THE ACCOUNTING REVIEW
Vol. 71, No. 2
April 1996
pp. 221–240

Using Decision Aids to Improve Auditors' Conditional Probability Judgments

Sarah E. Bonner
University of Southern California
Robert Libby
Mark W. Nelson
Cornell University

ABSTRACT: Prior research indicates that auditors encounter difficulty in applying experienced error frequencies to judgments of the probability that an audit objective is violated given a particular transaction cycle. This difficulty may occur because of a mismatch between the organization of this particular judgment task and the organization of auditors' knowledge. We test the effectiveness of two decision aids at counteracting this difficulty: (1) a checklist aid which facilitates knowledge retrieval and (2) a decomposition-and-mechanical-aggregation aid which facilitates both knowledge retrieval and aggregation. The checklist aid slightly improved the degree to which auditors' judgments reflected experienced frequencies and the mechanical-aggregation aid greatly improved auditors' judgments, completely counteracting the effect of the task organization-knowledge organization mismatch.

Key Words: *Decision aids, Knowledge organization, Audit judgment, Probability judgment.*

Data Availability: *Contact the authors.*

I. INTRODUCTION

A basic motivation for research examining the role of memory in audit judgment has been to inform the design of decision aids aimed at overcoming specific memory-related judgment deficiencies (e.g., Libby 1985, 649; Weber 1980, 216). Developing these aids

We thank KPMG Peat Marwick for supporting this research with a Research Opportunities in Auditing Grant and for providing additional support to Mark Nelson through a KPMG Peat Marwick Faculty Fellowship. We also appreciate comments received from Jean Bedard, Marlys Lipe, Bill Waller, participants at the 1995 annual meeting of the American Accounting Association, and workshop participants at the University of Arizona, Boston Accounting Research Colloquium, University of Colorado, University of Memphis, University of North Carolina at Chapel Hill and the University of Southern California.

Submitted December 1994.
Accepted October 1995.

requires research which identifies the specific source(s) of judgment error and tests the effectiveness of decision aids designed to target those source(s) of judgment error (Abdolmohammadi 1987; Ashton and Willingham 1988; Messier 1995). Mismatches between the organization of an audit task and the organization of the auditor's knowledge are one source of judgment error receiving increased attention in recent research (e.g., Frederick 1991; Ricchiute 1992). This study examines the effectiveness of two types of decision aids designed to alleviate the effects of one type of task-knowledge mismatch.

Knowledge can be organized along several dimensions. Studies in psychology indicate that when one dimension of a knowledge organization is more important than another dimension, people can apply the frequencies they have experienced in judgment tasks which are conditioned on the *more* important dimension, but have difficulty applying the frequencies they have experienced in judgment tasks which are conditioned on the *less* important dimension (Gavanski and Hui 1992; Sherman et al. 1992). Research in accounting indicates that such a mismatch in dimension importance may occur when auditors judge conditional probabilities in audit planning. Specifically, as auditors gain experience, they appear to develop a multidimensional knowledge organization in which audit objective becomes an increasingly important organizing dimension for knowledge of financial statement errors; however, detailed audit planning tasks tend to be organized primarily by transaction cycle. This mismatch between knowledge organization and task organization results in senior auditors' judgments of probabilities conditioned on transaction cycle, not reflecting the error frequencies they experience (Nelson et al. 1995).

We use the abstract audit planning setting developed in Nelson et al. (1995) to test the effectiveness of two types of decision aids that target hypothesized effects of this mismatch in dimension importance between knowledge organization and task organization. One aid is a list that reminds auditors of the membership of different financial statement errors in transaction cycle and audit objective categories. This aid (hereafter called the "list aid") provides auditors with retrieval cues with which they can recall and combine the frequency knowledge necessary to estimate conditional probabilities. The second aid decomposes the estimation of conditional probabilities into assessments of individual error probabilities, then aggregates those component judgments mechanically to assess the conditional probability. This aid (hereafter called the "mechanical-aggregation aid") assists in both the retrieval and aggregation of component judgments. The list aid provides an accurate structure for retrieving the necessary knowledge components and leaves the aggregation of that knowledge to the auditor, while the mechanical-aggregation aid directly assists retrieval and performs the aggregation process mechanically for the auditor.

To examine the effectiveness of these aids, we conducted an experiment in which auditors acquired frequency knowledge through experience, judged conditional probabilities and judged the probability of individual errors. The results of the experiment were as hypothesized. Consistent with the results of Nelson et al. (1995), auditors had difficulty applying experienced frequencies to conditional probability judgments of the form P(objective error | cycle) when no decision aid was provided. Both the list aid and the mechanical-aggregation aid improved the degree to which these conditional probability judgments reflected experienced frequencies. However, the mechanical-aggregation aid was much more effective than the list aid, indicating that, as hypothesized, difficulties in both knowledge retrieval and aggregation contribute to poor application of experienced frequencies to these judgments. The benefits of the mechanical-aggregation aid were dramatic enough to completely reverse the effects of the mismatch in dimension importance between knowledge organization and task organization. Should these results be confirmed in more realistic implementation tests, they would suggest that mechanical-aggregation aids may be more effective remedies for this type of mismatch. They would also

suggest the usefulness of the general approach of developing decision aids designed to specifically target hypothesized sources of judgment deficiencies.

The rest of this paper proceeds as follows. Section II proposes hypotheses based on prior literature in psychology and accounting and on an analysis of the task organization and knowledge organization related to conditional probability judgments. Section III describes the method. Section IV provides information about analyses and results, and the final section discusses the results.

II. RELATED LITERATURE AND HYPOTHESIS DEVELOPMENT

Fischhoff (1982) suggests that the source of decision error should determine the appropriate strategy for improving decision making. Similarly, Ashton and Willingham (1988) and Messier (1995) describe the ultimate goal of audit judgment research as improving the effectiveness and/or efficiency of audits by providing a scientific basis for improving audit decisions. All of these authors note that judgment research can be useful in identifying the sources of judgment error when there is some sort of mismatch between the decision maker and the decision maker's task.

Such mismatches can take many forms. For example, the format in which information is presented to decision makers (i.e., in graphs vs. tables) may not match the manner in which information is represented in memory, leading to diminished performance in interpreting and using the information in decision making (Stone and Schkade 1991; Vessey 1991). Similarly, the usefulness of a rule-based expert system for training may be diminished by a mismatch between the organization of the rule system and users' representation of the task domain (Pei and Reneau 1990). More directly related to auditing, the primarily transaction-cycle-based organization of audit planning tasks may not match the complex knowledge organizations and judgment processes of experienced auditors, which may have other important dimensions such as the temporal sequence of transactions within internal control systems (Frederick 1991) and the causal explanations relevant to important judgments such as determining a client's ability to continue as a going concern (Ricchiute 1992).

We examine the effects of a mismatch between: (1) the relative importance of transaction cycles and audit objectives in the organization of audit planning tasks and (2) the relative importance of these dimensions in auditors' knowledge organizations. In the remainder of this section, we review literature on the organization of audit planning tasks and the organization of auditors' knowledge, the effects of differences between the two on judgment, and the effects of decision aids designed to target those judgment issues.

