Chow test indicated that there was no significant change in decision rule after change in the accounting method $\bar{R}_s^2 = 0.85$
Marchant (1979)	180 undergraduate accounting students	Set product prices from 3 cues	Regression, 30 simulated cases	Information about cue attributes (IE2c) Uniformity over cases (IE3d) Stability (IIB3) Predictability (IIIA5)	No effect for amount of information about cue attributes Change in accounting method had no effect on the decision rules of a majority of subjects $\bar{R}_s^2 = 0.925$
Harrell (1977)	75 Air Force officers	Evaluate the performance of 32 training wings represented by 5 dichotomous cues before and after receiving one of five different feedback combinations	ANOVA (2 ⁵ factorial)	Feedback (IE4)	Task-properties feedback and outcome feedback were highly effective

TABLE 1 contd.

Study	Type of decision maker	Task	Modeling techniques	Variables of interest	Results
Ashton (1981)	36 undergraduate students, 56 MBA students and 46 consortium fellows	To estimate product prices based on 3 cues, with varying feedback on task and accuracy	Regression, 60 simulated cases	Feedback (IE4)	Initial outcome feedback effective, supplemental task-properties feedback no effect Both G and R_e increased as R_e increased
Kessler and Ashton (1981)	69 MBA students	To classify bonds into 6 rating categories based on 3 ratio profiles	Regression, 34 real cases	Feedback (IE4) Accuracy (IIIA1) Predictability (IIIA5)	Two treatments, (1) univariate correlations between the cues and actual event and summary hit rates, (2) univariate correlations between cases and subject responses; univariate correlations between cues and actual event and summary hit rates, increased achievement and matching in session 2, then leveled out; the other 2 types had no effect; no effect on predictability
Moriarty (1979)	A. 277 introductory accounting students	A. To predict the failure of 22 discount retail firms, half of which had failed, based on 4 presentations of data		Format (ID1) Accuracy (IIIA1)	Subjects were more accurate using the schematic faces as opposed to financial balances or ratios
	B. 20 practicing accountants	B. To predict the failure of 11 discount retail firms		Information about cue attributes (IE2c) Speed (IIIA2)	Explanations had no effect Faces produced faster responses
Harrell and Klick (1980)	166 Air Force colonels	Rate captains for promotion a 9-point scale based on 5 cues	ANOVA	Presentation format (ID1) Cue usage (IIB2)	Greater weight was placed on the cost cue when presented in dollars as opposed to months to train

conjoint measurement techniques (see e.g. Green & Wind, 1973) to extend an earlier study by Boatsman & Robertson (1974) of auditors' materiality judgments. Conjoint measurement techniques first categorize *ordinal* judgments by decision rule form (e.g. additive, multiplicative, distributive, etc.) and then determine cue weights (usually called part worths). In practice results normally are close to the ANOVA model which analyzes *interval* judgments and assumes an additive or combination additive/multiplicative model. In the first (1976) study, 15 partners from eight large CPA firms ranked 18 cases ($3 \times 3 \times 2$ factorial ANOVA) according to the materiality of an error in estimate of depreciable life causing a decrease in earnings of \$0.5 million. The cases were represented by financial statements, and the net income earnings trend and asset size were varied by choosing arbitrary size multiples. Eleven of the subjects were classified as using additive decision rules and the remaining four appeared to use a number of cues interactively. This finding is consistent with the computationally simpler ANOVA studies. As in all prior studies, the net income effect was by far the most important. Moriarity & Barron also point out a number of problems faced in using the technique, including the large number of cue values necessary to accurately determine functional form, failure to use cross-validated measures of model fit and the assumption of error free data (see also Messier & Emery, 1980).

This problem was made even more clear in the second (1979) study, which assumed an additive model (like the main-effects ANOVA model). Their goal was to determine the size of the effect and the shape of the function of five cues in "overall preaudit materiality" judgments. In the study, no background information was presented to the subjects and the judgment of interest, overall preaudit materiality, was left undefined as it is in the auditing literature. Five audit partners from one firm completed the 30 experimental cases. To varying degrees, each indicated lack of familiarity with the task. While methodological problems limit the interpretability of the data (see Swicinga, 1979), it is interesting to note that the income effect was again strongest.

Litigation. Schultz & Gustavson (1978) studied the factors that contribute to the risk of litigation against CPAs. Because of the shortage of empirical data, the authors turned to the expert judge for insight. They studied the cue usage, consensus, and self-insight of five actuaries representing five of the six U.S. insurers of accounting firms. Each actuary judged the "probability of a valid claim" in 36 cases (2^5 factorial design plus 4 repeat cases) represented by five dichotomous cues which included the number of accountants in the firm, the percentage of "write-up work" performed, the rotation of accountants among clients, the size of clients and the financial condition of clients. These cases were presented in the context of extensive background information concerning the firm, its practice and the other terms of the insurance. While the responses were highly predictable and the subjects exhibited high self-insight, consensus among the five actuaries was surprisingly poor ($\bar{r} = 0.12$). More striking is the fact that all five agree only on the more risky level of *one* cue—client condition. On the other hand, the responses were highly predictable and the subjects exhibited a high degree of insight into their cue weightings.

Internal auditing. Gibbs & Schroeder (1979) studied the relative importance of various factors to the expert evaluation of the competence of an internal audit staff and the consensus of their judgments. The major contribution of the study is a detailed list of 54 criteria developed from an extensive survey. In the experiment, 146 partners and managers judged 32 cases, formed from a 2^5 factorial design, on a 4-point competence scale. The cues, varied across cases, were continuing education, educational background, knowledge of company operations, knowledge of new trends and techniques in auditing and the amount of supervision. Knowledge of company operations and supervision were most important on average. Unlike most such studies, only a group model (as opposed to individual models) was constructed. The high portion of group variance accounted for (68.5%) indicates substantial agreement across participants.

Reasonableness of forecasts. Danos & Imhoff (1982) analyzed the determinants of auditors'

judgments of the reasonableness of financial forecasts. Each of 40 auditors evaluated the reasonableness of forecasts resulting from two sets of 16 cases (1/2 replication of 2⁵ design). Each case was represented by five cues (track records in forecasting sales and income, bias tendency, sensitivity to industry activity and percentage forecasted increase in net income). The cases were presented in the context of two different sets of extensive background information about the company providing the forecast. The results suggest that the two track record cues were most important on average. The change in background information affected the importance of the "percentage forecasted increase in net income" variable, which indicates that the auditors evaluate this one differently for different industries.

Audit reports. Libby (1979a) compared 30 "Big 8" audit partners' and 28 "money center" commercial lenders' perceptions of messages intended to be communicated by different audit reports. Allegations of different perceptions had formed the rationale for suggested changes in the audit reporting framework. Each subject evaluated the similarity of the messages intended by all pairs of 10 different audit reports (unqualified and different types of uncertainty and scope qualifications and disclaimers) and rated the reports on 13 adjective rating scales. An MDS algorithm called INDSCAL was used to build representations of the participants' perceptual structures and the auditors and bankers were compared. Contrary to the beliefs of a number of policy makers, all measures indicated highly similar perceptions between the auditors and bankers. The two observed dimensions were tentatively identified by the researcher as "need for additional information" and amount of "audit judgment" required. Differences between the qualified and disclaimer opinions were twice as great as distances between the unqualified and qualified reports. The source of the scope limitation (client versus circumstances imposed) appeared important while the source of the uncertainty (asset realization versus litigation) appeared to be of little consequence.

Uncertainty disclosures. Libby (1979b) tested the effect of uncertainty disclosure and the incremental effect of the auditor's qualification on

lending decisions. Thirty-four commercial loan officers from four money center banks participated in the study. Using extensive background data and case specific information, they evaluated a \$2 million term loan request from a medium sized family-owned paperboard fabricating company. While ANOVA was used as the method of case construction, a number of modifications were made to achieve a more representative design. First, four basic cases were formed by combining two levels of complete financial statements and verbal management evaluations. These four cases were then combined with uncertainty disclosure-supplemental data combinations. Because consultation with the participating banks suggested that the litigation disclosure was *always* followed by a supplemental in-house investigation, these two variables were purposely combined into one three-level cue: (1) no disclosure, (2) disclosure combined with a supplemental report predicting a positive outcome, and (3) disclosure with supplemental report predicting a negative outcome. The subjects were then split into two groups depending on the type of audit report issued when an uncertainty was disclosed (unqualified or "subject to" qualification). Unlike prior studies, this factor was made a between-subjects factor to mask the principal purpose of the study — the test of the audit report variable. Both the financial statement and management evaluation manipulations were significant. While the uncertainty disclosure-supplemental report variable had a large significant effect on their judgments, the type of audit report seemed to have no effect. These initial conclusions were conditioned on the assumption that the loan officers would *not* change their information search behavior as a function of the form of the audit report — an assumption in need of further research.

Policy making. The other two studies employing MDS attempted to model the accounting policy preferences of major participants in the policy making process. Rockness & Nikolai (1977) analyzed APB voting patterns in a search for similarities associated with affiliation and possible client pressures. They compiled the voting records of all members and transformed them into similarity measures between each pair of members. The

three dimensional solutions computed using the ALSCAL algorithm suggested few systematic patterns except what appeared to be a conceptual-pragmatic dimension with academics and a few similarly inclined practitioners on the conceptual side separated from a compromise and pragmatic group on the other. Over time, placement of firm representatives in the patterns shifted quite drastically. No grouping based on "Big 8" affiliation or other obvious patterns emerged.

Brown (1981) performed a significantly more detailed analysis of the accounting policy preferences of respondents to FASB discussion memoranda. He identified 9 major issues resulting in standards and 27 respondents (mainly including the sponsoring organizations of the FASB, large CPA firms and large industrial companies) who commented on seven or more of the issues. The FASB position was also used to generate a hypothetical respondent. From the discussion memoranda, 51 individual policy questions were derived and similarity measures based on answers to these were computed for each pair of respondents. The ALSCAL method was used to generate an overall two-dimensional map. The sponsoring organizations of the FASB (AICPA, FEI, AAA, NAA, and FAF) were spread to all four corners of the map. There appeared to be a strong separation between the preparer and attestor respondents. Only one cluster was evident including four of the "Big 8" firms and the New York Society of CPAs. Not only did the FASB not side with the "Big 8" firms as has been alleged in Congress but the FASB often took an outlier position which was highly similar to the Financial Analysts Federation position only. This suggests that the FASB pays more than lip service to a user orientation. Further, when individual issue maps were produced, they indicated major changes in coalitions from issue to issue.

Loan classification. In the final policy capturing study, Holt & Carroll (1980) used discriminant analysis to model 24 federal bank examiners' loan classification decisions. A complex method was used to combine five financial variables and seven "exogenous" variables into 20 cases. Each subject classified the cases into five groups and a single

(group) discriminant analysis model was constructed on the basis of 360 of the 480 observations. A step-wise procedure included six of the original variables, four of which were exogenous (past due status, knowledge of credit situation, documentation and last year's classification). The model accurately classified 75.8% of the 120 case holdout sample into two categories (pass and all others). While the multicollinearity among the cues raises interpretation questions, the results are among the first to shed light on the effect of regulators' preferences on data gathering by loan officers.

Accuracy

Accountants' interest in serving the needs of users has motivated the study of the accuracy of predictions made from accounting data and the causes of discovered prediction error. While earlier studies in psychology had indicated that inconsistency and misweighting of cues often lead to low judgmental achievement, several accounting studies have indicated higher levels of achievement. In these studies, cases are usually constructed by sampling past real-world examples where outcomes are known. Judgmental accuracy is measured by the correspondence between predictions and outcomes and is often compared with the accuracy of mechanical decision rules. In addition, judgmental consistency, consensus and predictability are sometimes measured. Decision processes are normally modeled by regressing the judgments on the cues presented in the experimental cases or through use of discriminant analysis. Business failure prediction and security return prediction have received attention in prior research. Achievement has generally been high when compared with the predictive ability of the data. A number of decision aids including different types of regression models and mathematical "composite" judges have also been demonstrated.

Failure prediction. Two independent studies made similar extensions of Libby's (1975a, b) failure prediction study. In Libby's study, commercial loan officers predicted business failure on the basis of five-ratio, single-period, financial profiles. One-half of the firm sample had actually failed within three years of the financial statement

date. The results indicated that the banker's predictions were quite accurate, some approaching the accuracy of an environmental linear model.² The officers were also consistent over a one week time period, predictable and exhibited high consensus.

Major issues addressed in the two extensions were the effects of multi-period financial profiles, level of task predictability, disclosure of the base-rate of failure and generality across different subject populations. In Casey's (1980c) extension, participating bankers evaluated three-year six-ratio financial profiles. His subjects were not apprised of the highly unrealistic base rate of failure (50%) and the predictive ability of the data was slightly lower than Libby's. Results indicating high consensus and the importance of leverage, profitability and liquidity mirrored Libby's findings.³ However, judgmental accuracy was quite low (56.7%). The fact that, on average, 86.7% of the nonbankrupt firms and only 26.7% of the bankrupt firms were accurately predicted suggests that the use of base rates not related to the sample proportions may have overpowered the validity of their cue combination rules. Decreased predictability of the data, particularly in the third year (73.3%), may also have contributed.

This issue of the effect of priors was assessed in an independent yet surprisingly similar study of Australian bankers' and students' ability to predict failure for Australian firms which was carried out by Zimmer (1980). The major difference between this study and Casey's was that Zimmer's subjects were told in advance that half of the firms had failed. The predictive ability of the three year, five ratio data was also somewhat higher (88.1%). Zimmer's results almost exactly mirrored Libby's supporting the suggested causes of Casey's conflicting results. Of additional interest were findings that the bankers (like Libby's, 1976a) were more accurate for judgments in which they had greater confidence and that the part-time students' performance was very similar to the bankers'.

A third study of business failure prediction is of particular interest. Abdel-khalik & El-Sheshai (1980) have taken a step towards separating the impact of information choice on achievement and its use. As indicated above, studies of judgmental accuracy often compare human performance to that of mathematical models. In particular, they compare three types of processors: (1) human processors (HP); (2) "models of men" where mathematical representations of the subjects (from the right side of the lens) replace the subjects themselves (MP_s) and environmental or optimal mathematical models from the left side of the lens (MP_e). However, in all studies examined, a small number of cues were preselected for the subjects. To disentangle the effect of selection from processing Abdel-khalik & El-Sheshai considered two potential selection techniques; human (HS) and mechanical (MS). By examining the validity of the 6 combinations of selection and processing, conclusions concerning the contribution of both sub-processes can be drawn. In this initial study, four of the combinations are investigated. Twenty-eight commercial lenders evaluated 32 firms, one-half of which had defaulted on debt. Subjects could purchase a maximum of 4 cues from a list of 18 ratios and trends based on an explicit cost function. The participants were then given the opportunity to purchase up to four more cues before being asked to evaluate the firms a second time. On average, 3.5 cues were purchased in the first round and an additional 1.5 in the second. The most frequently purchased items in the first round were earnings trend, current ratio, cash flow to total debt and the trend in cash flow to total debt. Even though additional cues were purchased, there was no difference in accuracy between the two evaluations. The average subject responses were highly predictable (84%). The average accuracy for the four information choice/use combinations were: HS/HP = 62.5%; HS/ MP_s = 62.5%; HS/ MP_e = 67.5%; and MS/ MP_e = 90.6%. The fact that the change in *processing* strategies increased accuracy by only 5 percentage

² The environmental model is constructed by relating the cues to the actual event using discriminant analysis.

³ Detailed numerical results are presented in Table 1.

points, while switching from human to mechanical *selection* (given optimal processing) increased accuracy by 23 percentage points suggests that the choice of cues is crucial while the weighting is of lesser consequence. This conclusion is consistent with that of Dawes & Corrigan (1974), Einhorn & Hogarth (1975) and others. Although the complete six-celled matrix was not analyzed and is necessary to confirm these conclusions, this is the first study known to the authors to directly test how well individuals choose cues.

Security analysis. Ebert & Kruse (1978) investigated whether security analysts' predictions of rate of return could be "bootstrapped". Bootstrapping occurs when linear models of the decision maker outperform the decision maker himself. Bootstrapping will occur when the loss in accuracy caused by the judge's lack of reliability is greater than the improvement in accuracy gained by utilization of the information which is not captured by the linear model. In all but Libby's (1976a) failure prediction study, bootstrapping was the rule, not the exception. Ebert & Kruse (1978) asked five security analysts to estimate the 12-month rate of return on 35 securities (and 15 repeats) on the basis of 21 cues related to the economy, the industry, and the firm. Bootstrapping again was the norm (4 of 5 judges). Where the average achievement of the analysts was 0.23, the average model of man achieved 0.29.

Task characteristics, learning and achievement

Even though management accountants and information systems designers are responsible for determining much of the content and format of management reports, accountants have expended little research effort investigating the relationships of these variables to learning and achievement. However, psychologists have developed a considerable literature aimed at determining the impact on achievement of many of the information characteristics listed in Fig. 1. The attributes receiving the greatest attention from psychologists include task predictability, the functional form of cue criterion relationships, the number of cues, cue validity distributions and intercorrelations and feedback type. In the accounting literature four

issues have been addressed: (1) the impact of accounting changes, (2) feedback methods, (3) report format and (4) cue presentation.

Impact of accounting changes. Two studies extended Ashton's (1976) examination of whether judges adjust their cue weighting rules to changes in the accounting rules used to produce the cues. In a three cue product pricing task, Ashton measured the change in the regression model of subject responses resulting from a change from variable to full costing. The results generally suggested a change in processing.

The two new studies attempted to eliminate a number of alternative hypotheses proposed by Libby (1976b). Both Swieringa *et al.* (1979) and Marchant (1979) made a number of common changes: (1) they told subjects only that a change in accounting method had taken place, *not* that a change in *decision rule* was appropriate, (2) they used more meaningful statistical tests, and (3) subjects in the change and no-change conditions both evaluated cases that were otherwise common. In addition Marchant (1979) provided cases drawn from the same distribution during both halves of the experiment, while Swieringa *et al.* (1979) used different distributions. Marchant's (1979) subjects were also marginally more sophisticated than the other two groups and he cut the case sample size in half. Swieringa, *et al.* found that a large number of subjects in the experimental and control groups changed their decision rules, and by only one of three measures did more subjects experiencing the accounting change exhibit more processing changes than the controls. Marchant found few subjects changing their decision rules and no difference between the accounting change and non-change groups. These results and research in progress (Swieringa, August 1981, personal communication) suggest that the accounting change has no effect and that the large number of decision rule changes in Ashton's and Swieringa *et al.*'s studies were due to the change in the cue distributions between the first and second halves of the cases. This suggests that subjects were applying different markups to different priced items (non-linear processing).

Feedback. A major goal of management control systems is to provide performance feedback which results in improved performance in future periods. In a particularly novel study, Harrell (1977) investigated the impact of two vehicles for management control: organizational policies and feedback given by immediate superiors. In his research, Harrell recognized the parallel between these two motivational techniques and two forms of feedback which have been studied in the multiple cue probability learning literature: (1) "task properties" or policy feedback, where "optimal" or company policy weights for multiple cues are directly presented and (2) outcome feedback, where the superior's actual preferred judgment in each case is presented as feedback. In the experiment 75 air force officers evaluated the performance on an 8-point scale of 32 training wings formed from a factorial design. Each case was described by five dichotomous (satisfactory-unsatisfactory) cues including cost per pilot, quality of pilots graduated, highly competent all-volunteer force, compliance with regulations and aircraft maintenance. Each participant evaluated the cases twice, both before and after receiving one of five different feedback combinations: (1) no feedback, (2) policy feedback only, (3) policy feedback and *consonant* outcome feedback, (4) policy feedback and *dissonant* outcome feedback and (5) policy feedback and *random* outcome feedback. Judgments of group 2 (policy only) were more like the policy than were judgments of group 1 (no feedback). Group 3 judgments (policy plus *consonant* outcome feedback) were even more like the policy than were the judgments of group 2. Group 4 judges (policy plus *dissonant* outcome feedback) appeared to ignore the policy and follow the outcome feedback indicating their superior's preferences. Group 5 participants (policy plus random outcome feedback) were able to discern the random nature of the outcome feedback and to ignore it — performing the same as Group 2. While no subject followed the policies exactly this seems reasonable since these experienced officers would have prior beliefs as to the appropriate responses and would consider feedback but not ignore prior beliefs. These results conflict sharply with those of many psychological studies where

outcome feedback was of little use and often detrimental to performance. We will discuss the reasons for these differences at the end of this section.

Two other studies analyzed the effects of different types of feedback on the learning of environmental relationships. The studies were of similar general purpose to that of Harrell (1977) but did not directly address an accounting problem. Ashton (1981) used Ashton's (1976) product pricing task to examine the effects of two different types of feedback and three levels of environmental predictability (R_e) on the learning of an equal-weighting decision rule. Undergraduate, MBA, and Ph.D. student participants evaluated three sets of 30 cases with different types of feedback in between. The participants were assigned to one of three levels of environmental predictability. Most subjects appeared to learn the task from initial outcome feedback (observing correct prices for a sample of 30 cases); additional task properties feedback had no incremental effect. This finding is inconsistent with the psychological literature. However, this conclusion is open to question as a control group receiving no feedback was not employed and the default decision rule, equal weights, was optimal for the task. Lack of environmental predictability, which indicates the amount of "error" or randomness in the environment, was again shown to be a detriment to learning.

Kessler & Ashton (1981) analyzed the effectiveness of four types of feedback on the learning of a more realistic financial analysis task. Sixty-nine participants used 3 ratios to predict the ratings of a set of 34 bond issues 4 times, 3–4 days apart, receiving feedback between sessions. Environmental predictability (R_e) was 0.74. The subjects received 1 of 4 types of feedback: (1) summary hit rates; (2) univariate correlations between cases and subject responses to earlier cases which indicate the judges cue weighting policy plus summary hit rates; (3) univariate correlations between the cues and the actual event (task properties feedback) plus summary hit rates and (4) both types of correlations plus summary hit rates. The results suggest that only task properties feedback was effective. Note that the effectiveness of outcome feedback was not evaluated in this study. As a

group, these three studies support the effectiveness of task properties feedback. However, contrary to findings in the psychological literature, they suggest that humans are able to learn from outcome feedback in meaningful environments and even determine when different types of feedback are consistent or inconsistent. This issue is discussed further later in this paper.

Report format. The next two studies are of particular interest because they address the important issue of the relationship of data presentation to prediction accuracy and cue usage. Questions concerning data presentation, though a natural concern for management accountants and information systems designers, have received little attention from researchers.

Multidimensional graphics have been suggested as an aid to the human's ability to follow trends in related variables (such as financial statement data). The particular method investigated was Chernoff's (1973) schematic faces. In two experiments Moriarity (1979) evaluated the use of multidimensional graphics in place of standard financial statement presentations. In the first 277 introductory accounting students predicted the failure of 22 discount retail firms (half of which had failed) on the basis of 1 of 4 presentations of 6 years' data: (1) schematic faces with no explanation; (2) schematic faces with an explanation of what the features represented; (3) selected financial statement balances needed to calculate the Dun and Bradstreet key ratios, and (4) the key ratios themselves. The schematic faces were based on simple transformation (i.e. one financial variable controls the length of the nose, another the width, etc.). Financial variables were assigned to features on the basis of the author's judgment of their importance. Average errors out of 22 were 7.3, 7.09, 7.49 and 8.62 respectively. The only significant difference was that the "key ratio" group was less accurate than the other three. However, response times for the schematic faces groups were significantly lower. A second experiment compared the judgments of 20 practicing accountants based on the ratio and faces presentations. Each participant evaluated half of the firms on the basis of each presentation. The order was reversed for half of the subjects. The subjects judged an

average of 6.5 cases incorrectly using the ratios and only 4.7 using the faces.

Monetary vs. non-monetary cue presentation. In a novel experiment, Harrell & Klick (1980) determined whether cue usage is affected by monetary vs. non-monetary cue presentation. In a personnel evaluation task, 166 senior air force colonels evaluated 36 hypothetical captains ($2^2 \times 3^2$ factorial) for promotion based on five cues, three of which were varied. One cue, the training cost of replacing the officer, was also presented in three alternative forms: dollars, months, and dollars and months. The derived weight placed on the cost cue for the "dollars" and "months" cases were compared. Only the weight placed on the replacement cost cue was significantly different indicating that a greater emphasis was placed on the cue when it was measured in dollars. This result suggests either that the costs of training pilots for a certain time period were greater than the subjects expected or that the presentation metric caused the effect.

Research contribution

The above mentioned studies include a number of replications and several new directions. Many of the studies followed the dominant theme established in earlier research, producing descriptions of state-of-the-art decision making in various accounting and auditing contexts. However, a number of researchers moved beyond these preliminary descriptions to the development and testing of predetermined hypotheses concerning important accounting issues.

Replications and extensions. The research involving substantial replication or marginal extension of prior studies of audit judgments, materiality judgments, business failure predictions and pricing decisions for the most part confirmed prior findings concerning accuracy, consensus, cue usage and adjustments in cue usage resulting from changes in computational algorithms. These results indicate the generality of prior conclusions across minor changes in the task, experimental design and analytical technique. While the logic underlying many of the extensions is far from clear, this research has helped change many practitioners' attitudes toward the scientific study of professional

judgment. Earlier efforts were often met with objections by participants whose judgment was being scrutinized (see Elstein, 1976 and Dawes, 1979 for a discussion of similar responses in other fields). Now it seems that some practitioners are becoming convinced of the usefulness of these efforts. Studies of consensus in audit decisions have had a particularly large impact which will be discussed in the final section of this paper.

In addition, four new contexts were examined using methodologies established in the accounting literature, two major extensions in experimental paradigms were made, and two new analytical techniques were introduced. Schultz & Gustavson (1978), Gibbs & Schroeder (1979), Holt & Carroll (1980), and Danos & Imhoff (1982) investigated the determinants of the risk of litigation against CPAs, the quality of internal audit work, loan classification, and the reasonableness of accounting forecasts. These issues are of significant current interest to the profession.

Methodological issues. A number of methodological issues have also been addressed. First, concern over the representativeness of experimental paradigms used in this research was addressed in particular by Mock & Turner (1979) and Libby (1979b). They demonstrated how more realistic case material could be constructed within the constraints of the ANOVA design in their internal control evaluation and commercial lending settings. The reactions of practitioners to the Mock & Turner study suggest that, while the results remain essentially unchanged, a more realistic experimental paradigm made the results much more convincing to practicing accountants. Strong preferences by practitioners for more realistic paradigms has also been suggested in the field of medicine (Elstein *et al.*, 1978, p. 284). Second, Moriarity & Barron (1976, 1979) and Rockness & Nikolai (1977), Libby (1979a), and Brown (1981) introduced analytical techniques new to the decision making literature in accounting. The results of Moriarity & Barron (1976, 1979) question whether the increased complexity of conjoint measurement leads to significant benefits over the simpler ANOVA approach more prevalent in the literature. However, multidimensional scaling appears to show some promise as a tool for

analyzing less structured experimental situations and more importantly for analysis of archival data concerning decision making. Finally, Libby (1979b) used a combination of the within-subjects and between-subjects designs to eliminate "demand characteristics" resulting from the former. Demand characteristics often result from the within-subjects design because knowledge of the experimental manipulation allows the subject to uncover the experimenter's hypotheses and to behave accordingly.

New issues. Five new issues of interest were examined using lens model related approaches. The first two of the issues relate to recent regulatory action. First, allegations concerning the influence of different interest groups on accounting policy decisions were analyzed in two studies (Rockness & Nikolai, 1977 and Brown, 1981). Contrary to recent allegations, no "Big-8" block of votes surfaced in either analysis, nor was "Big-8" dominance of the FASB in evidence. In fact, to the contrary, alignments seemed to vary greatly from issue to issue and the FASB not only did not side with either industry or professional interest groups, but took unpopular outlying positions on a number of important issues. This may explain the mounting criticism facing the board. A split between practitioners and industry representatives on some issues was also in evidence. This is consistent with recent research analyzing the incentives of different parties in the policy making process.

Second, two studies (Libby, 1979a, 1979b) evaluated part of the formal communication process between CPAs and commercial lenders and the impact of one type of qualification on the lenders' decisions. Contrary to opinions expressed by some policy making organizations, little miscommunication between the two groups was in evidence. Recognition by the bankers of other sources of information concerning uncertainties appeared to make the auditors' qualification redundant in this situation.

The last three new issues have important implications for psychologists as well as accountants. First, three studies (Harrell, 1977; Ashton, 1981; Kessler & Ashton, 1981) investigated the impact of different types of feedback on learning a rule

for combining multiple cues into global judgments. Earlier psychological studies had presented a fairly dismal picture of human capabilities in this domain. However, most psychological studies use generic tasks where cues and judgment have no real world referents. The lack of such referents may prohibit storage of outcomes in the form of previously determined stereotypes — a common strategy suggested by research in memory. This substantially complicates the learning task. The three studies reviewed here, particularly Harrell (1977), took place in meaningful task environments. Subjects in this experiment were more experienced at drawing these types of inferences, and were probably more highly motivated. Determination of which of the above reasons or others explain these highly contradictory results awaits further research, but the higher rate of learning exhibited in this study creates questions about the validity of a whole body of literature. While certain relationships between environmental conditions and learning were consistent with prior studies, participants' ability to learn from outcome feedback was much greater. As most performance appraisal systems employ outcome feedback it is comforting to know that the results of the psychological research may be overly pessimistic.

Second and third, Abdel-khalik & El-Sheshai (1980), Moriarity (1979), and Harrell & Klick (1980) address parts of two extremely important issues which have been neglected by accounting researchers. They investigated the importance of both the selection of cues and the format of data presentation on the quality of judgment. These issues have been touched upon by the ill-defined construct "information load" in the cognitive style literature which we discuss later (see e.g. Lusk, 1979), but no rigorous definitions or presentation of their relationship to performance had been presented. While such studies may have less immediate reader appeal, they may have the greatest potential practical impact. We will return to this issue later.

PROBABILISTIC JUDGMENT

The idea of using normative decision theory in auditing (Kinney, 1975), management control

(Dyckman, 1969) and information system selection (Demski, 1972) has prompted a considerable volume of accounting research into the human processing of probabilistic information. Most models suggested for the accountant's use involve selection of an action which will maximize the decision maker's expected utility under circumstances in which the payoff or consequence to the decision maker is conditioned upon his action choice and the occurrence of some state of nature. Conceptually, such models require the decision maker to (1) specify all possible states of nature and feasible alternative actions, (2) define the payoffs or consequences and assign utility measures to them, (3) evaluate information and form a subjective probability distribution over the possible states and (4) choose the optimal action. The decision maker is assumed to be an expected utility maximizer and a Bayesian processor of information. Although these models are conceptualized as sequential, in practice we may be able to observe only the final action choice. To avoid the apparent confounding problems most research in this area has attempted to study separately specific components of the models. Probability estimation has received by far the most attention.

The seven such studies reviewed in LL-77 were for the most part replications and tentative extensions to business contexts of research appearing in the psychology literature. Three of these studies investigated the aggregation issue in accounting. These studies suggest that combined information systems are easier to use than joint or disaggregated systems and that decision makers frequently exhibit preferences for certain probability sequences in violation of simple expected value maximization. Three of the studies found that decision makers tend to use simplifying heuristics in their processing of information but, in contrast to earlier studies in psychology, such use may be sensitive to task and situation variables. One study showed the feasibility of using the Bayesian model to study information use in the analysis of financial information.

With so many variables of interest and so few accounting studies, no general conclusions were drawn in LL-77 from these prior studies. We suggested further research to match heuristics with

situations; to analyze the sensitivity of decisions to errors in probability estimates and to develop decision aids to compensate for the limitations in human information processing. Significant steps have subsequently been taken in these directions.

Our review of the current literature is arranged differently from the 1977 review, reflecting the diversity of recent work. The first section deals with the choice of techniques used to elicit subjective probabilities. The second section deals with departures from normative decision behavior. This section is further divided into (1) studies which seek to identify and evaluate common heuristics and biases and (2) studies which test the ability of decision makers to perform the role of an information evaluator. The third section includes studies which use a normative decision model as a framework to examine expert judgment.