Organization of Audit Planning Tasks

The audit planning process typically requires auditors to allocate audit effort (i.e., the type and extent of audit tests) to satisfy various audit objectives for each transaction cycle. The probability that an audit objective is violated depends on the transaction cycle in question (e.g., validity errors occur more often in the sales and receivables cycle, while completeness errors occur more often in the acquisitions and payments cycle). Audit workpapers are organized primarily by transaction cycle, auditors assess the quality of accounting systems and internal controls at the transaction cycle level (Arens and Loebbecke 1994; KPMG Peat Marwick 1993), and time budget modifications often occur at the cycle level, so it is natural to condition probability judgments on transaction cycle, i.e., to judge P(objective error | cycle).

Knowledge of the frequency with which errors have occurred under similar circumstances would be useful when making such conditional probability judgments. However, Nelson et al. (1995) demonstrate that auditors' conditional probability judgments and related allocations of

audit effort in abstract audit planning tasks are positively influenced by experienced frequencies only when the judgments are of the form P(cycle error | objective). When judgments are of the form P(objective error | cycle), experienced frequencies do not have a positive effect, even though judgments of the form P(objective error | cycle) are more important to the audit planning process. Judgments which do not reflect underlying error frequencies could lead to lower efficiency through over-emphasis on the detection of low frequency errors and lower effectiveness through under-emphasis on the detection of high frequency errors (Nelson et al. 1995).

Content and Organization of Auditors' Knowledge

Given that auditors appear to have difficulty applying experienced frequencies to probability judgments that are conditioned on cycle, it is appropriate to seek a remedy. To identify possible remedies, it is important to consider two possible sources of the judgment problem. First, some specific elements of knowledge required to estimate probabilities might be lacking or incorrect. Second, knowledge may be sound, but a mismatch between knowledge organization and task organization may impede the processing of that knowledge.

Knowledge Required for Conditional Probability Estimation

It appears that the average, experienced auditor possesses the knowledge needed to estimate conditional probabilities. First, auditors must understand the manner in which errors in the recording of transactions result in errors in financial statements. This basic knowledge of accounting theory and systems is developed over an auditor's college education and early years of experience (Bonner and Lewis 1990; Libby and Tan 1994; Nelson 1993), such that experienced auditors have a highly developed understanding of financial statement errors. Second, auditors must know both the cycle and objective category memberships of errors, since conditional probability judgments involve these categories. This knowledge also develops over early years of auditors' experience (C. Davis 1994; Frederick et al. 1994; Tubbs 1992), such that experienced auditors are relatively accurate in sorting financial statement errors by objective or cycle when instructed to do so (Frederick et al. 1994; Nelson et al. 1995). Third, auditors must know the frequency of occurrence of individual errors. Previous research indicates that auditors can learn fairly accurate relative error frequencies from experience (Butt 1988; Nelson 1993) and that experienced auditors display fairly accurate knowledge of relative frequencies of individual errors (Ashton 1991; Libby and Frederick 1990; Kaplan et al. 1992; Libby 1985). Fourth, auditors need to know the frequency of occurrence of categories of cycle errors, because the probability judgments are conditioned on cycle categories. Given that auditors possess knowledge of cycle category memberships very early in their careers, we would expect them to be able to estimate category-level relative frequencies, and evidence suggests that they can (Nelson et al. 1995).

Thus, it appears that auditors possess, to a large degree, each of the individual knowledge elements necessary to utilize experienced frequencies when making probability judgments conditioned on cycle. Also, when asked to estimate probabilities conditioned on *objective* (i.e., probabilities of the form P(cycle error | objective)), auditors *do apply* their experienced frequencies in the estimation process (Nelson et al. 1995). The knowledge necessary for conditional probability estimation apparently is present in auditors' memories, but not applied to judgments of the form P(objective error | cycle). Therefore, we focus our examination on potential *processing problems* which might be created by knowledge organizations inconsistent with judgments of P(objective error | cycle).

Knowledge Organization

Prior research in psychology indicates that the organization of knowledge in memory changes over time in response to domain-specific experience (see, e.g., Chase and Simon 1973;

Chi and Glaser 1982; Chi et al. 1982; Larkin et al. 1980; Reimann and Chi 1989), with a causal dimension emerging as an important organizing dimension in a number of task domains (Lien and Cheng 1990). Research also suggests that a more important dimension of a knowledge organization forms a "natural sample space" which is accessed spontaneously in judgments of conditional probabilities (Gavanski and Hui 1992; Sherman et al. 1992). A natural sample space facilitates the application of experienced frequencies to the estimation of probabilities which are conditioned on the more important dimension of the knowledge organization, but hinders the application of experienced frequencies to the estimation of probabilities which are conditioned on less important dimensions of the knowledge organization.

Prior research in accounting shows that audit objective eventually becomes a more important dimension than transaction cycle for organizing experienced auditors' knowledge of financial statement errors (C. Davis 1994; Frederick et al. 1994; Nelson et al. 1995).[1] This organization may serve a variety of important functions, such as highlighting the causes of error (Tubbs 1992) and identifying general substantive testing approaches which can be used to satisfy audit objectives in all transaction cycles (Arens and Loebbecke 1994). However, as mentioned previously, this knowledge organization facilitates some conditional probability judgments and hinders others (Nelson et al. 1995). For example, imagine that an auditor is asked to estimate a probability conditioned on objective, e.g., "Given that you are concerned with violations of the validity objective, what is the probability that the violation will occur in the sales and receivables cycle (as opposed to some other cycle)?" An auditor with this knowledge organization should find it relatively easy to access the validity objective category and compare the relative numbers of sales and receivable cycle errors which violate the validity objective to the number of errors from other cycles which violate the validity objective. However, if asked to estimate a probability conditioned on cycle (e.g., "Given that you are concerned with the sales and receivables cycle, what is the probability an error will violate the validity objective (as opposed to some other objective)?"), an auditor with this knowledge organization must access and compare the number of sales and receivable errors occurring in all the objective categories, which requires that several objective categories and cycle subcategories be identified, accessed and combined, rather than only several subcategories within one objective category. This mismatch between auditors' knowledge organizations and the task organization required by judgments of P(objective error | cycle) results in an "unnatural sample space," which diminishes the degree to which these conditional probability judgments reflect experienced frequencies (Nelson et al. 1995).