Probability elicitation

In order to study the probability component of the judgment process, quantified representations of subjective probability estimates must be elicited from the decision maker. In attempting to measure this unobservable state of belief, we are concerned with how good the measurement is. Reviews of the psychology literature by Chesley (1977); Lichtenstein *et al.* (1977); and Slovic *et al.* (1977) have identified two major research directions. The first direction has been the investigation of various definitions of "goodness". Normative goodness refers to the extent that the elicited probabilities conform to probability axioms and correspond to the decision maker's state of belief; substantive goodness reflects the amount of knowledge of the topic area contained in the elicited probability and calibration refers to the long run appropriateness of levels of confidence. In general, the results of this research direction indicate (1) that most decision makers are overconfident, (2) that training seems to improve performance and (3) experts sometimes perform very well. The other research direction has been examination of the effect of different elicitation methods on the "goodness" of the measurement. This line of research has failed to identify a best method for eliciting probabilities.

Since several accounting studies have dealt with the comparison of different elicitation methods and since terminology in the literature is inconsis-

ent, it might be useful to describe briefly some commonly used methods. Methods can be conveniently classified as either direct or indirect. The most common direct methods include: fractile estimation, in which subjects assign values of the continuous variable to predetermined probability levels or fractiles of the cumulative density function (CDF) or the probability density function (PDF); bisecting techniques, in which subjects repeatedly bisect a range of the continuous variable into equally likely subdivisions; fixed interval methods, in which subjects assign probabilities to fixed partitions of the continuous variable in either the CDF or PDF; and curve fitting methods, where subjects draw a graph of the PDF. Indirect methods, where probabilities must be inferred from responses, include: the mean-variance method, in which subjects must specify the mean and variance of a normal distribution; equivalent prior sample (EPS), where subjects relate their feelings of uncertainty to having seen r occurrence in n trials; odds estimation where subjects give the ratio of the likelihood of two events and behavioral methods, where probabilities are inferred from the betting behavior of subjects in standard lotteries.

Convergence of methods in auditing. Three related accounting studies attempted to assess the convergent validity of different methods; that is the similarity of responses from two or more different elicitation techniques. Corless (1972) presented auditors with case descriptions about the internal controls in payroll preparation. Two methods of elicitation were used to assess their belief about the error rate in payroll preparation: (1) a beta distribution was constructed from responses to the bisecting method and (2) a discrete distribution was constructed from the responses to the fixed interval method. For each auditor, these distributions were compared on their medians and interquartile ranges. Although auditors were apparently quite willing to provide the necessary information, there was considerable discrepancy between the two distributions for most auditors. Felix (1976) compared a bisecting method and the EPS technique. After a brief training session on probability, auditors assessed prior probabilities for error rates in two attributes of an order-receiv-

ing, shipping and billing system. The two distributions were compared on the basis of quartile values; the results indicated somewhat smaller differences than those found by Corless.

Both Corless and Felix compared distributions on the basis of the average *difference* of quartile values as a percentage of the average quartile value. Crosby (1981) augmented this design with statistical tests of significance. Using direct estimation of fractiles and the EPS technique, Crosby's subjects assessed probability distributions for their beliefs about the error rate for one attribute. Although no training was given to the auditors, explanations of the methods were provided and consistency checks were incorporated to encourage participants to reexamine their fractile estimates for conformity to their beliefs. A beta distribution was constructed from the estimated fractiles and the goodness of the fit was assessed by a χ^2 -test. Note that prior studies merely assumed a good fit. The two distributions were compared with respect to their central tendencies and dispersions. Using both a paired *t*-test and a signed rank test, the null hypothesis of no difference in means, medians, variance and 90% credible intervals could not be rejected. However, the hypothesis of no difference in the 50% credible interval was rejected.

These results are not as encouraging as they might first appear. As Crosby noted, the range of possible error rates is small, from zero to about 10%; hence we would not expect much difference in the 90% credible interval. Also, the case material provided the previous year's compliance testing error rate which apparently became the mean estimate for the current year's estimate by participants. This was probably a problem with the Felix study as well, although Felix did not report the error rate provided to subjects. Finally, a quick calculation from Crosby's data indicates relative percentage differences even larger than those found by Felix.

Accuracy of methods in auditing. The studies reviewed thus far have examined the convergence of responses from different elicitation methods. However, two methods which yield poor normative and substantive probability responses may, nonetheless, have high convergent validity. Lack of

convergence is even more difficult to interpret. Two approaches have been taken to solve this problem. In an extensive series of experiments, Chesley (1976, 1977, 1978) developed an accuracy measure to objectively compare methods. His studies are also distinguished by the use of joint, nondichotomous distributions and by the fact that he tested hypotheses developed from psychological theory.

The main thrust of the three experiments was to examine certain theories which would explain why one elicitation method might be better than another. Torgerson (1958) described scale difficulty as a function of the number of cognitive scale elements (i.e. units, origin, distance). This theory would predict that a bisecting technique would be easier to use than direct estimation of fractiles. Chesley (1976, 1977) found the direct method to be superior in performance. Winkler (1967) and Slovic (1972) suggested that the ease of a response model is a function if its congruity with the way the information is mentally stored by the subject. Chesley (1977) found, however, that congruency of data presentation and response mode had no significant effect on performance. One last possibility, familiarity with the response mode, was tested (Chesley, 1978). Using five different response modes, Chesley was unable to find differences among them. This last experiment was hampered by small sample sizes and lack of an effective way of blocking by measures of familiarity.

Effect on audit decisions. Even if objective criteria for judging probability estimates are available, determining the practical effect of differences requires a measure of the effect of assessment differences on decisions. Two studies approached the question of the effect of different elicitation techniques on audit decisions. Crosby (1980) compared Bayesian sample sizes using input from both EPS and direct fractile methods. The subject auditors and the case materials were those described in Crosby (1981). Results of this study indicated that the normatively derived sample sizes were significantly dependent upon which method of elicitation was used. EPS generated smaller sample sizes than the fractile method. Both methods, in turn, provided smaller samples

than judgmental and classical sample sizes. Although there was no real benchmark by which to decide if a sample size was "too small", the results may suggest that the overconfidence (tight distributions) found by Lichtenstein *et al.* could lead to insufficient sample sizes and increased risk for auditors.

Kinney & Uecker (1979) examined the effects of different methods of eliciting subjective evaluations on compliance sampling results. Their methods differed only in the form of questions used to assess fractiles. Using methods similar to Tversky & Kahneman (1974), they asked auditors to evaluate one of four sample results and to assess either the 95th percentile population error rate or the probability that the population error rate was greater than 8%. Results of prior studies in psychology predict that the first method would yield confidence intervals that are too narrow while the second method would yield intervals that are too broad, presumably because the implied anchor points are different. In an audit context of evaluating sample results, these judgmental "errors" would be equivalent to increasing beta and alpha risk, respectively. For comparison purposes, Kinney & Uecker used classical evaluations of the sample results and counted the number of times subjects accepted the results (given an upper acceptable limit and confidence level) when they were not justified by classical evaluation. A χ^2 -test indicated significant dependence on the elicitation method. The direct fractile method was more likely to accept results more often than justified.

Heuristics and biases

One possible reason that different elicitation methods yield different distributions is that the different methods induce subjects to use different simplified processing rules, or heuristics. This explanation prompted the study by Kinney & Uecker (1979) cited above. It is becoming increasingly apparent that heuristic use is also dependent on task characteristics. In a review of the literature of heuristics and biases, Biddle & Joyce (1981) laid a base for an extensive series of experiments with the ultimate goal of suggesting to practitioners the conditions under which specific heuristics are likely to be employed; when errors

in audit judgment will result from the use of an heuristic; and methods of avoiding these situations. This section includes studies aimed toward that goal.

Representativeness in auditing. The representativeness heuristic (Tversky & Kahneman, 1974) generally posits that an assessment of the likelihood that A comes from population B will often be based on the extent to which A is similar to B. Frequently this process will lead decision makers to ignore normatively relevant data such as base rates, data reliability and predictability. Joyce & Biddle (1981b) tested for auditors' neglect of base rates and insensitivity to reliability in situations in which this heuristic could lead to systematic departure from normative responses. In two experiments auditors were asked to estimate the probability of management fraud given information about base rates, manager personality profiles and nonconclusive company descriptions. While auditors performed better than subjects in previous studies, they still underweighted base rate information in arriving at estimates exceeding Bayesian probabilities. In both experiments, lack of appreciation of base rate information is more pronounced when base rates are low. The potential impact on auditing is quite serious in certain areas where base rates are typically low and consequences are high (e.g. management fraud).

In experiments aimed at testing the effect of source reliability Joyce & Biddle (1981b) asked auditors to judge the probability of collection of an overdue account on the basis of a credit report from either a credit agency or the credit manager of the client. Results indicated that in a between-subjects design, the auditors did not differentially weight the source of information. In a within-subjects design, however, where each subject was sensitized to the two sources the auditors weighted the credit agency as more diagnostic. The authors suggest that explicit comparisons of the credibility of different sources could be built into audit programs.

Bamber (1980) developed a formal probabilistic definition of source credibility in an experiment to test whether audit managers differentially weight the work of different audit seniors. A normative Bayesian model was expanded to include measures

of sampling error and judgmental error (source credibility). Results indicated a highly significant main effect for source credibility. These results are consistent with the within subjects findings of Joyce & Biddle (1981b), but provide no information on a between-subjects basis. Again, the idea of making source credibility explicit in the judgment process is recommended.

Another related aspect of representativeness is insensitivity to the relationship between sample size and sampling error. Prior researchers have found that subjects frequently choose sample error rates most representative of their criteria (Uecker & Kinney, 1977) or that they choose samples with a larger sampling fraction (ratio of sample size to population size) despite the fact that a smaller sample drawn from another size population can have a smaller sampling error (Bar-Hillel, 1979). Biddle & Joyce (1979) ran a series of experiments to test auditors' appreciation of the role of sample size information. Auditors were asked to (1) evaluate two samples from different sized populations; (2) evaluate different sized samples from the same population; (3) evaluate sample results in isolation without population information. Results indicate that while more than half of the auditors performed normatively, a large number appear to have based their decisions on sampling fractions, or at least to have overemphasized sampling fraction information. Another large subset of the auditors conformed to neither the normative rule nor the representativeness heuristic.

Anchoring in auditing. Another common heuristic cited by Tversky & Kahneman (1974) is referred to as anchoring and adjustment in which decision makers choose some initial starting point from prior experience (a best guess, a random number, etc.) and then make adjustments from this anchor on the basis of additional information. Psychological research has shown that such adjustments are typically in the right direction but of insufficient magnitude. Again, since the audit process can be viewed as the updating of beliefs on the basis of current information knowledge of whether and in what situations auditors make these kinds of errors is important. Although several recent studies have addressed this issue

results have been inconclusive.

Joyce & Biddle (1981a) conducted three experiments to detect the use of anchoring and adjusting by auditors. The first experiment replicated a typical Tversky & Kahneman (1974) task using auditors and audit words, to provide a baseline measure of auditors' performance. Given a normatively irrelevant anchor, auditors were asked to estimate the incidence of management fraud. Results showed that the estimates of the group with a high anchor exceeded those of the group with a lower anchor. A second experiment asked auditors to make extent-of-audit judgments, given information that controls are either weak, changing from strong to weak or changing from weak to strong. Anchoring was not in evidence. The results showed some evidence of a contingent adjustment strategy where subjects made large adjustments when controls became weak and made small adjustments when controls became stronger. As the authors note, this behavior is consistent with a conservative approach to auditing.

In a third experiment auditors were asked to judge the probability of successful introduction given certain necessary elementary events. The experimental manipulation was to phrase the question in either conjunctive form (success requires all elementary events) or disjunctive form (failure results if at least one elementary event does not occur). The auditors were then asked to suggest an opinion on the client's financial statements. Results showed that the probability assessments were unaffected by the manipulations but that opinions varied widely. For example, one subject recommended an unqualified report based on a probability assessment of 0.5 while another subject who assessed the probability of success at 0.8 chose a disclaimer.

Kinney & Uecker (1979) reported evidence of anchoring by auditors in an analytical review application. Subjects were given audited sales, cost of goods sold, gross profit and gross profit percentage information for the prior two years. They were also given unaudited book values for the current year and were asked to provide a range of values beyond which they would investigate a change in the gross profit percentage. For one group of subjects, book values showed a significant increase

in this percentage; for the second group, there was a significant decrease. The mean upper and lower control limits set by subjects were significantly higher for the group with higher book values.

A second experiment by Kinney & Uecker (1979), discussed in the preceding section, recommended the use of a risk assessment elicitation method over a direct fractile approach in a compliance testing situation. Although the results indicate that auditors using the fractile method might be more likely to accept sample results when they are not justified, we must consider several issues. One might question the appropriateness of using a classical statistical evaluation as the benchmark rather than a Bayesian model which accounts for differences in priors. Further, the data show that if we define accuracy as percent deviation from the statistical evaluation, the auditors using the fractile assessment method were more accurate in 3 out of 4 experimental cases. It is not clear that an elicitation technique that is less accurate should be recommended. Finally, while the audit sampling issue is accurately addressed, the data really present *no* evidence relating to the anchoring phenomenon. Since we can only guess what anchor subjects may have used, we can only guess as to the direction and magnitude of the adjustments.

Anchoring in management control. While most of the research in this area has searched for generalized heuristic use, Magee & Dickhaut (1978) hypothesized that decision makers choose heuristics on the basis of situational variables. In a cost variance investigation case, they predicted that subjects under different compensation plans would exhibit different problem-solving strategies. Graduate business students made 24 investigation decisions based on cost reports and knowledge of the means and variances of the in-control (state 1) and out-of-control (state 2) probability distributions as well as the probability of state occurrence. Noting that the subjects lacked the means to solve dynamic programming problems or to explicitly perform Bayesian revisions, the authors predicted the use of a control chart approach. Such an approach would involve a lower limit L , below which investigation would never take place; an upper limit U , above which one would always in-

vestigate; and an interval between L and U that would trigger an investigation only after some number, N , of repeated observations. Subjects were paired into two different compensation plans. The author used a decision tree questionnaire to elicit heuristics used by the subjects. The experimental hypotheses were supported in that (1) most subjects used a control-chart strategy and (2) the compensation plan significantly affected the specific strategies used. Under each plan, subjects tended to choose the control-chart strategy consistent with maximization of their own compensation.

Brown (1980) also used a variance investigation task to examine the opportunity cost of sub-optimal behavior. In eight situations created by manipulating a statistical parameter, costs, and information levels, subject investigation strategies were only slightly more costly than a Bayesian model.

Anchoring in financial analysis. In assessing the accuracy of subjective probability judgments, Wright (1979) had students generate probability distributions for the systematic risk of securities. For each of fifteen firms, subjects received a measure of earnings variability and a debt to equity measure. At both the aggregate and the individual levels, there was evidence of conservative revision of probabilities, i.e. revision in an appropriate direction but to an inadequate degree. Subjects were more accurate for single cue versus joint cue position distributions. In a postexperimental questionnaire, subjects reported that, in the joint cue tasks, they focused on the variability of earnings cue and "adjusted" their estimate for the value of the debt to equity cue.

Sequence effects. Two studies provided further insight into Ronen's (1971) finding of a sequence effect in problems involving disaggregated probabilistic information. Ronen had found that most subjects prefer higher initial state probabilities when joint probabilities are equal and, in many cases, even when the other alternative has a higher joint probability. Hirsch (1978) extended the Ronen study by using both a chance task and a business task, by manipulating more independent variables and by incorporating a personality variable. A factorial design manipulated the

differences in initial probabilities, the joint probability magnitude and the difference in joint probabilities. Results showed that in both tasks, when the joint probability difference was zero, the sequence effect existed at all levels of the other variables. As joint differences increased, progressively higher levels of the other variables were required to produce the sequence effect. The deviations from expected value maximization were much greater in the chance task than the business task. In the business task, subjects who scored as internals on a locus of control scale were significantly more prone to the sequence effect than externals who were almost unanimously expected-value maximizers.

Snowball & Brown (1979) also used a business context to examine bank trust officers' use of disaggregated probabilities. They set up a business task capable of distinguishing expected value maximization, preference for high initial step probabilities, preference for high second stage probabilities and anti-expected value maximization. Although nearly two-thirds of the responses were consistent with the normative model, the next most preferred response (18.5) was a preference for higher initial stage probabilities. Another 11% of the responses showed a preference for higher second stage probabilities. As in Hirsch (1978), nonnormative behavior decreased as joint differences increased. Results also showed that sub-optimal strategies were more prevalent among those subjects with a higher disposition toward risk.

Information evaluation

The studies in the preceding section sought evidence of specific simplifying heuristics to explain departures from normative standards. In contrast, the studies in this section are concerned only with the question of whether or not decision makers have the ability or can learn to perform as required by normative decision models. Three related studies have dealt with the accountant's role as an information evaluator in choosing an information system for another decision maker who will make an action decision. Each of the three studies involved a number of urns which contained varying proportions of black and white marbles.

A simulated decision maker was to guess the proportion of black marbles in an urn (selected at random), given the prior probability distribution and the results of a sampling of the marbles in the urn. The task required of the subjects (all students) involved the choice of an appropriate sample size (i.e. to choose an information system) knowing the payoffs. Subjects were monetarily rewarded in such a way as to encourage expected value maximization.

Uecker (1978) used two different simulated decision makers, one Bayesian and one Conservative-Bayesian, to test subjects' ability to learn the optimal information system to provide the decision makers. Using a fixed per-unit cost of sampling, each subject performed fifty trials with feedback with each decision maker. Results showed that the subjects were apparently able to distinguish between the two simulated decision makers since average sample size choices for the two decision makers were significantly different. Moreover, on average, subjects were closer to optimal sample size for the Bayesian decision maker. Compared to a normative model, however, the subjects did not tend to converge toward the optimal sample sizes for either decision maker. In both cases and regardless of the order in which the decision makers were presented, no significant amount of learning occurred over 50 replications.

In another version of this experiment Uecker (1980) described a simulated decision maker to half the subjects to see if explicit knowledge of the decision rule would increase their ability to choose an optimal information system. Results indicated no difference in performance between those who received information about the decision maker and those who did not. An important confounding feature in these two experiments was the fact that the actual curve relating sample size and expected net gain from sampling was rather erratic. This meant that it is possible that subjects found themselves in a position from which both increases and decreases in sample size would make them worse off. In fact, as the author points out, a sample size of 40 may have had a better payoff than a sample size of 24 even though the optimal sample size was 22.

Hilton *et al.* (1981) tested the extent to which

subjects correctly perceive the effect of accuracy on information value. Subjects were given knowledge of the decision maker and were presented with a series of trials. For each trial they were offered a particular sample size at a specific price. By varying prices and sample sizes over time, they were able to compute a demand value of information for each subject. The normative responses would show information value increasing in accuracy with declining marginal returns. On average, the subjects were very close to normative values, both in terms of absolute amounts and in recognizing the declining marginal value of increased sample sizes. Only one individual subject, however, exhibited monotonically decreasing marginal increments in information value.

Normative framework

The research dealing with heuristic processing of information and with the ability of students to perform the information evaluation functions has compared actual performance with some objective or normative standard. The three papers in this section represent a more descriptive approach to the study of decision making. In these studies, normative decision theory is used not as a standard of performance but as a framework for examining elements of the decision process. All three papers deal with the materiality construct in auditing.

Materiality. From the extensive history of conceptual and empirical research on materiality, Newton (1977) was the first to explicitly address the effect of uncertainty on materiality judgments. Audit partners were presented with a case involving a decline in value of marketable securities. Each subject was asked for a dollar amount of decline which, if permanent and not written down by management, would be material enough in relation to net income to warrant a qualified opinion. Note that this "certainty equivalent" is the end product of most prior materiality studies. Subjects were then presented with several dollar value declines and asked for the minimum probability that the decline would be permanent which would justify issuance of a qualified opinion. The purpose of these standard lottery questions was to estimate a utility curve for each subject over the

range of values in the case. A final question provided a specific dollar decline and a probability of decline and asked whether the subject would qualify the audit report. Responses from the elicitation phase were used to predict the answers to the final question. Although some subjects would qualify without regard to probability, results indicated that most of the audit partners seemed to use probabilities in their judgments. Other results indicated that most auditors were risk averse and that judgments were consistent with expected utility maximization. Some subjects exhibited invariance of probabilities over different dollar amounts, a result which Newton viewed as a violation of utility theory. But note that such behavior does conform with a model constrained by absolute aversion to risk of all losses in excess of some cut-off point (see Libby & Fishburn, 1977).

Audit decisions. Most studies which have examined the degree of consensus among auditors in extent of audit decisions (see e.g. Joyce, 1976) have found significant individual differences. To add insight into the causes of disagreement, Lewis (1980) viewed the audit decision process within an expected utility framework and suggested that specific elements of the process could be examined in isolation: utilities and subjective probability distributions over the set of states. Both Lewis (1980) and Ward (1976), who implicitly used a similar model, investigated the homogeneity of utility functions.

To see if auditors considered the same factors in a materiality decision, Ward asked audit partners and managers to rank the importance of 24 factors in making materiality judgments. These factors included elements of the legal, technical, professional, personal and environmental influences on the auditor. Results of this ranking indicated significant (Kendall's $W = 0.386$, $p \leq 0.01$) but not overwhelming agreement among auditors. Ward also examined the perceived relationship between the size of an audit error and the expected loss to the auditor. There was little agreement about the functional form of the relationship. Although 12 of 24 subjects chose either logistic or exponential relationships, all the forms were

chosen by at least one subject and five subjects provided their own tracings.

Lewis (1980) chose an audit case involving disclosure of a contingent liability and in which both the states and the actions were given. The purpose of the study was to examine the degree to which auditors have homogeneous utilities. Practicing CPA's, mostly supervisors and managers, were asked to express their preferences for the outcomes associated with a two-state, three-action decision. The preferences, shown on an 11-point scale, were used as interval scale utility measures. A between-subject design was employed by assigning subjects to either a high or low materiality situation. Homogeneity was measured as the average pairwise correlation of the utility measures among all auditors in each case. Results suggested that the homogeneity condition is significantly more likely as the level of materiality increases.

Research contribution

Table 2 summarizes the studies reviewed in this section and relates them to our classification of information processing variables. A number of promising avenues for further research have been developed. In elicitation theory, most of the studies have tested and confirmed results of psychological studies which indicate low convergent validity of elicitation techniques (Corless, 1972; Felix, 1976; Crosby, 1981) and that training may increase convergence (Felix, 1976). Chesley (1976, 1977, 1978) examined the effect on accuracy of different elicitation methods, congruity of data and response mode and personal characteristics and has generally found that none of these variables significantly affects accuracy. From a procedural view, however, he has found that the use of multiple stage elicitation, with reconciliations, improves accuracy. Two studies have shown that the low convergent validity among techniques could lead to significantly different sample size recommendations (Kinney & Uecker, 1979; Crosby, 1980). The implications of this important finding are discussed in the final section of this paper.

Studies of heuristics and biases indicate that the search for generalized simplifying strategies is not a simple task. We have noted in several cases that results interpreted in terms of preconceived

heuristics are also open to radically different interpretations. Further, some studies indicate that accountants have developed their own heuristics for certain tasks (e.g. Joyce & Biddle, 1981a). What is clear from the results is that many of the rules of probability theory are not well understood by accountants. Generally, it appears as though some auditors are nearly normative, some auditors act as if they use the Kahneman & Tversky's heuristics and some do something else. Magee & Dickhaut (1978) provide some evidence that situational variables, such as reward structures, cause decision makers to "choose" among simplified problem strategies. Similar explanations for task variables may explain the absence of any generalized heuristic use. We will return to this issue in a later section. Bamber (1980) provided a more objective approach by developing a formal probabilistic definition of source credibility which would seem to have a wider use in measuring an auditor's perception of the credibility of audit evidence in general.

The application of normative decision models to auditing, management control and information system selection has prompted several lines of research. One benefit of using a normative model is the explicit requirement to separate components of the decision process. Three studies dealing with the materiality construct have provided that separation by examining elements of auditors' utilities (Ward, 1976), the effect of changes in materiality levels on utilities (Lewis, 1980) and the effect of probabilities and risk on materiality judgments (Newton, 1977). In experiments aimed at evaluating the ability of decision makers to perform normatively with IE models, results are mixed. Hilton, *et al.* (1981) found that subjects apparently did not perceive the declining marginal increases in information value as accuracy increases. Uecker (1978) found that subjects were unable, over 50 replications, to converge on optimal sample sizes for a simulated decision maker even when the decision model of the DM was known. In these three studies, it is obvious that the subjects did not have the means to compute the expected net gains from sampling nor to do even simple Bayesian revisions.

If we already have normative models to select

TABLE 2. Summary of probabilistic judgment studies

Study	Type of decision maker	Task	Variables of interest	Results
Corless (1972)	Practicing auditors	Respond to two different subjective probability elicitation methods in estimating error rate in payroll preparations	Response mode (IE3b) Convergence (IIIA3c) Consensus (IIIA3b)	Within subjects different techniques produced different distributions; between subjects, very little agreement
Felix (1976)	Practicing auditors	Receive brief training session and respond to two different elicitation methods in estimating error rates in shipping-billing function	Response mode (IE3b) Convergence (IIIA3c) Prior experience (training) (IIA4a)	Smaller distributional differences than Corless, measured by quartile differences; possible training effect
Crosby (1981)	42 audit seniors	Respond to two different elicitation techniques in estimating error rate in billing-collection function	Response mode (IE3b) Convergence (IIIA3c)	Statistical tests of central tendency and dispersion indicate no differences in 90% intervals, but significant differences in 50% intervals
Chesley (1976)	28 IE and 28 accounting undergraduates	Respond to two elicitation methods to describe prior distribution or likelihood distribution based on manufacturing case material	Type task (IE3a) Response mode (IE3b) Prior experience (IIA4a) Accuracy (IIIA1)	Nature of task and background of student had no effect on accuracy; significant elicitation technique effect on accuracy in first trial, but subsequent trials showed no effect
Chesley (1977)	28 MBAs	Same as (1976), describing prior distributions	Response mode (IE3b) Format of data (ID1) Cognitive structure (IIA3c) Accuracy (IIIA1)	Ease of response mode not a function of congruity of data and response mode; contradicts theory that method requiring more scale elements would cause greater cognitive difficulty; significant correlation between performance on embedded figures test and elicitation task

TABLE 2 contd.

Study	Type of decision maker	Task	Variables of interest	Results
Chesley (1978)	47 CA students and 42 MBAs	Respond to five elicitation methods to describe marginal likelihood and joint likelihood distribution using (1976) case materials	Response mode (IE3b) Accuracy (IIIA1)	No significant differences in performance using any of the five methods
Crosby (1980)	42 audit seniors	Priors elicited in Crosby (1981) were used to compute Bayesian sample sizes; subjects asked to provide judgmental sample size	Response mode (IE3b) Response bias (IIIA4) Convergence (IIIA3c)	Different elicitation methods for priors result in significantly different sample sizes; Bayesian samples smaller than classical or judgmental samples
Kinney and Uecker (1979)	179 audit seniors	Set up investigation region for analytical review; use one of two elicitation methods to evaluate sample results	Response mode (IE3b) Heuristics (IIB4) Response bias (IIIA4)	Book-value "anchors" affect size of investigation region; different elicitation techniques provide different evaluation of sample outcomes
Joyce and Biddle (1981b)	Large number of practicing auditors spread over three experiments	Judgments of probability in a variety of audit settings ranging from management fraud prediction to collectibility of receivables	Information content (IC1,2) Heuristics (IIB4)	Base rates are not ignored, but are typically under-weighted, especially when base rates are low; reliability of source of data was ignored in between-subjects design; in within-subjects design, subjects were sensitized to the differential reliability
Bamber (1980)	35 audit managers	Judge likelihood that client controls are adequate; revise probability based on senior's report of compliance testing	Information content (IC1,2) Heuristics (IIB4)	Subjects were sensitive to changes in reliability of audit seniors as source of data

TABLE 2 contd.

Study	Type of decision maker	Task	Variables of interest	Results
Biddle and Joyce (1979)	Large number of practicing auditors spread over three experiments	Evaluate sample evidence and make sample-size choices in audit settings	Information content (IC2) Heuristics (IIB4)	Large proportion of subjects tend to overemphasize the sampling fraction, ignoring sample size; on the other hand, subjects generally understood inverse relationship of sample size and sample error
Joyce and Biddle (1981a)	Large number of practicing auditors spread over three experiments	Make probabilistic judgments in variety of audit decision settings	Information content (IC3) Type of task (IE3a) Heuristics (IIB4) Response bias (IIIA4)	Normatively irrelevant anchors affect judgments only in unfamiliar tasks; in more familiar tasks, anchor and adjustment strategy could not explain degree of adjustment
Magge and Dickhaut (1978)	38 graduate business students	Make variance investigation decisions	Costs and rewards (IE2b) Heuristics (IIB4)	Nature of incentive system influenced choice of investigation strategy
Brown (1980)	86 senior undergraduate and graduate business students	Make variance investigation decisions	Distribution (IB2) Costs and rewards (IE2b) Heuristics (IIB4) Response bias (IIIA4)	Assuming a mid-point anchor, adjusting strategy not consistent with anchor and adjustment heuristic
Wright (1979)	11 MBA and Ph.D. students	Estimate probability distribution for the systematic risk of a security given single or joint accounting risk measures	Number of cues (IB1) Form of decision rule (IIB1) Heuristics (IIB4) Accuracy (IIIA1)	In aggregate and individually, subjects appeared to be conservative Bayesian; self-report indicated sequential process for joint cues, anchoring on one and adjusting based on the other; subjects more accurate with single cue

TABLE 2 contd.

Study	Type of decision maker	Task	Variables of interest	Results
Hirsch (1978)	48 manufacturing corporation managers	Choose between sets of sequenced events in business and nonbusiness settings	Sequence (ID2) Type of task (IE3a) Heuristics (IIB4) Cognitive structure (IIA3c)	Sequence effect (Ronen, 1971) was found; as joint probability difference increased, subjects required higher difference in initial stage probabilities; effect more pronounced in nonbusiness task and by subjects with internal locus of control
Snowball and Brown (1979)	37 trust officers	Choose between sets of sequenced events in business settings	Sequence (ID2) Cognitive structure (IIA3c) Heuristics (IIB4)	Nearly 2/3 subjects were normative; 1/5 showed preference for higher initial probabilities; sequence effect diminished as joint probability differences increased; risk-seeking subjects were more suboptimal
Uecker (1978, 1980)	40 and 41 undergraduate students, respectively	Choose appropriate sample size for simulated decision maker (costly sampling) in colored marble and urn problem	Information about cue attributes (IE2c) Stability (IIB3) Heuristics (IIB4) Accuracy (IIIA1)	Subjects could discern difference in decision rule of DM, but did not converge toward optimal choice, even when decision rule was known; normatively irrelevant anchor affected sample-size choices; no significant learning over multiple trials

TABLE 2 contd.

Study	Type of decision maker	Task	Variables of interest	Results
Hilton <i>et al.</i> (1981)	104 MBA students	Decide whether to purchase sample sizes at a given price for simulated DMs use in Uecker (1978) type task	Costs and rewards (IE2b) Accuracy (IIIA1)	On <i>average</i> , subjects perceived monotonic effect of accuracy on information value and their perceptions approximated the normative model, although few <i>individuals</i> were that good
Newton (1977)	19 CPAs, partners	Decide whether to qualify opinions in three circumstances given dollar magnitudes and probabilities; standard lotteries were used to estimate utility curves	Form of decision rule (IIB1)	Most subjects used probabilities in their judgments; most were risk averse and were consistent with expected utility maximization
Lewis (1980)	72 CPAs, mostly supervisors and managers	Rate desirability of outcomes in disclosure decision	Consensus (IIIA.3b)	Utilities for audit outcomes are more homogeneous across auditors as materiality increases
Ward (1976)	24 CPAs, managers and partners	Rank importance of 24 materiality factors; indicate perceived functional relationship between size of audit error and expected loss of auditor	Consensus (IIIA.3b) Perception of characteristics of information set (IIB3)	Statistically significant, but weak, agreement in rankings of materiality factors; half of subjects specified either logistic or exponential relationship

information systems for known decision makers, we might ask why we are concerned with whether a student subject behaves normatively. There are several possible answers to this question. One is that we should not be concerned; we should use the normative model. On the other hand, there are probably no real-life situations in which a human decision maker's model is completely specified; nor are we likely to find situations in which all information systems and their potential signals are known. By looking at idealized cases, where normative solutions exist, we can get a better idea of how people may be performing in those actual situations and how we may help them to perform more effectively.