Specific Processing Problems Caused By Mismatch Between Task Organization and Knowledge Organization

The unnatural sample spaces resulting from the mismatch between audit task organization and auditors' knowledge organizations could create at least two processing problems. First, auditors may have difficulty identifying and retrieving all the individual errors in the various categories and subcategories and/or the frequencies of those errors needed to compute a probability that is conditioned on an unnatural sample space. When judging a probability conditioned on a natural sample space, all the relevant individual errors are members of a single category, so identification and retrieval processes are simplified. Second, auditors may be able to identify and recall individual error frequencies, but have difficulty aggregating them across

[1] There may be a number of other important dimensions of auditors' knowledge organizations besides audit objective and transaction cycle. We restrict our focus to objective and cycle in this paper because prior psychology research indicates that these are the dimensions whose importance *relative to each other* in auditors' knowledge structures should influence the application of experienced frequencies to judgments of P(objective error | cycle).

categories and making all the calculations necessary for estimating probabilities conditioned on an unnatural sample space. In response to these difficulties, auditors might reverse the conditional and estimate instead a conditional based on a related natural sample space (e.g., estimating P(cycle error | objective) instead of P(objective error | cycle)). Prior research in psychology supports this explanation for the influence of unnatural sample spaces on conditional probability estimation (Sherman et al. 1992).[2]

Therefore, the keys to improving the degree to which auditors' judgments of P(objective error | cycle) reflect experienced frequencies appear to be facilitating the retrieval of individual errors and their frequencies and facilitating the aggregation of those frequencies to form conditional probabilities. The next section considers the effectiveness of decision aids in achieving these objectives.

Decision Aids Designed to Target Processing Problems

List Decision Aid

One approach to decision aiding decomposes a judgment into components which the decision maker can assess and subsequently combine (Raiffa 1968). Prior psychology research indicates that even simple checklists improve decisions by helping decision makers to consider systematically all information relevant to their decision problem (MacGregor et al. 1988; Armstrong et al. 1975).

Previous accounting research has provided mixed results regarding the effects of checklist aids on auditors' performance. For example, E. Davis (1994) found that auditors using a checklist did not make better judgments of the probability of a going-concern problem than auditors not using an aid. Similarly, Pincus (1989) and Eining et al. (1993) found that the performance of auditors judging the probability of management fraud did not improve in the presence of a checklist of red flags. On the other hand, Heiman (1990) and Kennedy (1995) found that simple memory prompts reduced memory-related biases, and Butler (1985) and Heiman-Hoffman (1992) found improved judgment due to such prompts.

We expect a list aid to increase the degree to which conditional probability judgments reflect experienced frequencies. The list aid (shown in table 2) provided to our subjects specifies each individual error's respective objective and cycle category membership. The aid corrects any possible miscategorizations and provides retrieval cues to help auditors recall individual errors, and, thus, may help them judge individual error frequencies. To the extent that auditors estimate relative frequencies rather than absolute frequencies, the listing of all errors experienced may facilitate individual error frequency estimation. By providing the cycle category membership of each error, the list should insure that auditors can recall which individual error frequencies are needed to obtain category-level frequencies, as well as providing the category names as retrieval cues for any frequency information that might be stored at the category level. This aid is similar to the standard checklist decision aids used in practice for many audit tasks (e.g., lists of internal controls used in control evaluation, lists of red flags used for management fraud judgments).

Our first hypothesis is:

H1: Auditors' judgments of P(objective error | cycle) which are aided with a list of errors and their category memberships better reflect experimental error frequencies than do auditors' unaided judgments.

---

[2] This is not merely the result of semantic confusion over the meaning of a conditional probability, because subjects asked to estimate conditionals that are consistent with their knowledge organization do not make similar errors.

Mechanical-Aggregation Decision Aid

Another approach to decision aiding is to decompose a decision into its component judgments and mechanically aggregate those components for the decision maker (Edwards 1966; Einhorn 1972; Lyness and Cornelius 1982; for a summary, see Dawes et al. 1993). Mechanical aggregation of components can be more accurate than a single overall judgment when: (1) the component judgments are estimated with high accuracy, and (2) significant error is introduced through non-mechanical aggregation. However, if component judgments are systematically biased, that bias can accumulate in the mechanical aggregation process to produce a highly inaccurate judgment, while a single global judgment would be less biased and more subject to reasonableness checks (Kleinmuntz 1990; Ravinder et al. 1988).

As with checklist aids, the results of previous auditing research concerning the effectiveness of mechanical-aggregation aids are mixed. Jiambalvo and Waller (1984) and Daniel (1988) produced results which indicate that mechanically aggregated component judgments decrease performance in audit risk assessment, with mechanically aggregated probabilities often exceeding 1.0. Kachelmeier and Messier (1990) found a similar result with respect to sample size determination, with sample sizes larger but much more variable than intuitively estimated sample sizes. On the other hand, Libby and Libby (1989) found that mechanically aggregating component judgments improved judgments of internal control reliance, and E. Davis (1994) provided similar results for going-concern judgments. Eining et al. (1993) also found that auditors' judgment performance using decomposition-and-mechanical-aggregation aids was better than that for unaided global judgments.

We expect mechanical aggregation to increase the degree to which conditional probability judgments reflect experienced frequencies. Specifically, our subjects assess the probabilities of individual errors, then these probability judgments are combined mechanically to form conditional probability estimates. Previous research has found that auditors learn individual error frequencies relatively well from experience (e.g., Libby 1985). The mechanical-aggregation aid assists at least three processes: (1) retrieving the appropriate individual errors which serve as component judgments (because subjects are forced to provide answers for the correct components), (2) aggregating the individual error frequencies necessary to form category-level frequencies, and (3) combining individual and category-level probabilities to make the conditional probability judgment. An important feature of this mechanical aggregation process is that there is no possibility of auditors reversing the conditional (i.e., estimating P(cycle error | objective) when P(objective error | cycle) was requested), because the components combined to estimate the conditional probability are identified and combined mechanically. This aid is also similar to those used frequently in practice. For example, some aids mechanically combine judgments of inherent risk, control risk and analytical procedures risk to determine tests of details risk and the consequent required sample size (for a summary of these applications, see Bamber et al. 1995; Libby 1981).

We propose the following hypothesis:

H2: Judgments of P(objective error | cycle) obtained by mechanically aggregating auditors' unaided judgments of individual error probabilities better reflect experimental error frequencies than do auditors' unaided global conditional probability judgments.

List Aid Versus Mechanical-Aggregation Aid

Finally, we examine the degree to which each decision aid facilitates judgments of P(objective error | cycle) in the presence of the other aid. Recall that the decision aids were chosen

to address two possible process explanations for difficulties in judging P(objective error | cycle): the list aid was chosen to reduce difficulties in retrieval, while the aggregation aid was chosen to reduce difficulties in both retrieval and aggregation. The list aid should not significantly improve aggregation-aided judgments of P(objective error | cycle), since the process facilitated by the list aid (retrieval) is already facilitated by the aggregation aid. On the other hand, the aggregation aid should significantly improve list-aided judgments, since it assists retrieval *and* mechanically performs the aggregation process (a process not directly facilitated by the list aid). The following hypotheses result:

H3: Mechanically aggregated judgments of P(objective error | cycle) that are aided with a list of errors and their category memberships do *not* better reflect experimental error frequencies than do mechanically aggregated judgments not aided by the list.

H4: Judgments of P(objective error | cycle) obtained by mechanically aggregating auditors' list-aided judgments of individual error frequencies better reflect experimental error frequencies than do auditors' list-aided global conditional probability judgments.