PREDECISIONAL BEHAVIOR

In our 1977 paper we shared Einhorn's (1976) concern that most accounting experiments investigated highly structured repetitive situations where the task was well defined, the subject was presented with information and the possible responses were prespecified. Many important accounting decisions involve ill-defined tasks where the decision maker must search for information and generate and evaluate possible responses. Accountants have recently begun to investigate the dynamics of problem definition, hypothesis formation and information search in these less structured situations by using measurement techniques designed to examine *predecisional behavior*. The techniques used in lens model and probabilistic judgment research, for the most part, measure initial inputs and final outputs from which their functional relationship may be inferred. The techniques discussed in this section, which are often called process tracing methods, require a large number of intermediate responses which allow a more detailed sequential set of relationships to be assessed. The principal benefits of these methods are a richer level of *detail* and the ability to provide *sequential* measures of decision behavior.

These techniques were developed or refined by Newell & Simon (1972) and their associates to help build and test their theory of human problem solving. Payne *et al.* (1978) discuss the three data

collection techniques most often used to study predecisional behavior: verbal protocols, explicit information search, and eye movements. Verbal protocols are usually gathered by requiring participants to "think-aloud" into an audio or video recorder while performing the task. The tape is then transcribed and the protocols are further classified into predetermined formal categories relevant to the researcher's hypothesis. The results of the codings are often displayed as tree graphs, matrices and computer programs. Explicit information search measurements are familiar to the accounting literature (see e.g. Pankoff & Virgil, 1970). These methods require the decision maker to acquire each piece of information separately so that an accurate record can be made of each acquisition. The technical sophistication of the data gathering equipment can vary from piles of cards to computers. The resulting data are used to measure cue usage and search sequence. Recording of eye movements and fixations can also provide data concerning cue usage and search sequence. Russo (1978) provides a more detailed discussion of this technique.

As was noted in LL-77, Clarkson (1962) provided an early extension of Newell & Simon's theory of problem solving into the realm of decision making under uncertainty and in particular to financial analysis. While the basic literature in problem solving continued to evolve, few applications in this area immediately followed. Since Einhorn's (1976) synthesis and our first paper, there has been a renewal of interest in this research. Two studies appear to have had the greatest influence on this revival. Payne's (1976) study of the impact of task complexity on choice of decision strategy was noteworthy for its use of rigorously defined operational hypotheses tested in an internally valid experimental design and multiple measurement methods which increase the external validity of the results. He also employed theoretical developments drawn from Simon's work *and* research aimed at developing algebraic representations of judgment which were discussed in the lens model section of this paper. The second influential study, reported by Elstein, *et al.* (1978), involved a series of medical diagnosis experiments which combined the measurement technology and

models of memory developed by Simon and his associates with the prescriptive orientation and theory of heuristics and biases developed in the probabilistic judgment literature. While much of the problem solving literature is primarily descriptive, Elstein *et al.* tried to relate different strategies for hypothesis formation and information search and interpretation to the accuracy of judgment. The advances represented in these two studies were a major step in the direction of overcoming many objections concerning the validity of this research and indicated how it could be related to the accountant's interest in *improving* the quality of decisions.

Development of accounting research using this approach is in its early stages. The basic psychological theory and analytical methods are also in earlier stages of development. However, several promising studies concerned with financial statement analysis, performance evaluation and audit planning have been completed. Some of these studies examined the information search and cue combination strategies used by experts, others have attempted to compare experts with novices to determine the unique elements in expert strategies and one study has begun to investigate the impact of task characteristics on strategy choice. Like the early research using the lens model framework, these studies attempted to describe the state of the art in decision making in a number of accounting contexts. Most involved straightforward replications of studies in other contexts.

Financial analysis

Four studies have been conducted aimed at modeling expert financial analysts. The first such study, which we referred to earlier, is Clarkson's (1962) attempt to construct a model of a bank trust officer's portfolio selection process. After gathering background information through interviews, observations of meetings and examination of documents, verbal protocols were taken from one trust officer as he selected securities to be

included in new client portfolios. Based on the protocols and prior evidence, a computer program was intuitively derived and tested. Separate models for income and growth portfolios were constructed. The security selection portion of the models primarily involved a conjunctive process where each security was subjected to a series of up to 15 binary tests until one security in an industry was found to meet all the tests. Additional industries were then subjected to the same process until the available funds were invested. The data used in the model included financial statements, stock prices and forecasts. The ability of the models to predict the portfolio selections was tested on four new accounts not used in construction of the models. The predictions were quite accurate and were superior to random and naive single variable models. The accuracy of the underlying representation was tested by a rough comparison with the protocols produced while the subject evaluated the new accounts. While the author judged the fit to be good, alternative representations could easily be suggested.⁴

Based on Payne's (1976) research, Biggs (1979) attempted to develop and test more objective criteria for discriminating between models with different functional forms which might be used in a financial analysis task. Eleven experienced financial analysts thought aloud as they selected the company with the highest earnings power from a group of five. Each company was represented by extensive multiperiod financial statements. The protocols were categorized as reflecting one of three types of operators and the operator sequences were interpreted as evidence of one of four processing models: additive compensatory, additive difference, conjunctive or elimination by aspects. At least one subject appeared to be using each of the rules. However, the different models usually led to the same conclusions. The additive compensatory and elimination by aspect models were most frequently observed. Subjects using the compensatory models took much more time to

⁴ In fact we would describe the process as involving a strong compensatory component as evidenced by the fact that positive scores on other variables can offset failure to meet a criterion value. See Clarkson & Meltzer (1960) for an alternative representation (additive difference) which fits the output decisions at least as well as this model.

complete the task. An attempt to use a post-experimental questionnaire to validate the results was partially successful.

Bouwman (1980) compared expert and novice financial analysts in an attempt to determine the differences that education and experience produce in their decision making strategies. Fifteen accounting students and three professional accountants thought aloud while they analyzed four extensive financial cases to determine any underlying problem areas. The protocols were converted into problem behavior graphs which present knowledge states as nodes and operators as arrows between nodes. The difficulties in interpreting and summarizing protocol data and the small sample size precluded valid statistical comparisons. However a comparison of the graphs of a single student and accountant provided some potentially useful insights. The student appeared to follow a simple undirected sequential strategy where the information was evaluated in the order presented until a single problem was uncovered. Information was frequently examined based on very simple trends (e.g. sales are up). The information was used to form a series of simple relations which were internally consistent but may have been inconsistent with one another. When an observed fact was identified as a "problem", little additional information was gathered. On the other hand, the expert seemed to follow a standard checklist of questions. Data were often examined in terms of complex trends. He appeared to develop a general overall picture of the firm and classify it under a general category such as "expanding company" based on the initial information acquired. When the stereotype was violated, an in-depth examination to uncover significant causes would be initiated. The problems seemed to be recognized based on a set of common problems or hypotheses associated with patterns of cues in long-term memory.

In the final study of financial analysis, Stephens (1979) asked 10 bankers to think aloud while evaluating one of two commercial lending cases. He found that the lending officers spent a great deal of time computing and analyzing ratios and ratio trends. No evidence was available that adjustments were made for differences in inventory or depreciation method.

Managerial accounting

Shields (1980a,b) has begun to study the general strategies used by managers in performance report evaluation and the impact of certain attributes of task complexity on these strategies. Twelve executive MBA graduates thought aloud while they analyzed performance reports in order to estimate the cause of the observed behavior and to predict future behavior. The four cases differed in the number of responsibility centers and performance parameters which included standard accounting variances and nonaccounting data (e.g. absenteeism). Data was presented to the subjects on information boards which contained an envelope with data cards enclosed for each performance cue. By collecting the cards in the order chosen, an accurate measure of information search is also provided. The verbal protocols were coded into 15 categories. Goal statements appeared to direct information search during the first half of the process. This was followed by hypothesis generation which organized additional information search in the third quarter. The fourth quarter primarily involved development of causal attributions and predictions. A smaller percentage of the data was searched as the number of responsibility centers and performance parameters was increased. The variability in the percentage of information searched increased with the number of responsibility centers but not with the number of cues. Ex post measures of cue importance and order of presentation both affected search order.

Auditing

The audit sample selection research of Mock & Turner (1979) discussed earlier was extended in the first protocol analysis in audit decision making. Biggs & Mock's (1980) goal was to describe auditors' sample selection processes in terms of overall patterns and use of specific information and to make a preliminary comparison of the impact of experience on these patterns. In the experiment, two experienced and two inexperienced audit seniors thought aloud while they made sample size selections for Mock & Turner's (1979) detailed sample selection case. The recorded protocols were converted to flowcharts and abstracts for analysis. As in Bouwman's (1980) study, the

subject sample was too small for reliable comparisons. However, the authors suggest that the more experienced subjects employed a significantly different decision strategy from that of their less experienced counterparts. The experienced seniors appeared to build an overall picture of the company and then make the four required decisions. The two new seniors employed a serial strategy involving a separate search for information relevant to each decision. The results are similar to Bouwman's (1980) financial analysis study. All of the subjects attended to a much greater proportion of the available information than was indicated in the decision rationale memos produced by Mock & Turner's (1979) subjects. This suggests a potential audit documentation problem relating to lack of self-insight. There were also major between-subject differences in the sample size decisions and the proportion of the available information attended to.

In what is probably the most novel experiment discussed in this section, Weber (1980) tested whether there is consensus among EDP auditors in the way they structure computer controls in memory. This study is quite different from the others in that it is based on more established research in cognitive psychology and employs more traditional experimental designs and procedures. Weber hypothesized that the lack of consensus in internal control evaluations discovered in some studies was caused in part by differences in the way in which these cues are structured in memory. The ability of a group of expert EDP auditors to recall and properly cluster a series of EDP controls was compared with a student control group's performance on the same task. A list of 50 computer controls (10 from each of 5 categories) was read in random order to the 7 auditors and 6 student participants. Three seconds after the list was read, the subjects were instructed to recall and write down as many of the controls mentioned as possible. The auditors were able to recall significantly more controls than the students. Among the auditors, the external auditors outperformed the internal auditors. The auditors' clustering of the controls was also more similar to the *a priori* model than that of the students, indicating a significant degree of consensus. Frequency of

recall also provides a measure of cue importance, the results suggesting that "management and organizational controls" are more important than the other four categories (data preparation, input, processing and output).

Research contribution

The accounting studies of predecisional behavior are summarized in Table 3. Their preliminary nature precludes any general conclusions relating to information processing behavior or accounting policy issues. At the same time they do illustrate potentially useful directions for further research and areas in need of methodological improvement. Of particular interest to both accounting practitioners and educators are the studies by Biggs & Mock (1980) and Bouwman (1980) which attempted to determine the components of expertise. Such studies show promise for providing direction for education and continuing professional training and indicating the relative advantages of teaching different topics in the classroom or in the field. Attempts to tie these results to similar lens model studies (e.g. Slovic, Fleissner & Bauman, 1972 and Ashton & Kramer, 1980) may prove beneficial.

Biggs & Mock's (1980) finding that auditors appear to attend to a much greater number of cues than they actually use in making their decisions points out an important distinction between attention and use which helps to clear up a number of questions concerning decision makers' self insight. It also may suggest changes in audit documentation procedures which would provide more accurate records of cue processing. Weber (1980) has illustrated an innovative methodology and set the stage for developing a better understanding of auditor consensus. Biggs (1979) and Shields (1980a, b) have illustrated that decision rules with different surface structures are often used both by the same decision makers over time and across decision makers. Further, these different rules often produce the same solutions. This latter point is also illustrated by Clarkson (1962) and Clarkson & Meltzer (1960). The results suggest the problematic nature of determining the "true" decision rule (see Einhorn, *et al.* 1979) and the need for research which suggests when different rules will

TABLE 3. Summary of studies of predecisional behavior

Study	Type of decision maker	Task	Data collection and analysis methods	Variables of interest	Results
Biggs (1979)	11 financial analysts	Selection of company with highest earnings power among a group of 5 from extensive financial data	Verbal protocols, operator classification	Form of decision rule (IIB1) Decision speed (IIIA2)	Of four models, the additive compensatory and elimination by aspects were the models most often observed; the compensatory models took more time
Bowman (1980)	15 accounting students and 3 professional accountants	Analyze 4 cases containing extensive financial information to determine any underlying problem areas	Verbal protocols, problem behavior graphs	Prior experience (IIA4a) Heuristics (IIB4)	Student followed simple undirected sequential strategy; information was evaluated in order presented until a single problem was uncovered; expert followed directed strategy based on standard checklists, complex trends, and stereotypes; developed overall company picture
Stephens (1979)	10 bank loan officers	Evaluate one of two bank lending cases	Verbal protocols	Cue usage (IIB2)	Spent great deal of time computing and analyzing ratios and ratio trends; no adjustment made for differences in inventory or depreciation methods
Shields (1980a, b)	12 executive MBA graduates	Analysis of 4 performance reports for attribution of cause of behavior and expectancy of future managerial behavior	Verbal protocols and information boards	Number of cues (IIB1) Context (number of cases) (IE) Heuristics (search) (IIB4)	Goal statements directed information search during first half of process; followed by hypothesis generation, which led to further information search; followed by causal attributions and predictions; variability of search increased with number of cases

TABLE 3 contd.

Study	Type of decision maker	Task	Data collection and analysis methods	Variables of interest	Results
Biggs and Mock (1980)	2 experienced and 2 inexperienced audit seniors	To make sample selections for Mock and Turner's (1979) sample selection case	Verbal protocols, flowcharts	Prior experience (IIA.4a) Heuristics (IIB4)	Substantial difference between experienced and inexperienced Experienced built overall picture of the company; then made four decisions; inexperienced followed serial strategy involving separate search for information for each decision Attended to more cues than indicated
Weber (1980)	67 EDP auditors and 96 students	Recall EDP controls	Recall protocols and clusters	Consensus (IIIA 3b) Cue usage (IIB2)	Auditors recalled more cues than students and clustered cues more than a <i>priori</i> model Recalled more cues from "management and organizational controls" category

be selected and the significance of their selection for the accounting issue under study.

COGNITIVE STYLE

Most of the accounting studies in this section focus on the impact of personal characteristics of the decision makers on the qualities of the judgment or the impact of varying information load on qualities of the judgment. Many of the studies can be traced to Schroder *et al.* (1967), who theorized that the level of information processing of decision makers is a single-peaked function of the environmental complexity and that maximum processing takes place at some optimal level of complexity. Further development of this research by Driver & Lintott (1973) revealed that, across levels of complexity, individuals with different decision styles processed different amounts and different types of information.

Accounting studies based on these developments attempted to classify users of information by their cognitive structure and to design information systems best suited to the individual style of the decision maker. Constructs used to classify individuals included decision style and various personality measures. We noted in our previous review of this literature that the ambiguity of the results to date argued against the feasibility of tailor-made information systems and that perhaps new research approaches should be used to examine the link between cognitive structure and decision behavior. Although a few studies continue the search for a meaningful connection between cognitive structure and decision behavior, most of the recent work has emphasized the importance of task characteristics and cognitive structure as well as their interaction.

Cognitive structure

Vasarhelyi (1977) used a planning context to further explore the relationships between decision style and performance, information utilization and decision speed. In the experiment, 50 subjects (average 7.4 years of business experience) made business planning decisions using an elaborate case study and an interactive decision support system.

Subjects were classified as either heuristic or analytic using a test for cognitive style. The task included both structured and non-structured phases and both quantitative and qualitative data. Subject performance was measured by the ranking of plans by a panel of judges. Information utilization (kind and quantity) was measured by a self-report questionnaire. Results indicated no difference in performance (overall or structured vs. quantitative information). There was weak support for heuristics using less information overall and making faster decisions.

In an audit task, Weber (1978) examined the relationships of a personality measure (dogmatism), risk-taking propensity, and experience to the accuracy variability of auditors' decisions and the degree of confidence in those decisions. Of twelve hypotheses tested only three showed significant results and two of these three were in an *unexpected* direction. The only hypothesis confirmed was that the extent of audit plan decreased when risk-taking propensity increased.

Neither of these studies can be viewed as encouraging the idea of tailor-made information systems. Nor have they provided much insight into how people make decisions. As Vasarhelyi (1977) noted, the formidable measurement problems faced in this research area are the likely cause. These problems, coupled with the likelihood that any existing relationships are probably weak to begin with, further question the possibility of finding meaningful links between cognitive characteristics and overt behavior. The remaining studies in this section deal with this problem by more closely relating the research to underlying theories.