Note that we also examine the effects of the aids on judgments of the form P(cycle error | objective), although we do not propose a hypothesis about these effects. Nelson et al. (1995) provide evidence that, because judgments of P(cycle error | objective) are facilitated by the natural sample spaces provided by auditors' knowledge organizations, they reflect experienced frequencies more accurately than do judgments of P(objective error | cycle). For these judgments, the decision aid could have no effect, or could even have a negative effect by encouraging auditors to disregard a useful knowledge organization. By examining effects on these judgments, we examine a potential cognitive cost of the aids.

Summary of Hypotheses

Figure 1 provides a summary of the four hypotheses stated above. H1 compares global unaided judgments to global judgments that are aided by the list alone (cell 1 versus cell 3). H2 compares global unaided judgments to mechanically aggregated judgments that are unaided by

FIGURE 1
Summary of Hypotheses

	No Mechanical-Aggregation Aid (Human Judgments)	Mechanical-Aggregation Aid (Mechanically Aggregated Judgments)
No List Aid	Cell 1	Cell 2
List Aid	Cell 3	Cell 4

H1: Cell 3 > Cell 1
H2: Cell 2 > Cell 1
H3: Cell 2 = Cell 4
H4: Cell 4 > Cell 3

the list (cell 1 versus cell 2). H3 and H4 examine the incremental effects of each aid when the other is present. H3 compares mechanically aggregated judgments made in the presence of the list aid to mechanically aggregated judgments made without the list aid (cell 2 versus cell 4) to determine whether the list aid helps once the mechanical-aggregation aid has been applied. H4 compares global list-aided judgments to mechanically aggregated list-aided judgments (cell 3 versus cell 4) to determine whether mechanical aggregation improves judgments after the list aid has been applied.

III. METHOD

Subjects

Subjects were 105 auditors from a single Big 6 firm who were required to participate in the experiment as part of firm training sessions. The auditors had an average of 3.3 years of experience, which is enough to have developed a knowledge organization for financial statement errors in which objective is a more important organizing dimension than is cycle (Frederick et al. 1994; Nelson et al. 1995). Thus, these auditors were expected to form natural sample spaces for judging conditional probabilities of financial statement errors on the basis of audit objectives, hindering their abilities to judge probabilities conditioned on transaction cycle. Further, auditors at this level of experience are accustomed to using computerized decision aids, so the subjects provided with the list aid should have found it familiar.

Overview, Design and Procedures

Overview

Subjects completed the experiment on Macintosh computers in rooms configured for computerized instruction. The computerized procedure standardized the timing of frequency presentations and facilitated randomization of items and treatments (to be discussed later) and data recording. Subjects were not allowed to speak with each other throughout the experiment. One of the researchers was available at all times to supervise and answer subjects' questions.

During the experiment, subjects proceeded through several tasks. This sequence of tasks is shown in table 1. First, subjects completed a free sort task. Next, subjects observed individual presentations of financial statement errors designed to convey frequencies. Following a distracter task, subjects judged conditional probabilities and completed another distracter task. Then, subjects estimated individual error probabilities and completed another distracter task. Finally, subjects completed an unconditional probability judgment task and two directed sort tasks. Subjects took an average of 32.2 minutes to complete the experiment.

Design

The experiment employed a $2 \times 2 \times 2 \times 2$ design. To test whether the frequencies presented early in the experiment were reflected in subjects' later conditional probability judgments, each subject estimated two conditional probabilities, one which the presented frequencies indicated should be high and one which the presented frequencies indicated should be low, such that the two conditional probability questions become the first within-subjects variable. The treatment effect (i.e., the difference between the answers to these two questions) measures the degree to which experienced frequencies affect conditional probability estimates. To determine the effects of the list and mechanical-aggregation decision aids (both individually and in tandem) on the degree to which subjects' conditional probability judgments reflected experienced frequencies, provision of the list aid was manipulated between subjects, and provision of the mechanical-aggregation aid

TABLE 1
Sequence of Experimental Procedures

1. Free sort nine individual errors
2. Presentation of nine individual errors in varying frequencies (total of 49)
3. Distracter task (4 math problems)
4.. List aid introduced to list-aided subjects (available to them until step 10)
5. Conditional probability questions (2 questions conditioned on either cycle or objective)
6. Distracter task (4 demographic questions)
7. Individual error probability questions (used for mechanical-aggregation aid; 9 questions)
8. Distracter task (2 demographic questions)
9. Unconditional probability questions (6 questions—3 each for cycles and objectives)
10. Directed sort by cycles
11. Directed sort by objectives

NOTE: Order of 10 and 11 determined randomly.

was manipulated within subjects (the 2 × 2 formed by crossing the list aid and mechanical-aggregation aid variables is shown in figure 1). To examine the effects of the decision aids, both when knowledge organization matched task organization and when it did not, the type of conditional probability estimated by the subjects (either P(cycle error | objective) or P(objective error | cycle)) was also manipulated between subjects. Subjects were randomly assigned to one of the four combinations of the two between-subjects variables.

Materials and Procedure

As indicated in table 1, subjects first performed a free sort task to confirm that objective was a more important organizing dimension than cycle in their knowledge organizations. This task required subjects to sort nine financial statement errors into three categories on the basis of "how they thought the errors best go together." Subjects accomplished this sorting by clicking on either "1," "2," or "3" beside each error to indicate how they would group the errors. The order of presentation of the nine errors in the free sort task was randomized and held constant across subjects. The free sort task was the same as that used by Nelson et al. (1995).

In the second part of the experiment, subjects viewed individual presentations of nine financial statement errors in varying frequencies, for a total of 49 presentations. The nine financial statement errors and frequencies were the same as those used by Bonner et al. (1995) and Nelson et al. (1995); they are shown in tables 2 and 3. The errors were selected to fully cross three transaction cycles with three audit objectives.[3] To provide subjects with reasonably well-defined categories, the errors were chosen to be the best members of their respective categories. The frequency of presentation of each of the nine individual errors was determined by randomly assigning the three transaction cycles and audit objectives to the rows and columns shown in table 3 for each subject. Subjects were told that they would see a series of errors found during the audits of medium-sized manufacturers by a major accounting firm during the last quarter of 1993, and that their task was to remember the errors to the best of their ability. No instructions regarding frequency learning were given; subjects were told only that each error would remain on the

[3] See Nelson et al. (1995) for a more detailed explanation of the error selection process.

TABLE 2
List Aid

Error Matrix

	Trade Receivables, Sales and Collections Cycle	*Inventory, Purchasing and Cost of Goods Sold Cycle*	*Investments Cycle*
Proper Cutoff Objective	Next period's sales were included in the current year's revenue and receivables.	Raw materials were improperly shown as received after year-end.	Purchases of treasury bills were recorded in the wrong fiscal period.
Validity Objective	Billings to legitimate customers were booked twice.	More finished goods were recorded as received than were actually received.	Fictitious investments were included in the account balance.
Valuation Objective	The bad debt expense and allowance were underestimated.	Obsolete inventory was not written down to net realizable value.	Marketable securities were not reduced to lower of cost or market.

computer screen for ten seconds. This method of presenting error frequencies is the same as that used by Bonner et al. (1995) and Nelson et al. (1995), and is similar to that used by Butt (1988).[4]

After viewing the individual error presentations, subjects performed a distracter task consisting of two addition and two division problems. Next, subjects were asked to imagine that they worked for the firm from which the list of errors previously shown was obtained, and to assume that they had been exposed to all of those audits, such that the errors constituted their experience. Following this, subjects assigned to the list aid treatment received instructions regarding its presence and purpose and were required to examine it before proceeding. The program was set up such that subjects could not proceed without accessing the list aid. After this, subjects were able to access it at any time by clicking on an "error matrix" button on each screen. The aid which appeared when subjects clicked on the "error matrix" button is shown in table 2. To control for the possibility that people visualize categories better as columns versus rows, half of the list subjects saw cycles as rows and half saw cycles as columns. The list aid was available for access by subjects assigned to the list treatment for all probability judgments (conditional,

[4] We expect that the effect of the individual error presentations should not be influenced by auditors' pre-existing frequency knowledge. Frequency knowledge is time-tagged in memory such that experimental frequencies can be discriminated from pre-existing frequency knowledge (Butt 1988). Also, our random assignment of cycles and objectives to experimental frequencies should ensure that any influence of pre-existing frequency knowledge is spread across treatments. Finally, as discussed further in section IV, our results indicate that the aggregation aid produces judgments of P(objective | cycle) which *do* reflect experimental frequencies. Any interference from pre-existing frequency knowledge would influence the judgments of individual error probabilities on which the aggregation aid is based, so the correspondence between aggregation-aided judgments and experimental frequencies indicates that results are not driven by interference from pre-existing frequency knowledge.

TABLE 3

Individual and Category-Level Frequencies of Errors and Frequency-Based Answers to Conditional Probability Questions

Audit Objective	*Transaction Cycle* A	B	C	*Total — Objective Category*	
I	1	2	4	7	P(hi | low)=4/7
II	2	4	8	14	
III	4	8	16	28	P(low | hi)=4/28
Total — Cycle Category	7	14	28	49	
	P(hi | low)=4/7		P(low | hi)=4/28		

individual error and unconditional), and subjects voluntarily accessed the aid an average of 6.6 times (mean 7.6 times including the mandatory initial access; median 8 times; range 1–17 times) for an average total access time of 1.97 minutes (median 1.47 minutes; range 0.28–6.37 minutes).

Next, all subjects made two conditional probability judgments. The type of conditional probability judgments made depended on the random assignment of the subject to either P(cycle error | objective) or P(objective error | cycle).[5] Subjects assigned to P(cycle error | objective) estimated conditionals of the form P(high-frequency cycle | low-frequency objective) and P(low-frequency cycle | high-frequency objective), while subjects assigned to P(objective error | cycle) estimated conditionals of the form P(high-frequency objective | low-frequency cycle) and P(low-frequency objective | high-frequency cycle). As noted in table 3, these conditionals differ the most in the level of the conditional probability implied by the frequencies presented to subjects (4/28 for P(low | high), 4/7 for P(high | low)). As such, they allow detection of an effect without requiring a large number of judgments by the subjects. The specific cycles and objectives (e.g., sales, validity) referenced in the conditionals depended on the random assignment of the errors in table 2 to the error frequencies in table 3. Subjects answered these questions by sliding a pointer on a scale numbered from 1 to 100; the instructions explained that these numbers indicated probabilities. The order of the two questions (P(high | low) and P(low | high)) was randomized between subjects. After estimating conditional probabilities, subjects completed a distracter task which consisted of four demographic questions about age, whether English was a first or second language, years of experience in a CPA firm and years of experience in the audit department.

Next, each subject estimated the probability of occurrence of each of the nine financial statement errors (again, with or without the assistance of the list aid).[6] In the instructions to this task, subjects were told that the total number of instances of errors presented was 49; this was done

[5] For example, a subject assigned to P(objective error | cycle) might be asked the following: "Given that you are dealing with an error that occurred in the TRADE RECEIVABLES, SALES AND COLLECTIONS CYCLE, what is the probability that the error violated the VALUATION OBJECTIVE?"

[6] For example, each subject was asked: "What it the probability that the following error occurred? 'NEXT PERIOD'S SALES WERE INCLUDED IN THE CURRENT YEAR'S REVENUE AND RECEIVABLES.'"

to hold constant subjects' knowledge of the population frequency. As before, subjects made their probability judgments by moving a pointer on a sliding scale with endpoints of 1 and 100. The order of the nine judgments was randomized between subjects. These nine judgments constituted the components that were mechanically aggregated to form aided, conditional probability judgments. Again, after this task, subjects completed a short distracter task consisting of two demographic questions (title of position, business unit).

After this distracter task, subjects estimated six unconditional probabilities (one for each of three cycles and objectives, with or without the list), and again were told that the total number of instances in the population was 49. These questions were similar to the unconditional frequency questions used in Nelson et al. (1995), except that they were stated in terms of probabilities, requiring subjects to answer on a scale of 1 to 100. Finally, subjects performed two directed sort tasks. These sort tasks were the same as the free sort task, except that they directed subjects to sort according to either transaction cycle or audit objective. Whether a cycle sort or objective sort appeared first was randomized between subjects. The directed sort tasks and unconditional probability questions provided information that was used in ruling out alternative explanations in our replication of the results of Nelson et al. (1995).

IV. RESULTS

Verification of Objective as More Important Than Cycle in Organizing Auditors' Knowledge of Financial Statement Errors

Based on the results of C. Davis (1994), Frederick et al. (1994), Nelson et al. (1995) and Tubbs (1992), we expected that our subjects would, on average, possess a knowledge organization in which objective was a more important organizing dimension for financial statement errors than cycle. We verified this with the same technique employed by Nelson et al. (1995). Subjects' free sorts of errors were compared to an "ideal" free sort that subjects would have made had they sorted perfectly according to either audit objective or transaction cycle. Each subject's free sort was coded in a 9×9 similarity matrix with a "1" if errors were grouped together and a "0" if they were not. Similarity matrices of this type were also formed for the "ideal" cycle and objective sorts shown in table 2. Each subject's similarity matrix was compared to these "ideal" matrices by correlating the vectors formed from the lower left triangle of each matrix. This measure of the similarity of matrices is called a "cophenetic correlation" (Sneath and Sokal 1973).

The average cophenetic correlation between subjects' similarity matrices from their free sorts and the predetermined objective (cycle) similarity matrix was 0.58 (–0.13), which is similar to Nelson et al.'s (1995) results of 0.53 (–0.04). The average correlation with the "ideal" objective matrix is significantly higher than that with the "ideal" cycle matrix ($t = 10.01$, $p = .0001$) and indicates that, on average, the audit objective dimension is more important than the cycle dimension for organizing financial statement errors for our group of auditor subjects.