Differential peaking

Lusk (1979) argued that main effects alone do not imply that different systems should be provided for different users. Only an interaction between cognitive characteristics and information stimuli can justify individualized information systems. This interaction is often termed "differential peaking". In an experiment, undergraduate students were to complete a questionnaire, the answers to which could be abstracted from an information report. Each student received one of five tabular or graphic reports which represented

increasingly complex transformations of the same data. Individual difference groups were formed based upon performance on an embedded figures test. The mid-point of scores on this test divided the subjects into high and low analytic groups. As expected from cognitive field theory main effects were both significant. That is, high analytics significantly out-performed low analytics; and individuals using less complex reports significantly out-performed those using more complex reports. The interaction, however, was *not* significant and hence there was no evidence of differential performance peaking. In other words, each group achieved optimal results using the same reports. One important limitation of this study, noted by Lusk, is the lack of reports less complex than type A.

Benbasat & Dexter (1979), on the other hand, did find evidence of differential peaking. A total of 48 undergraduate students, faculty members and professional accountants were classified as either high or low analytics using the same embedded figures test used by Lusk (1979). The subjects participated in a multi-period game involving decisions of production level and inventory management. One half of each cognitive group received structured/aggregate information reports. The other half had access to a data base representing the collection of the raw data which formed the basis for the aggregated reports. Over 15 decision periods, high analytics significantly out-performed low analytics and required less time. This result is consistent with Lusk (1979). There was no main effect on performance for type of information system, but there was a significant interaction of cognitive type and information systems. Results indicate that structured reports were better for high analytics and the data-base inquiry system is better for the low analytics. Presumably, low analytics were unable to "break-up" the structured reports.

Overload

Because of recent trends in accounting policy which advocate expanded disclosure, two studies concentrated on the information overload question in financial reporting; i.e. can expanded disclosure have a negative effect on decisions. Casey

(1980a) asked experienced loan officers to make predictions of bankruptcy for each of 10 firms using one of three levels of information. Group I used a 3 year set of six financial ratios; Group II used the same ratios plus balance sheets and income statements. Group III was given notes to the financial statements in addition to the information given to Group II. Both a panel of experts and the subjects themselves agreed that the information loads for the three groups were significantly different. Loan officers in Group II were more accurate and spent no more time than those in Group I. Bankers in Group III, however, used more time but were no more accurate than Group II. The notes to the financial statement may have represented an overload of information. However, they may have merely lacked additional information content.

Snowball (forthcoming) extended this area of inquiry by including the effect of user expertise as well as different levels of information load. Students with three levels of accounting training (the expertise manipulation) were asked to estimate next period's cash flows of a company, using current financial reports. Information load was defined in two ways: level of disclosure (detailed vs summarized footnotes) and time allowed to complete the task (restricted, moderate, unlimited time). Significant results indicated that increasing expertise was associated with less confidence and a wider dispersion of point estimates. With respect to information load, the only significant effect showed that reduction in available time increased the subject's confidence in the estimate. One apparent weakness in the study was the failure of the information load manipulation. Subject perceptions of the disclosure and time conditions did not reflect significantly different levels. This would tend to make the experimental results difficult to interpret.

Other issues

Some of the conceptual problems discussed in this section are addressed in the remaining papers, all of which provide more complete research frameworks and two of which emphasize the importance of task characteristics. Dirsmith & Lewis (1980) note that there are many focal points of

decision research and that it may be unrealistic to try to directly link cognitive characteristics with behavior. They point out, in contrast to prior accounting studies, that the psychological research upon which these studies are based indicates that cognitive style is relevant at the perceptual subsystem as opposed to the executive subsystem. Rather than trying to explain behavior, Dirsmith & Lewis attain conceptual congruency by relating a perceptual independent variable (level of ambiguity intolerance) with a perceptual dependent variable (perception of information use). In their study, they sought to relate cognitive style to the information inductance hypothesis (Prakash & Rappaport, 1977) which they defined as a pre-decisional orientation toward the use of financial accounting information in internal decision making. Industrial business managers and auditors were classified as cognitively closed or open based upon response to an ambiguity intolerance scale. Subjects were given an extended strategic planning case study and were asked to complete a questionnaire which assessed their perceptions of the extent to which both external and internal parties rely on financial accounting information in their decision making. Significant support was found for hypotheses that predicted that cognitively closed individuals are more likely to perceive external users as relying primarily on financial accounting information and more likely to perceive managers as emphasizing such information in their own decision making, particularly in the evaluation of alternative solutions. Failure to find similar relationships in studies seeking to explain behavior may simply be attributable to the fact that the relationships are swamped by the effects of contextual variables, different utility functions, different decision rules or other differences.

Pratt & Waller (1979) used a Skinnerian reinforcement contingency model to highlight the person-task interaction. In this model the stimulus (task, information load) interacts with the reinforcement history of the decision maker and activates information processing. This process involves information search and the conceptual organization of information. Varying levels of complexity of conceptual organization of information (termed conceptual level) are contained in

the information processing system of the model and, rather than remaining a constant personal characteristic, conceptual level is contingent on the nature of the task. The remainder of the model involves a decision rule, a behavioral response and an updating of the reinforcement history.

In an experimental application of the model, Pratt & Waller had ninety evening students predict earnings based on one of three annual reports which represented three levels of stimulus complexity. The subjects were classified by a personality measure, occupation and investment experience. Conceptual level was measured by a multi-dimensional scaling of similarity judgments made by the subjects with respect to the information content of twelve sections of the annual report. Information search (use) was measured by a self-report allocation of weights to the twelve report sections. In line with the hypothesis, variation in conceptual level was explained, in order, by complexity of the stimulus, investment experience, occupation and personality (however, only the complexity main effect was significant). Correlation analysis also showed significant but weak links between conceptual level and perceived use of information, between conceptual level and earnings predictions and between perceived use of information and earnings predictions. The statistical weakness of the results likely reflects the inability of the operational measures to fully capture the complex constructs involved in the model. The authors emphasize this point with respect to nearly all components of the model.

Research contribution

It would appear that the difficulties faced in the search for a direct link between personality or cognitive structure and decision behavior are more a reflection of the complexity of the relationships involved than a depreciation of prior research. Neither Vasarhelyi (1977) nor Weber (1978) could establish a reasonable relationship between cognitive measures and decision behavior. As we noted earlier, this does not mean the relationships are non-existent. They may simply be swamped by other, unmeasured, intervening variables. While Lusk (1979) noted that a necessary condition for the development of tailor-made information

systems, that of differential performance peaking, was not evidenced in his study, Benbasat & Dexter (1979) did find such differential peaking. And, consistent with earlier research, both of these studies showed high analytics out-performing low analytics.

Two studies looked at the issues of information overload. While this issue is of practical importance, related research appears to be hampered by poor definition of concepts. Casey (1980b) called for methodological improvements in information load studies by refining the definitions and measurements of key variables in the model. Specifically, he suggests validation of the load manipulations, multi-dimensional measures of load and more attention to the effects of load on the *processing* of information.

The less than overwhelming results of earlier research have led some researchers in new directions. Dirsmith & Lewis (1980) stopped short of trying to explain behavior by linking a cognitive independent variable to a cognitive dependent variable. Their efforts are consistent with the cognitive style background literature which predicts that cognitive characteristics have their impact on the perceptual rather than the executive subsystem. Also Pratt & Waller (1979) have formulated a reinforcement contingency framework for the integrated study of the processor, the task, their interaction and their effect on the behavioral responses. Although the results of these studies are not particularly strong, the frameworks are somewhat promising.

DIRECTIONS FOR FUTURE RESEARCH

Certain generalizations about human decision making are emerging from the accounting and psychology literatures. Much of the lens model research has indicated that accountants and other experts may not be as proficient at certain aspects of decision making as once was thought. Inaccuracies appear to result from both inconsistency in application of decision rules and misweighting of evidence. The probabilistic judgment literature has suggested that misweighting of evidence results from use of simplified decision rules often called

heuristics. However, heuristic use seems to be very sensitive to seemingly unimportant task attributes. Studies of predecisional behavior indicate that these context effects are related to basic cognitive properties such as the structure of short and long term memory which interact with attributes of the task to determine how problems are represented and solved. Problem representations based on intuitive causal models or frames seem to drive many decision making strategies.

An apparent merging of these three research approaches has accompanied two trends in the study of decision making. In the past four years, studies in accounting and psychology have become more theoretical as attempts are made to structure theories around basic cognitive processes. At the same time, the research has reflected increasing concern for the external validity and application of results. While these directions may at first appear contradictory, they are in fact closely related. The failures of more *ad hoc* theories to predict significant context effects discovered in recent applied research has demonstrated the need to search for more basic principles related to higher order mental processes such as research into the role of problem representation in learning and judgment (e.g. Tversky & Kahneman, 1980 and Einhorn & Hogarth, 1981).

The impact on accounting practice

Questions concerning the impact of accounting research on practice are a favorite topic for accounting conferences (see e.g. Abdel-khalik & Keller, 1978). When LL-77 was written, academic accountants were only beginning to become aware of human information processing and decision making research. A very different picture emerges today. The results of this research appear to be having a direct effect on accounting practice, particularly in auditing and financial analysis.

Behavioral decision theory has contributed to two recent trends in audit decision making: (1) the use of "expert measurement and mechanical combination" (Einhorn, 1972) to perform certain audit decisions and (2) the use of statistical sampling and regression for analytical review. Lens model research suggesting that different staff members make widely differing decisions in the

same circumstances have been of increasing concern to auditors (see Holstrum, 1980). Given the lack of criteria for judging the accuracy of many audit decisions, the notion of "general acceptance" or consensus among experts has become the most often cited criterion for judging decision quality. While consensus does not insure accuracy, its absence may be interpreted as *prima facie* evidence of inaccuracy by courts and regulatory authorities. This concern increases if one believes that more costly errors are caused by highly idiosyncratic decisions. More systematic review procedures and detailed procedure manuals were early responses to this problem. Most recently, a number of firms are developing more formal consistency aiding devices which rely on research relating to the causes of decision error and the ability to model judgment.

Einhorn (1972) demonstrated that considerable error is often added to decisions during the act of combining individual factors into a global judgment. In response, he suggested uniting the experts' ability to select and code non-numerical cues with a mechanical method for combining cues. In his application to a medical diagnosis problem, decisions were made by having physicians evaluate a series of attributes which had been preselected by an expert panel and then combining these ratings into a global judgment using a mathematical model. A very similar approach is now being employed by a number of large CPA firms in audit decision making. The evidence cited earlier concerning auditor consensus suggests that auditors substantially *agree* on judgments of attributes such as internal control quality but disagree on the way in which these attributes are combined into audit decisions. By replacing men with models in this latter part of the process, Einhorn's approach should eliminate this major source of disagreement. In practice, a variety of attribute judgments such as internal control quality, the required level of assurance, the results of analytical review, the expected amount of monetary error, and materiality are input into decision tables or equations which combine these factors with statistical sampling models to determine the extent of substantive tests. The weighting rules implicit in the tables or equations are usually

determined by senior policy makers or by observing average behavior in the firm. While these approaches are often criticized for their arbitrariness, they can be no more arbitrary than the individualistic procedures they replace and are likely to reduce significant causes of error.

The most important impact of probabilistic judgment research has been to speed the acceptance of statistical techniques for sampling and analytical review. Findings that auditors are poor intuitive statisticians, failing to understand the implications of simple concepts such as the relationship between sample variability and sample size and between fractile estimates and error risk, are being taken seriously by auditing policy makers. At least one firm has incorporated materials illustrating common errors into their training programs and a number of firms have responded by requiring use of statistical sampling in many situations. The recently issued exposure draft audit standard entitled *Statistical Sampling* (AICPA, 1980) moves Generally Accepted Auditing Standards substantially in this direction.

The ability of models of man to eliminate inconsistency and of environmental regression models to eliminate both inconsistency and misweighting in intuitive judgments has not gone unnoticed in the area of financial analysis. In consumer credit analysis, environmental models have completely replaced human loan officers at many institutions. Bankruptcy models (e.g. Altman *et al.* 1977) are being used by commercial bankers in their lending decisions and by auditors in their going concern evaluations. Tax officials and regulatory authorities use similar models in tasks ranging from selection of tax returns for audit to classifying problem banks. While some of these applications actually preceded much of the judgment research, recent findings have increased their acceptance among many.

While these initial effects are receiving increased recognition, a number of new directions promise even greater contributions. As the field matures, both suggestions and predictions for future change become more tenuous. With this caution in mind, the remainder of this section will delineate what we believe to be constructive directions for future research. These, of course, are incomplete and are

limited by our imagination, biases and interests.

Lens model research

Studies of cue weighting and consensus continue to be of major interest to academics and practitioners. Cue weighting studies meet the increasing need to make judgmental policies explicit, communicate these policies to others, and evaluate adherence to stated policies. Consensus studies often indicate potential problem areas in need of further study. While internal control evaluation and its relationship to audit planning continues to receive attention it appears that inadequate attention has been directed to planning the nature of the extensions. This should not be a random process. Attempts to refine prior results and eliminate alternative hypotheses require careful development of the conceptual basis of the study. Also, relatively few novel applications have been attempted. Contributions can continue to be made in other areas of audit and managerial decision making in both abstract and more realistic settings. Further, regression related measures should be combined with other measurement techniques designed to measure predecisional behavior to adequately account for the numerous dimensions of "cue usage".

Behavioral decision theory research is often presented as an alternative to earlier experimental studies of accounting policy related issues. This contrast appears to have deterred applications of these more powerful research techniques and psychological theories to accounting and reporting issues. Many suggestions for changes in accounting and disclosure such as recent suggested changes in the standard audit report rest on assumptions about the effects of these changes on individual behavior. Both normative models and psychological theories can help refine these predictions and the related methodologies can be used to test the predictions in settings where archival data is unavailable.

Lens model type descriptive studies have provided the groundwork for the development of a number of decision aids to be used in accounting contexts. However, accounting researchers have had little direct involvement in their development or testing. The literature suggests that relatively

minor changes in the way in which judgments are elicited can result in radically different judgments (see e.g. Einhorn & Hogarth, 1981). Research investigating these potential effects could help ensure the desired outcome of decision aid applications. Researchers should also be involved in the testing of aids to determine whether they have produced the desired result.

More basic accounting research into the effects of information characteristics such as cue interrelationships (e.g. Gibbins, 1980) and data format (e.g. Moriarity, 1979) on learning and decision accuracy show great promise for developing a scientific basis for managerial accounting and information systems design. It appears as though our ability to produce information has far outstripped human abilities to process information. This research should be closely related to studies of basic processes such as problem recognition, hypothesis generation and information search discussed in the predecisional behavior section.

A number of recent studies (e.g. Rockness & Nickolai, 1977 and Brown, 1981) have opened the door for the use of lens model related techniques in the analysis of archival decision data. Most accounting researchers appear to have forgotten that a number of classic studies in the psychological literature such as Dawes' (1971) graduate admissions study analyzed archival information resulting from actual decisions. While such studies create additional problems related to data availability and experimental control their ability to address severe external validity problems in some situations makes the approach attractive.

Our final suggestion relates to the types of accounting issues addressed in previous studies. Audit and financial analysis applications have dominated existing research. However, related problems are encountered in the managerial accounting area. Decisions relating to planning and control have many of the same characteristics as these other tasks. Harrell's (1977) study of performance evaluation presents only one of many possible examples of contributions to this field. Recent interest in the study of the accounting policy making process also provides important opportunities for decision researchers. In fact, decision researchers appear to be uniquely qualified

TABLE 4. Summary of cognitive style studies

Study	Type of decision maker	Task	Variables of interest	Results
Vasarhelyi (1977)	50 students and businessmen	Make business planning decisions in multi-period financial planning game	Type of task (IE3a) Cognitive structure (IIA3c) Accuracy (IIIA1) Speed (IIIA2) Subjective cue usage (IIIB1)	Cognitive style could not explain differences in performance or use of qualitative vs quantitative information Heuristic decision makers used less information and made faster decisions
Weber (1978)	40 practicing CPAs	Assess possible distribution of dollar errors in inventory based on case materials	Type of task (E3a) Cognitive structure (IIA3c) Accuracy (IIIA1) Speed (IIIA2)	Use of decision aid increased performance, but required more time Dogmatism and risk-taking could not explain differences in performance
Lusk (1979)	300 undergraduates	Answer questions from tabular and graphic reports	Format (ID1) Cognitive structure (IIA3c) Accuracy (IIIA1)	High analytics outperformed low analytics Both groups performed best on least complex report. No significant interaction of report and cognitive structure
Benbasat and Dexter (1979)	48 students, faculty, professional accountants	Make production and inventory management decisions in multi-period game	Aggregated vs disaggregated (ID3) Cognitive structure (IIA3c) Accuracy (IIIA1) Speed (IIIA2)	High analytics outperformed low analytics and took less time Significant interaction of information system and cognitive structure
Casey (1980a)	122 bank loan officers	Make ten predictions of bankruptcy judgments	Number of cues (IB1) Accuracy (IIIA1) Speed (IIIA2)	Subjects with medium and high loads outperformed low information load High load took more time but did not outperform medium load

TABLE 4 contd.