Verification of Effect of Knowledge Organization on Conditional Probability Judgments

Next we verified that *unaided* conditional probability judgments of the form P(cycle error | objective) better reflected experimental frequencies than did those of the form P(objective error | cycle). This constituted a replication of the main result found by Nelson et al. (1995) and required the same analysis performed in that study. The analysis was a planned contrast within a 2×2 ANOVA on the *global* conditional probability judgments made in the *without-list* group, with level of conditional probability (either P(high | low) or P(low | high)) a within-subjects variable and type of category conditioned upon (either cycle or objective) a between-subjects variable. A

replication of previous results would produce a difference between P(high frequency *cycle* error | low frequency *objective*) and P(low frequency *cycle* error | high frequency *objective*) that is significantly greater than the difference between P(high frequency *objective* error | low frequency *cycle*) and P(low frequency *objective* error | high frequency *cycle*).

The mean conditional probability judgments are shown in table 4. Consistent with the results of Nelson et al. (1995), the contrast was significant ($t = 3.20$; one-tailed $p = .001$). In fact, the results are even stronger than those of Nelson et al. (1995), because auditors' unaided global probability judgments were positively influenced by experimental frequencies when conditioned on objective ($t = 2.61$; one-tailed $p = .006$), and *negatively* associated with experimental frequencies when conditioned on cycle ($t = -1.94$; one-tailed $p = .029$). In other words, as in prior psychology studies (e.g., Sherman et al. 1992), experimental frequencies were applied inversely when knowledge organization did not match task organization. Nelson et al. (1995) found a significant positive correlation between experimental frequencies and judgments conditioned on objective, but did not find a significant negative correlation between experimental frequencies and judgments conditioned on cycle (i.e., the slope of that line was essentially zero).[7]

Tests of Hypotheses

H1 and H2

H1 predicted that auditors' conditional probability judgments which are aided with a list-type aid will better reflect experimental frequencies than will unaided judgments. H2 predicted that conditional probability judgments formed by mechanically aggregating auditors' individual error probability judgments will better reflect experimental frequencies than will auditors' unaided judgments. These predictions were made for probability judgments of the form P(objective error | cycle), because these judgments are inconsistent with a knowledge organization in which objective is a more important organizing dimension than cycle.

Each of these two hypotheses was tested by comparing the judgments aided with a single decision aid, either the list or mechanical-aggregation aid (cells 3 and 2, respectively, in figure 1), to unaided judgments (cell 1). These tests were structured as planned contrasts within an overall $2 \times 2 \times 2$ ANOVA with judgments of the form P(objective error | cycle) as the dependent variable and level of conditional probability (P(high | low) or P(low | high)), list aid (present or not) and mechanical-aggregation aid (global judgments versus mechanically aggregated judgments) as independent variables. The planned contrasts isolated the cell-by-cell comparisons necessary for testing the hypotheses.[8]

H1 would be supported if the difference between judgments of P(high | low) and P(low | high) is greater with the list than without the list. This contrast was significant ($t = 1.96$; one-tailed

---

[7] As in Nelson et al. (1995), we used directed sort data and judgments of unconditional probabilities of subjects assigned to the without-list conditions to rule out alternative explanations for results. Specifically, we used this data to insure that: (1) subjects' knowledge of the cycle and objective category memberships of errors did not differ, and (2) subjects' knowledge of unconditional probabilities of cycle errors was not significantly different from their knowledge of unconditional probabilities of objective errors. The results of these tests indicate that our results cannot be explained by differences in category knowledge or unconditional probability knowledge.

[8] All hypothesis tests are based on contrasts that examine how the aids affect the difference between the two conditional probability estimates (i.e., examining the degree to which the decision aids help subjects to discriminate P(high | low) and P(low | high)). An alternative approach is to conduct separate tests for P(low | high) and P(high | low) which examine $|P - P^*|$, where P is the judged conditional probability and P^* is the conditional probability calculated from the frequencies actually presented to subjects in the experiment (14.3 for P(low | high) and 57.1 for P(high | low)). This alternative approach provides the same results for our hypothesis tests as that provided by our analyses, so it will not be discussed further.

TABLE 4

Unaided Conditional Probability Judgments (n = 53)
Analysis of Conditional Type[1] × Presented Conditional Probability Contrast

Means (Standard Deviations)

Conditional Probability Indicated by Presented Frequencies	*Estimates of P(Objective Error / Cycle)*	*Estimates of P(Cycle Error /Objective)*
P(low | high) = 4/28 = 14.3%	44.6 (31.1)	39.0 (25.4)
P(high | low) = 4/7 = 57.1%	30.8 (21.3)	56.5 (27.7)

[1] "Conditional Type" is either P (cycle error | objective) or P (objective error | cycle).

p = .028). As discussed previously and shown in table 5, unaided judgments were significantly negatively associated with experienced frequencies (difference of –13.8 between P(high | low) and P(low | high)), while judgments aided with the list were in the right direction, although not significantly so (difference of 2.4; t = .41; one-tailed p = .340).

H2 would be supported if the difference between judgments of P(high | low) and P(low | high) is greater for mechanically aggregated judgments than for global judgments. This contrast was significant (t = 5.19; one-tailed p = .000). As shown in table 5, decomposition and mechanical aggregation of auditors' judgments changed the judgments from being significantly negatively associated with experienced frequencies to reflecting a significant positive difference (of 21.2) between the two levels of conditional probabilities (t = 4.98; one-tailed p = .000).

Thus, both H1 and H2 were supported. Each of the aids improved auditors' judgments of probabilities conditioned on cycle. However, the effect of the list aid was limited and the effect of the mechanical-aggregation aid was much larger.

TABLE 5

Conditional Probability Judgments of the Form P(Objective Error | Cycle) (n = 52)
Tests of Hypotheses 1–4: Effects of List Aid and Mechanical-Aggregation Aid

Means (Standard Deviations)

	Conditional Probability Indicated By Presented Frequencies	*No Mechanical-Aggregation Aid (Human Judgments)*	*Mechanical-Aggregation Aid (Mechanically Aggregated Judgments)*
No List Aid	P(low | high) 4/28 = 14.3%	44.6 (31.1)	23.2 (13.1)
	P(high | low) 4/7 = 57.1%	30.8 (21.3)	44.4 (17.8)
List Aid	P(low | high) 4/28 = 14.3%	34.8 (25.4)	24.0 (11.1)
	P(high | low) 4/7 = 57.1%	37.2 (24.4)	44.1 (17.2)

H3 and H4

Because the mechanical-aggregation aid directly assisted auditors with both retrieval and aggregation of frequencies, we expected that it would improve judgments which were already helped by the list aid (H4), while the list aid would not improve judgments that were already decomposed and mechanically aggregated (H3). To test H3, we compared the mechanically aggregated judgments of the without-list group (cell 2 in figure 1) to those of the list group (cell 4). As predicted, this contrast was not significant ($t = 0.20$; two-tailed $p = .850$). As can be seen in table 5, the difference between P(high | low) and P(low | high) was virtually identical for mechanically aggregated judgments with and without the list (20.1 versus 21.2).

To test H4, we compared the global judgments of subjects in the list group (cell 3) to their mechanically aggregated judgments (cell 4). As predicted, this contrast was significant ($t = 2.73$; one-tailed $p = .004$). When the list aid was present, the addition of the mechanical-aggregation aid changed the difference between P(high | low) and P(low | high) from a very small positive difference (2.4) to a large positive difference (20.1). When viewed in combination with the results for H1 and H2, these results indicate that both retrieval and aggregation are processing difficulties underlying auditors' inaccurate application of the frequencies they have experienced to judgments of P(objective error | cycle), consistent with the hypothesized effects of the task organization-knowledge organization mismatch.

Effects of Aids on Judgments of P(Cycle Error | Objective)

Finally, because of the practical concern that the use of a decision aid could impair judgments in tasks whose organization matches knowledge organization, we also examined the effects of the list and mechanical-aggregation aids on judgments of the form P(cycle error | objective). To examine these effects, we performed the same analyses that were used to test the four hypotheses, but examined judgments of the form P(cycle error | objective) rather than judgments of the form P(objective error | cycle). Table 6 presents the means and standard deviations for judgments of P(high | low) and P(low | high) for each of the four between-subjects conditions.

To examine whether the list aid affected judgments (as in H1), we examined the effect of the list aid when no mechanical-aggregation aid was present. As can be seen from table 6, the

TABLE 6

Conditional Probability Judgments of the Form P(Cycle Error | Objective) (n = 53)
Additional Analyses: Effects of List Aid and Mechanical-Aggregation Aid

Means (Standard Deviations)

	Conditional Probability Indicated By Presented Frequencies	*No Mechanical-Aggregation Aid (Human Judgments)*	*Mechanical-Aggregation Aid (Mechanically Aggregated Judgments)*
No List Aid	P(low | high) 4/28 = 14.3%	39.0 (25.4)	18.1 (10.3)
	P(high | low) 4/7 = 57.1%	56.5 (27.7)	46.2 (23.5)
List Aid	P(low | high) 4/28 = 14.3%	27.8 (21.0)	23.3 (15.2)
	P(high | low) 4/7 = 57.1%	38.4 (26.2)	44.1 (22.1)

difference between P(high | low) and P(low | high) was slightly smaller in the list group (10.6 as compared to 17.5), but not significantly so ($t = 0.76$; two-tailed $p = .450$). Next, as in our test of H2, we examined the effect of the mechanical-aggregation aid when no list aid was present. Although decomposition and mechanical aggregation increased the difference between P(high | low) and P(low | high) from 17.5 to 28.1, the increase was not significant ($t = 1.29$; two-tailed $p = .203$). Finally, we examined the incremental effects of each aid given the existence the other. First, we examined the effect of the list aid with the mechanical-aggregation aid already applied, as in our test of H3. The list aid slightly decreased the difference between P(high | low) and P(low | high) (28.1 versus 20.8), but this difference was not significant ($t = 0.88$; two-tailed $p = .384$). Second, we examined the effect of the mechanical-aggregation aid with the list aid already applied, as in our test of H4. The mechanical-aggregation aid slightly increased the difference between P(high | low) and P(low | high) (from 10.6 to 20.8), but this difference was not significant ($t = 1.18$; two-tailed $p = .244$). These results indicate that neither the list aid nor the mechanical-aggregation aid significantly influenced the degree to which conditional probability judgments of the form P(cycle error | objective) reflect experienced frequencies.

V. DISCUSSION

Authors in both psychology (Fischhoff 1982) and accounting (Ashton and Willingham 1988; Messier 1995) have advocated an approach to decision aid development which first identifies the psychological causes of judgment errors in a particular situation and then tests the effectiveness of decision aids which are specifically designed to target those causes. Our study followed that approach by testing the effectiveness of decision aids designed to target the difficulties that auditors have in applying their experienced frequencies to judgments of conditional probabilities when there exists a mismatch in dimension importance between the task organization and knowledge organization. Drawing on the results of Nelson et al. (1995), we examined two decision aids (list and mechanical-aggregation) that targeted retrieval and aggregation of frequencies in auditors' probability judgments of the form P(objective error | cycle).

Results were as predicted. The free-sort data indicated that the experienced auditor subjects possessed a knowledge organization for financial statement errors in which audit objective was a more important organizing dimension then transaction cycle. Also, a comparison of unaided judgments of the form P(cycle error | objective) to those of the form P(objective error | cycle) shows that the former reflect experimental frequencies better than do the latter (the conditional probabilities which are directly relevant to audit planning). This result replicates the results of Nelson et al. (1995). Second, both list-aided and mechanical-aggregation-aided judgments of the form P(objective error | cycle) better reflected experienced frequencies, suggesting that both decision aids could enhance auditor judgments; however, the aggregation aid provided much more improvement. Third, once the mechanical-aggregation aid was applied, the list aid provided no further improvement in conditional probability judgments, while the mechanical-aggregation aid improved judgments that were already aided by the list. Because the list aid targeted primarily the retrieval process, while the mechanical-aggregation aid targeted both retrieval and aggregation processes, these results suggest that difficulties in both retrieval and aggregation drive the observed difficulties in conditional probability judgments. In turn, this supports the theorized mismatch between knowledge organization and task organization which caused retrieval and aggregation difficulties. Thus, our development of decision aids which target specific hypothesized cause(s) of a judgment difficulty not only enabled us to test whether the difficulty can be alleviated, but also to test (indirectly) our underlying theory for why the difficulty occurs.

From an audit practice perspective, our results indicate that the mechanical-aggregation aid alone would eliminate the adverse effects of a mismatch between dimension importance in knowledge organization and task organization, with judgments and decisions conditioned on transaction cycle reflecting auditors' experienced frequencies to the same extent as do judgments and decision conditioned on objective. The list aid would have less effect, eliminating the inverse application of experienced frequencies to conditional probability estimates, but not resulting in judgments and decisions which reflect experienced frequencies to a very large extent. Whether either of these decision aids is preferable to existing practice depends on their relative costs, as well as on the potential for incorporating frequency information into the audit planning process in other ways. Given that neither decision aid had a significant effect on judgments of the form P(cycle error | objective), neither decision aid creates this possible cognitive cost. However, since mechanical-aggregation aids require many component judgments and specific algorithms for mechanical aggregation, they may be more costly in terms of time and effort and less flexible in application than are list aids such as the one employed in our experiment. Also, less experienced auditors may not be considered to have been exposed to a large enough sample of audits to possess sufficiently accurate knowledge of individual error frequencies for decision aids based on their experience to be useful, and it may be costly to use group assessments and/or training to enhance the accuracy of individual error frequency estimates (Abdolmohammadi 1994; Nelson 1993). As an alternative, archival frequencies could be considered explicitly in the planning process (or implicitly through the consideration of client characteristics which influence the frequency of error), such that a list aid which merely eliminates the inverse application of experienced frequencies may be sufficient. The relative cost effectiveness of these decision aids and alternatives for their implementation await future research.