Study	Type of decision maker	Task	Variables of interest	Results
Snowball (forthcoming)	30 intermediate-level 30 senior accounting majors and 30 recent accounting graduates	Estimate next period's operating cash flows from annual report data	Aggregated vs disaggregated (ID3) Task characteristics-time allowed (IE3) Prior experience (IIA4a) Consistency (IIIA3a) Perceived decision quality (IIIB2)	Reduced time increased subjects' confidence. No other significant main effects or interactions
Dirsmith and Lewis (1980)	144 manufacturing managers and auditors	Respond to questions concerning importance of financial accounting information in decisions made in case study	Cognitive structure (IIA3c)	Significant relation between perceived use of financial accounting information by external DMs and perceived importance for internal decision making. Subjects intolerant of ambiguity perceive more external reliance on accounting information
Pratt and Waller (1979)	90 evening accounting students	Make earnings prediction based on annual report	Aggregated vs disaggregated (ID3) Personality (IIA3b) Cognitive structure (IIA3c) Prior experience (IIA4a) Accuracy (IIIA1) Subjective cue usage (IIIB1)	Conceptual level related to complexity of stimulus. Perceived information search and conceptual level correlate with behavioral response (but weakly)

to conduct the positive research necessary to understand these processes better.

Probabilistic judgment

One of the major areas of interest to accounting researchers has been the elicitation of subjective probability. In particular, attention has been directed at a comparison of various elicitation methods. In a sense, this attention has been somewhat premature in that there is no natural benchmark against which to compare the methods. As we noted earlier, the convergence test of alternative methods is difficult to interpret and the use of classical sample sizes as a comparison measure is somewhat arbitrary. These problems are complicated further by the fact that the mapping of any of the error measures to utilities is far from clear. Existing analytical research provides behavioral researchers with little guidance. We may not sufficiently understand the decision theory approach to auditing to draw the crucial linkages between this research and practice applications.

A promising approach may be that suggested by Solomon, *et al.* (1980). As an appropriate benchmark, they suggest the measures of "goodness" employed by applied decision researchers in other disciplines. These empirical measures of goodness include mathematical scoring rules, calibration and sufficient extremeness of probability distributions. The elicitation methods that maximize these empirical measures could be viewed as "better" in a more meaningful way.

A second recommendation in the area of elicitation is to study the effect of *extensive* training. Much of the variation in elicited probabilities may simply be the result of the inconsistency born of ignorance. It should take more than a few hours to become comfortable with many of these elicitation techniques.

The study of heuristics and biases continues to be particularly popular because of their many implications for practitioners, especially in auditing. To a great extent, the motivation for studies of heuristic processing is the eventual development of decision aids to eliminate resulting biases. Unfortunately, the task of identifying generalized use of particular heuristics has not been as easy as might have been expected from the early studies in

psychology. It would appear that minor departures from the contexts of the classic experiments of Tversky & Kahneman (e.g. 1974) produce different results.

Future research in heuristics and biases should address the two critical assumptions of the decision-improvement motivation of the research. The first assumption is that a normative model is an appropriate benchmark from which to measure judgment bias. The second assumption is that we can obtain valid experimental evidence of the existence of common heuristics and of the conditions in which they might be used.

Einhorn & Hogarth (1981) point out the ambiguity of the interpretation of departure from normative solutions. Normative responses are generated from a context-free representation of a task environment. Human responses, on the other hand, are generated from an individual's representation of the task environment. Unless we can determine which is the better representation of the environment, comparisons of human and normative responses give ambiguous results. This problem actually leads to the difficulty with the second assumption, that of experimental validity. Since a decision maker's response is generated from his or her representation of the task, the design of our experiments and our interpretation of results must take into account possible alternative representations. Most of our experimental designs are structured using the normative decision theory model. Consequently, the success of our variable manipulations will often be dependent upon the congruence of the decision maker's view of the problem and the normative model. To enhance our knowledge of the individual's task representation we may need to turn to psychological research on more basic cognitive processes. As a supplemental source of information, process tracing techniques may prove to be quite valuable in this area.

Finally, we believe it is essential to move toward the study of more realistic experimental situations. Audit decisions are not made *in vacuo* and there is little reason to believe that brief vignettes will capture the important decisions in which we are interested. Again, the work of Elstein *et al.* (1978) in medical problem solving

provides an excellent example of the high fidelity achievable in a controlled experiment.

Predecisional behavior

The psychological theories, research methods and accounting applications of this approach are all at early stages of development. However, it is becoming clear that consideration of more basic cognitive characteristics such as the organization and capacity of memory will be necessary to complete the model of human decision making. The interaction of cognitive and task attributes appears to affect the way in which problems are represented in memory. The cognitive representation of the task, in turn, determines the way in which the problem is solved. As Einhorn & Hogarth (1981) note, this view is reflected in recent theories of probabilistic judgment (Tversky & Kahneman, 1980), similarity judgments (Tversky & Sattath, 1979) and preference reversals in gambling behavior (Grether & Plott, 1979).

Accounting studies in this area may also help fill gaps in existing research. For example, research examining the memory of experts might indicate explanations for differences between experts and novices demonstrated in prior research and might lead to development of training aids. The role of cognitive representation in choice of decision heuristics may provide insights into methods for redesigning management reports or audit programs to lead to proper heuristic choice. Studies of the interaction of memory and information search may lead to development of decision aids to be used at these important stages in less structured accounting situations such as variance investigation and audit client screening.

The related measurement techniques can also be used in exploratory studies of decision situations such as the audit client screening decision where little literature is available to guide the researcher. Measures of information search and attention can also assess additional dimensions of cue importance and possibly explain apparent differences between self-report and statistical measures of cue importance. Einhorn *et al.* (1979) discuss the meaning of various measures of cue importance in detail.

For this developing area to reach its full

potential, a number of pitfalls must be avoided. First, the ability to deal with less structured tasks may lead to the erroneous conclusion that less attention is necessary to the structure imposed by the principles of experimental design. All experimentation requires adherence to the principles of experimental design to insure the internal validity of results. Second, productive accounting research requires a well defined research objective and explicit consideration of the contribution of the research to the field. Most decision research in accounting and other applied disciplines is directed at the *improvement* of decisions. Even in policy capturing studies (e.g. Joyce, 1976) where results were purely descriptive, the desired end result of the research program is usually prescriptive. This focus is consistent with the majority of lens model and probabilistic judgment research in psychology where emphasis is placed on achievement and its components or causes. In contrast, psychological studies of problem solving aim to describe the dynamic processes involved and the supporting knowledge base. Often, little effort is expended in trying to relate these processes to performance measures. We believe that to reach its full potential, accounting researchers must move problem solving research in the direction of the accountants' interest in the *improvement* of decisions. Elstein *et al.* (1978) provide a good example to follow in this regard. Third, researchers should follow Weber's (1980) lead and take advantage of a broader spectrum of theory and methodology in cognitive psychology. Innovative combinations of knowledge from different subdisciplines may provide larger increments to our knowledge of accounting problems.

We also suggest that initial enthusiasm for the associated methodologies of protocol analysis and information search analysis be constrained by a number of cautions. First, even though the resulting data may provide more detailed sequential information relating to process, this does *not* imply a perfect matching of data and process. Unfortunately, this common misunderstanding is fostered by the label "process tracing" often associated with the techniques. Protocols at best provide an incomplete record of the contents of short-term memory. Further, the meaning of this

record is usually less than obvious. For example, the fact that a cue has been verbalized tells little about its importance to a decision process. Not only may this cue not be relied on in reaching a decision, but another unmentioned cue may have a significant effect. The greater level of detail provided by the resulting models should not be interpreted as indicating that they represent mental processes.

The relatively short history of their use and their added complexity also creates a unique set of methodological problems. Four problems are of particular importance. First, the validity of protocol data has been attacked on the grounds that people may not use it to gain access to their higher order mental processes (see e.g. Nisbett & Wilson, 1977 and Ericsson & Simon, 1980). This problem can be addressed in part by care in interpreting the meaning of protocol data. Second, objections have been raised concerning the objectivity of data coding methods — in particular those related to verbal protocols. The choices of coding categories, the choice of short phrases which serve as the unit of analysis and the assignment of phrases to categories are highly subjective. The relationship between original protocols and resulting computer programs are also often less than obvious. Third, tests of the goodness of fit of the resulting models usually require only that the computer model account for the verbalizations. The weakest form only requires that most protocols be “easily” coded within the coding scheme. More stringent tests which require predictions of protocols and final choices from independent samples are needed. Further, competing models should be tested for comparison purposes. Tests of between-group differences also are limited by the lack of well-developed statistical descriptors of protocol data. Researchers often must provide only intuitive comparisons of “typical” individuals. The inability to discriminate between reliable and unreliable responses contributes to these problems. Finally, both the volume of data produced and lack of simple statistical descriptors creates difficulties in the communication of results. Research reports involving use of these techniques are often

quite long and difficult to follow. Simpler experiments exhibiting great care in the operationalization of independent and dependent variables will go a long way toward solving these last two problems. These unresolved methodological problems reflect the early stage of development of this research. The studies reviewed here suggest that an important contribution can be made. However, given the developing nature of the discipline, an extra measure of great effort, forethought and attention to the tenets of scientific method is required.

Cognitive style

Despite extensive research conducted in this area, we know little of the role of cognitive structure in information processing. While some evidence exists to support the idea that differential performance in certain tasks is related to cognitive differences, only Benbasat & Dexter (1978) have found an interaction of cognitive structure and information system. Recognizing the measurement problems of identifying what may be weak relationships, researchers in accounting have begun alternative approaches to seek a better understanding of the effect of cognitive differences.

We recommend further development of theoretical frameworks of decision behavior which specify the cognitive components and how they interact with other components of the processing system. We also support the suggestions of Casey (1980b) that operational definitions of information load be amplified to include those variables in Fig. 1 relating to characteristics of the data set, experimental conditions such as context and time limitations and the nature of the task. Manipulation of these variables by the experiments must be perceived in the same way by the subjects. As Casey notes, knowledge of how the subject views the task may best be obtained through the analysis of predecisional behavior. However, further contributions to this area of accounting research may require further conceptual and operational developments in basic psychological research.

BIBLIOGRAPHY

- Abdel-khalik, A. R. & El-Sheshai, K., Information Choice and Utilization in an Experiment on Default Prediction, *Journal of Accounting Research* (Autumn 1980), pp. 325–342.
- Abdel-khalik, A. R. & Keller, T. (eds.), *Financial Information Requirements for Security Analysis* (Durham, N.C.: Duke University, 1978).
- AICPA, *Statistical Sampling*. Proposed Statement of Auditing Standards (1980).
- Altman, E. I., Haldeman, R. G. & Narayanan, P., ZETA Analysis, *Journal of Banking and Finance* (June 1977), pp. 29–54.
- Ashton, R. H., An Experimental Study of Internal Control Judgments. *Journal of Accounting Research* (Spring 1974), pp. 143–157.
- Ashton, R. H., Cognitive Changes Induced by Accounting Changes: Experimental Evidence on the Functional Fixation Hypothesis, *Studies on Human Information Processing in Accounting*, Supplement to *Journal of Accounting Research*, (1976), pp. 1–17.
- Ashton, R. H., A Descriptive Study of Information Evaluation, *Journal of Accounting Research* (Spring 1981), pp. 42–61.
- Ashton, R. H. & Brown, P. R., Descriptive Modeling of Auditor's Internal Control Judgments: Replication and Extension, *Journal of Accounting Research* (Spring 1980), pp. 1–15.
- Ashton, R. H. & Kramer, S. S., Students as Surrogates in Behavioral Research: Some Evidence, *Journal of Accounting Research* (Spring 1980), pp. 269–277.
- Bamber, E. M., Expert Judgment in the Audit Team: An Examination of Source Credibility. *Unpublished Manuscript*. The Ohio State University, 1980.
- Bar-Hillel, M., The Role of Sample Size in Sample Evaluation, *Organizational Behavior and Human Performance* (October 1979), pp. 245–257.
- Benbasat, I. & Dexter, A., Value and Events Approaches to Accounting: An Experimental Evaluation, *The Accounting Review* (October 1979), pp. 735–749.
- Biddle, G. C. & Joyce, E. J., The Role of Sample Size in Probabilistic Inference in Auditing. *Unpublished Manuscript*. University of Chicago, December 1979.
- Biddle, G. C. & Joyce, E. J., Heuristics and Biases: Some Implications for Probabilistic Inference in Auditing, in *Symposium on Auditing Research, IV* (Urbana, Illinois: University of Illinois, 1981).
- Biggs, S. F., An Empirical Investigation of the Information Processes Underlying Four Models of Choice Behavior, in T. J. Burns (ed.), *Behavioral Experiments in Accounting II* (Columbus, Ohio: College of Administrative Science, Ohio State University, 1979).
- Biggs, S. F. & Mock, T. J., Auditor Information Search Processes in the Evaluation of Internal Controls. *Working Paper 2-80-6*. University of Wisconsin-Madison, February 1980.
- Boatsman, J. & Robertson, J., Policy Capturing on Selected Materiality Judgments, *The Accounting Review* (April 1974), pp. 342–352.
- Bouwman, M. J., The Use of Accounting Information: Expert Versus Novice Behavior. *Unpublished Manuscript*. University of Oregon, April 1980.
- Brown, C., Human Information Processing for Decisions to Investigate Cost Variances. *Unpublished Manuscript*. University of Illinois, 1980.
- Brown, P. R., A Descriptive Analysis of Select Input Bases of the Financial Accounting Standards Board, *Journal of Accounting Research* (Spring 1981), pp. 62–85.
- Brunswik, E., *The Conceptual Framework of Psychology* (Chicago: University of Chicago Press, 1952).
- Brunswik, E., Representative Design and Probabilistic Theory in a Functional Psychology, *Psychological Review* (May 1955), pp. 193–217.
- Casey, C. J., Variation in Accounting Information Load: The Effect on Loan Officers' Predictions of Bankruptcy, *The Accounting Review* (January 1980a), pp. 36–49.
- Casey, C. J., The Cognitive Effects of Variation in Information Load: Directions for Future Research. *Unpublished Manuscript*. Harvard University, 1980b.
- Casey, C. J., Additional Evidence on the Usefulness of Accounting Ratios for the Prediction of Corporate Failure, *Journal of Accounting Research* (Autumn 1980c), pp. 603–613.
- Chernoff, H., The Use of Faces to Represent Points in K-Dimensional Space Graphically, *Journal of the American Statistical Association* (June 1973), pp. 361–368.
- Chesley, G. R., The Elicitation of Subjective Probabilities: A Laboratory Study in an Accounting Context, *Journal of Accounting Research* (Spring 1976), pp. 27–48.

- Chesley, G. R., Subjective Probability Elicitation: Congruity of Datum and Response Mode, *Journal of Accounting Research* (Spring 1977), pp. 1–11.
- Chesley, G. R., Subjective Probability Elicitation Techniques: A Performance Comparison, *Journal of Accounting Research* (Autumn 1978), pp. 225–241.
- Clarkson, G. P. E., *Portfolio Selection: A Simulation of Trust Investment* (Englewood Cliffs, N.J.: Prentice-Hall, 1962).
- Clarkson, G. P. E. & Meltzer, A. H., Portfolio Selection: A Heuristic Approach, *Journal of Finance* (December 1960), pp. 465–480.
- Corless, J., Assessing Prior Distributions for Applying Bayesian Statistics in Auditing, *The Accounting Review* (July 1972), pp. 556–566.
- Crosby, M., Implications of Prior Probability Elicitation on Auditor Sample Size Decisions, *Journal of Accounting Research* (Autumn 1980), pp. 585–593.
- Crosby, M., Bayesian Statistics in Auditing: A Comparison of Probability Elicitation Techniques, *The Accounting Review* (April 1981), pp. 355–365.
- Danos, P. & Imhoff, E. A., Auditor Review of Financial Forecasts: An Analysis of Factors Affecting Reasonableness Judgments, *The Accounting Review* (January 1982), in press.
- Dawes, R. M., A Case Study of Graduate Admissions: Application of Three Principles of Human Decision Making, *American Psychologist* (February 1971), pp. 180–188.
- Dawes, R. M., The Robust Beauty of Improper Linear Models in Decision Making, *American Psychologist* (July 1979), pp. 571–582.
- Dawes, R. M. & Corrigan, B., Linear Models in Decision Making, *Psychological Bulletin* (January 1974), pp. 95–106.
- Demski, J., Information Improvement Bounds, *Journal of Accounting Research* (Spring 1972), pp. 58–76.
- Dirsmith, M. & Lewis, B., The Cognitive Style of Information Inductance. *Unpublished Manuscript*. University of Pittsburgh, 1980.
- Driver, M. & Lintott, L., *Managerial Decision Diagnostics* (Los Angeles, Cal.: Graduate School of Business Administration, University of Southern California, 1973).
- Dyckman, T. R., The Investigation of Cost Variance, *Journal of Accounting Research* (Autumn 1969), pp. 215–244.
- Ebert, R. J. & Kruse, T. E., Bootstrapping the Security Analyst, *Journal of Applied Psychology* (February 1978), pp. 110–119.
- Einhorn, H. J., Expert Measurement and Mechanical Combination, *Organizational Behavior and Human Performance* (February 1972), pp. 86–106.
- Einhorn, H. J., Synthesis: Accounting and Behavioral Science, *Studies on Human Information Processing in Accounting*, Supplement to *Journal of Accounting Research* (1976), pp. 196–206.
- Einhorn, H. J. & Hogarth, R. M., Unit Weighting Schemes in Decision Making, *Organizational Behavior and Human Performance* (April 1975), pp. 171–192.
- Einhorn, H. J. & Hogarth, R. M., Behavioral Decision Theory: Processes of Judgment and Choice, *Annual Review of Psychology* (1981), pp. 53–88.
- Einhorn, H. J., Kleinmuntz, D. N. & Kleinmuntz, B., Linear Regression and Process Tracing Models of Judgment, *Psychological Review* (1979), pp. 465–485.
- Elstein, A. S., Clinical Judgment: Psychological Research and Medical Practice, *Science* (November 1976), pp. 696–700.
- Elstein, A. D., Shulman, L. E. & Sprafka, S. A., *Medical Problem Solving: An Analysis of Clinical Reasoning* (Cambridge, Mass.: Harvard University Press, 1978).
- Ericsson, K. A. & Simon, H. A., Verbal Reports as Data, *Psychological Review* (May 1980), pp. 215–251.
- Felix, W. L., Evidence on Alternative Means of Assessing Prior Probability Distributions for Audit Decision Making, *The Accounting Review* (October 1976), pp. 800–807.
- Gibbins, M., Some Analytical Evidence on the Problem of Choosing Among Statistically Interrelated Sources of Data in Making Judgments. *Unpublished Manuscript*. University of British Columbia, July 1980.
- Gibbs, T. E. & Schroeder, R. G., Evaluating the Competence of Internal Audit Departments, in *Symposium on Auditing Research III* (Urbana, Ill.: Dept. of Accountancy, University of Illinois, 1979).
- Green, P. E. & Wind, Y., *Multiattribute Decisions in Marketing: A Measurement Approach* (Hinsdale, Ill.: Dryden Press, 1973).

- Grether, D. M. & Plott, C. R., Economic Theory of Choice and the Preference Reversal Phenomenon, *American Economic Review* (September 1979), pp. 623–638.
- Hamilton, R. E. & Wright, W. F., The Evaluation of Internal Controls Over Payroll. *Unpublished Manuscript*. University of Minnesota, 1977.
- Hammond, K. R. & Stewart, T. R., The Interaction between Design and Discovery in the Study of Human Judgment, in *Program of Research on Human Judgment and Social Interaction* (Institute of Behavioral Science, University of Colorado, Paper No. 152, 1974).
- Harrell, A. M., The Decision-Making Behavior of Air Force Officers and the Management Control Process, *The Accounting Review* (October 1977), pp. 833–841.
- Harrell, A. M. & Klick, H. D., Comparing the Impact of Monetary and Nonmonetary Human Asset Measures on Executive Decision Making, *Accounting Organizations and Society* (1980), pp. 393–400.
- Hilton, R. W., Swieringa, R. J. & Hoskin, R. E., Perception of Accuracy as a Determinant of Information Value, *Journal of Accounting Research* (Spring 1981), pp. 86–108.
- Hirsch, M., Disaggregated Probabilistic Accounting Information: The Effect of Sequential Events on Expected Value Maximization Decisions, *Journal of Accounting Research* (Autumn 1978), pp. 256–269.
- Holstrum, G. L., Audit Judgment Under Uncertainty: Empirical Evidence and Implications for Audit Practice. *Unpublished Manuscript*. Deloitte Haskins and Sells, New York, June 1980.
- Holt, R. N. & Carroll, R. J., Classification of Commercial Bank Loans Through Policy Capturing, *Accounting Organizations and Society* (1980), pp. 285–296.
- Joyce, E. J., Expert Judgment in Audit Program Planning, *Studies in Human Information Processing in Accounting*, Supplement to the *Journal of Accounting Research* (1976), pp. 29–60.
- Joyce, E. J. & Biddle, G. C., Anchoring and Adjustment in Probabilistic Inference in Auditing, *Journal of Accounting Research* (Spring 1981a), pp. 120–145.
- Joyce, E. J. & Biddle, G. C., Are Auditors' Judgments Sufficiently Regressive? *Journal of Accounting Research* (Autumn 1981b).
- Kessler, L. & Ashton, R. H., Feedback and Prediction Achievement in Financial Analysis, *Journal of Accounting Research* (Spring 1981), pp. 146–162.
- Kinney, W. R., A Decision Theory Approach to the Sampling Problem in Auditing, *Journal of Accounting Research* (Spring 1975), pp. 117–132.
- Kinney, W. R. & Uecker, W. C., Overcoming the Effects of the Anchoring and Adjustment Heuristic in Audit Judgments: An Experiment. *Unpublished Manuscript*. University of Iowa, December 1979.
- Lewis, B. L., Expert Judgment in Auditing: An Expected Utility Approach, *Journal of Accounting Research* (Autumn 1980), pp. 594–602.
- Libby, R., The Use of Simulated Decision Makers in Information Evaluation, *The Accounting Review* (July 1975a), pp. 475–489.
- Libby, R., Accounting Ratios and the Prediction of Failure: Some Behavioral Evidence, *Journal of Accounting Research* (Spring 1975b), pp. 150–161.
- Libby, R., Man Versus Model of Man: Some Conflicting Evidence, *Organizational Behavior and Human Performance* (June 1976a), pp. 1–12.
- Libby, R., Discussion of Cognitive Changes Induced by Accounting Changes: Experimental Evidence on the Functional Fixation Hypothesis, *Studies on Human Information Processing in Accounting*, Supplement to *Journal of Accounting Research* (1976b), pp. 18–24.
- Libby, R., Bankers' and Auditors' Perceptions of the Message Communicated by the Audit Report, *Journal of Accounting Research* (Spring 1979a), pp. 99–122.
- Libby, R., The Impact of Uncertainty Reporting on the Loan Decision, *Studies on Auditing – Selections from the Research Opportunities in Auditing Program*, Supplement to *Journal of Accounting Research* (1979b), pp. 35–57.
- Libby, R., *Accounting and Human Information Processing: Theory and Applications* (Englewood Cliffs, N.J.: Prentice-Hall, 1981).
- Libby, R. & Fishburn, P. C., Behavioral Models of Risk Taking in Capital Budgeting, *Journal of Accounting Research* (Autumn 1977), pp. 272–292.
- Libby, R. & Lewis, B. L., Human Information Processing Research in Accounting: The State of the Art, *Accounting, Organizations and Society* (1977), pp. 245–268.
- Lichtenstein, S., Fischhoff, B. & Phillips, L. D., Calibration of Probabilities: The State of the Art, in Jungermann, H. & de Zeeuw, G. (eds.) *Decision Making and Change in Human Affairs* (Dordrecht-Holland: Riedel, 1977), pp. 275–324.

- Lusk, E. J., A Test of Differential Performance Peaking for a Disembedding Task, *Journal of Accounting Research* (Spring 1979), pp. 286–294.
- Magee, R. P. & Dickhaut, J. W., Effect of Compensation Plans on Heuristics in Cost Variance Investigations, *Journal of Accounting Research* (Autumn 1978), pp. 294–314.
- Marchant, G. A., Data Fixity: Some Further Empirical Evidence. *Unpublished Honors Thesis*. University of New South Wales, 1979.
- Messier, W. F. & Emery, D. R., Some Cautionary Notes on the Use of Conjoint Measurement for Human Judgment Modeling, *Decision Sciences* (October 1980), pp. 678–690.
- Mock, T. J. & Turner, J. L., The Effect of Changes in Internal Controls on Audit Programs, in Burns, T. J. (ed.), *Behavioral Experiments in Accounting II* (Columbus, Ohio: College of Administrative Science, Ohio State University, 1979).
- Moriarity, S., Communicating Financial Information Through Multi-Dimensional Graphics, *Journal of Accounting Research* (Spring 1979), pp. 205–223.
- Moriarity, S. & Barron, F. H., Modeling the Materiality Judgments of Audit Partners, *Journal of Accounting Research* (Autumn 1976), pp. 320–341.
- Moriarity, S. & Barron, F. H., A Judgment-Based Definition of Materiality, *Studies on Auditing – Selections from the Research Opportunities in Auditing Program*, Supplement to *Journal of Accounting Research* (1979), pp. 114–135.
- Newell, A. & Simon, H. A., *Human Problem Solving* (Englewood Cliffs, N.J.: Prentice-Hall, 1972).
- Newton, L. K., The Risk Factor in Materiality Decisions, *The Accounting Review* (January 1977), pp. 97–108.
- Nisbett, R. E. & Wilson, T. D., Telling More Than We Can Know: Verbal Reports on Mental Processes, *Psychological Review* (1977), pp. 231–259.
- Pankoff, L. D. & Virgil, R. L., Some Preliminary Findings from a Laboratory Experiment on the Usefulness of Financial Accounting Information to Security Analysts, *Empirical Research in Accounting: Selected Studies, 1970*, Supplement to *Journal of Accounting Research* (1970), pp. 1–48.
- Payne, J. W., Task Complexity and Contingent Processing in Decision Making: An Information Search and Protocol Analysis, *Organizational Behavior and Human Performance* (December 1976), pp. 366–387.
- Payne, J. W., Braunstein, M. L. & Carroll, J. S., Exploring Pre-decisional Behavior: An Alternative Approach to Decision Research, *Organizational Behavior and Human Performance* (February 1978), pp. 17–44.
- Postman, L. & Tolman, E. C., Brunswik's Probabilistic Functionalism, in Kock, S. (ed.), *Psychology: A Study of a Science, 1* (New York: McGraw-Hill, 1959).
- Prakash, P. & Rappaport, A., Information Inductance and Its Significance for Accounting, *Accounting, Organizations and Society* (1977), pp. 29–38.
- Pratt, J. & Waller, W., A Reinforcement Contingency Model of Decision Behavior: An Application in an Accounting Setting. *Unpublished Manuscript*. University of Washington, 1979.
- Rockness, H. O. & Nikolai, L. A., An Assessment of A.P.B. Voting Patterns, *Journal of Accounting Research* (Spring 1977), pp. 154–167.
- Ronen, J., Some Effects of Sequential Aggregation in Accounting on Decision Making, *Journal of Accounting Research* (Autumn 1971), pp. 307–332.
- Russo, J. E., Comments on Behavioral and Economic Approaches to Studying Market Behavior, in Mitchell, A. A. (ed.), *The Effect of Information on Consumer and Market Behavior* (Chicago: American Marketing Association, 1978).
- Schroder, H., Driver, M. & Streufert, S., *Human Information Processing* (New York: Holt, Rinehart & Winston, 1967).
- Schultz, J. J. & Gustavson, S. G., Actuaries' Perceptions of Variables Affecting the Independent Auditor's Legal Liability, *The Accounting Review* (July 1978), pp. 626–641.
- Shields, M. D., Some Effects of Information Load on Search Patterns Used to Analyze Performance Reports, *Accounting, Organizations and Society* (1980a), pp. 429–442.
- Shields, M. D., On the Use of a Performance Report. *Unpublished Manuscript*. University of North Carolina, 1980b.
- Slovic, P., Psychological Study of Human Judgment: Implications for Investment Decision Making, *Journal of Finance* (September 1972), pp. 779–799.
- Slovic, P., Fischhoff, B. & Lichtenstein, S., Behavioral Decision Theory, *Annual Review of Psychology* (1977), pp. 1–39.

- Slovic, P., Fleissner, D. & Bauman, W. S., Analyzing the Use of Information in Investment Decision Making: A Methodological Proposal, *Journal of Business* (April 1972), pp. 283–301.
- Snowball, D., Some Effects of Accounting Expertise and Information Load: An Empirical Study, *Accounting, Organizations and Society* (forthcoming).
- Snowball, D. & Brown, C., Decision Making Involving Sequential Events: Some Effects of Disaggregated Data and Disposition Toward Risk, *Decision Sciences* (October 1979), pp. 527–546.
- Solomon, I., Tomassini, L. & Romney, M., Empirical Evaluation of Auditors' Prior Probability Distributions. Presented at the Annual Meeting of the American Accounting Association, Boston, 1980.
- Stephens, R. G., Accounting Disclosures for User Decision Processes, in Ijiri, Y. and Whinston, A. B. (eds.) *Quantitative Planning and Control* (New York: Academic Press, 1979), pp. 291–309.
- Swieringa, R. J., Discussion of a Judgment-Based Definition of Materiality, *Studies on Auditing – Selections from the Research Opportunities in Auditing Program*, Supplement to *Journal of Accounting Research* (1979), pp. 139–147.
- Swieringa, R. J., Dyckman, T. R. & Hoskin, R. E., Empirical Evidence about the Effects of an Accounting Change on Information Processing, in Burns, T. J. (ed.), *Behavior Experiments in Accounting II* (Columbus, Ohio: College of Administrative Science, Ohio State University, 1979).
- Torgerson, W. S., *Theory and Methods of Scaling* (New York: John Wiley & Sons, 1958).
- Tversky, A. & Kahneman, D., Judgment under Uncertainty: Heuristics and Biases, *Science* (1974), pp. 1124–1131.
- Tversky, A. & Kahneman, D., Causal Schemes in Judgments under Uncertainty, in Fishbein, M. (ed.), *Progress in Social Psychology, Vol. 1*, 1980.
- Tversky, A. & Sattath, S., Preference Trees, *Psychological Review* (1979), pp. 542–573.
- Uecker, W. C., A Behavioral Study of Information System Choice, *Journal of Accounting Research* (Spring 1978), pp. 169–189.
- Uecker, W. C., The Effects of Knowledge of the User's Decision Model in Simplified Information Evaluation, *Journal of Accounting Research* (Spring 1980), pp. 191–213.
- Uecker, W. C. & Kinney, W., Judgmental Evaluation of Sample Results: A Study of the Type and Severity of Errors Made by Practicing CPA's, *Accounting, Organizations and Society* (1977), pp. 169–189.
- Vasarhelyi, M., Man-Machine Planning Systems: A Cognitive Style Examination of Interactive Decision Making, *Journal of Accounting Research* (Spring 1977), pp. 138–153.
- Ward, B. H., An Investigation of the Materiality Construct in Auditing, *Journal of Accounting Research* (Spring 1976), pp. 138–152.
- Weber, R., Auditor Decision Making on Overall System Reliability: Accuracy, Consensus and the Usefulness of a Simulation Decision Aid, *Journal of Accounting Research* (Autumn 1978), pp. 368–388.
- Weber, R., Some Characteristics of the Free Recall of Computer Controls by EDP Auditors, *Journal of Accounting Research* (Spring 1980), pp. 214–241.
- Winkler, R. L., The Assessment of Prior Distributions in Bayesian Analysis, *Journal of the American Statistical Association* (September 1967), pp. 776–800.
- Wright, W. F., Accuracy of Subjective Probabilities for a Financial Variable, in Burns, T. J. (ed.), *Behavioral Experiments in Accounting II* (Columbus, Ohio: College of Administrative Science, Ohio State University, 1979).
- Zimmer, I., A Lens Study of the Prediction of Corporate Failure by Bank Loan Officers, *Journal of Accounting Research* (Autumn 1980), pp. 629–636.