Another important direction for future research is to examine the generality of our results in more realistic audit settings. Our study obtains experimental control by conveying error frequencies absent other important characteristics of the audit setting. While we believe this is the correct approach for the first work in this area, future research could gradually relax this control to examine how these characteristics influence our results. For example, future research could examine the extent to which our results hold when auditors' estimates of conditional probabilities are influenced by client-specific inherent and control risk factors. Similarly, future research could examine the extent to which standardized audit programs mitigate the effects of inaccurate conditional probability estimates on audit planning decisions.

REFERENCES

Abdolmohammadi, M. J. 1987. Decision support and expert systems in auditing: A review and research directions. *Accounting and Business Research* 17 (Spring): 173–185.

———. 1994. Summary information decision aids as a means of providing error frequency knowledge in tests of controls. Working paper, Bentley College, Waltham, MA.

Arens, A. A., and J. K. Loebbecke. 1994. *Auditing: An Integrated Approach.* 6th ed. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Armstrong, J. S., W. B. Denniston, Jr., and M. M. Gordon. 1975. The use of the decomposition principle in making judgments. *Organizational Behavior and Human Performance* 14 (October): 257–263.

Ashton, A. H. 1991. Experience and error frequency knowledge as potential determinants of audit expertise. *The Accounting Review* 66 (April): 218–239.

Ashton, R. H., and J. J. Willingham. 1988. Using and evaluating audit decision aids. In *Auditing Symposium IX*, edited by R. P. Srivastava and J. E. Rebele. Lawrence, KS: University of Kansas.

Bamber, E. M., P. R. Gillett, T. J. Mock, and K. T. Trotman. 1995. Audit judgment. In *Auditing Practice, Research and Education: A Productive Collaboration*, edited by T. Bell and A. Wright. New York, NY: American Institute of Certified Public Accountants.

Bonner, S. E., and B .L. Lewis. 1990. Determinants of auditor expertise. *Journal of Accounting Research* 28 (Supplement): 1–20.

———, R. Libby, and M. W. Nelson. 1995. Audit category knowledge as a precondition to learning from experience. Working paper, Cornell University, Ithaca, NY.

Butler, S. A. 1985. Application of a decision aid in the judgmental evaluation of substantive test of details samples. *Journal of Accounting Research* 23 (Autumn): 513–526.

Butt, J. 1988. Frequency judgments in an audit-related task. *Journal of Accounting Research* 26 (Autumn): 315–330.

Chase, W. G., and H. A. Simon. 1973. Perception in chess. *Cognitive Psychology* 4 (January): 55–81.

Chi, M. T. H., and R. Glaser. 1982. The measurement of expertise: Analysis of the development of knowledge and skill as a basis for assessing achievement. In *Design, Analysis and Policy in Testing and Evaluation*, edited by E. L. Baker and E. S. Quelmalz. Beverly Hills, CA: Sage Publications.

———, ———, and E. Rees. 1982. Expertise in problem solving. In *Advances in the Psychology of Human Intelligence 1,* edited by R. Sternberg. Hillsdale, NJ: Lawrence Erlbaum Associates.

Daniel, S. J. 1988. Some empirical evidence about the assessment of audit risk in practice. *Auditing: A Journal of Practice & Theory* 7 (Spring): 174–181.

Davis, C. E. 1994. Experience and the organization of auditors' knowledge of financial statement errors. Working paper, Baylor University, Waco, TX.

Davis, E. B. 1994. Effects of decision aid type on auditors' going concern evaluations. Working paper, Baylor University, Waco, TX.

Dawes, R. M., D. Faust, and P. E. Meehl. 1993. Statistical prediction vs. clinical prediction: Improving what works. In *A Handbook for Data Analysis in the Behavioral Sciences: Methodological Issues*, edited by G. Keren and C. Lewis. Hillsdale, NJ: Lawrence Erlbaum Associates.

Edwards, W. 1966. Nonconservative probabilistic information processing systems. Institute of Science and Technology, University of Michigan, Ann Arbor, MI, Report Number ESD-TR-66-404.

Einhorn, H. J. 1972. Expert measurement and mechanical combination. *Organizational Behavior and Human Performance* 7 (February): 86–106.

Eining, M. M., D. R. Jones, and J. K. Loebbecke. 1993. An experimental examination of the impact of decision aids on the assessment and evaluation of management fraud. Working paper, University of Utah, Salt Lake City, UT.

Fischhoff, B. 1982. Debiasing. In *Judgment Under Uncertainty: Heuristics and Biases*, edited by D. Kahneman, P. Slovic, and A. Tversky. New York, NY: Cambridge University Press.

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

———, V. Heiman-Hoffman, and R. Libby. 1994. The structure of auditors' knowledge of financial statement errors. *Auditing: A Journal of Practice & Theory* 13 (Spring): 1–21.

Gavanski, I., and C. Hui. 1992. Natural sample spaces and uncertain belief. *Journal of Personality and Social Psychology* 63 (November): 766–780.

Heiman, V. B. 1990. Auditors' assessments of the likelihood of analytical review explanations. *The Accounting Review* 65 (October): 875–890.

Heiman-Hoffman, V. B. 1992. Experience Effects in Analytical Review: The Relation Between Knowledge and Judgment. Working paper, University of Pittsburgh, Pittsburgh, PA.

Jiambalvo, J., and W. Waller. 1984. Decomposition and assessment of audit risk in practice. *Auditing: A Journal of Practice & Theory* 3 (Spring): 80–88.

KPMG Peat Marwick. 1993. *Audit Program Software*.

Kachelmeier, S. J., and W. F. Messier, Jr. 1990. An investigation of the influence of a nonstatistical decision aid on auditor sample size decisions. *The Accounting Review* 65 (January): 209–226.

Kaplan, S., C. Moeckel, and J. D. Williams. 1992. Auditors' hypothesis plausibility assessments in an analytical review setting. *Auditing: A Journal of Practice & Theory* 11 (Fall): 50–65.

Kennedy, J. 1995. Debiasing the curse of knowledge in audit judgment. *The Accounting Review* 70 (April): 249–274.

Kleinmuntz, D. 1990. Decomposition and the control of error in decision-analytic models. In *Insights in Decision Making: A Tribute to Hillel J. Einhorn,* edited by R. M. Hogarth. Chicago, IL: University of Chicago Press.

Larkin, J. H., J. McDermott, D. P. Simon, and H. A. Simon. 1980. A model of competence in solving physics problems. *Science* 208 (June 20): 1335–1342.

Libby, R. 1981. *Accounting and Human Information Processing: Theory and Applications.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

———. 1985. Availability and the generation of hypotheses in analytical review. *Journal of Accounting Research* 23 (Autumn): 648–667.

———, and D. M. Frederick. 1990. Experience and the ability to explain audit findings. *Journal of Accounting Research 28* (Autumn): 348–367.

———, and P. Libby. 1989. Expert measurement and mechanical combination in control reliance decisions. *The Accounting Review* 64 (October): 729–747.

———, and H.-T. Tan. 1994. Modeling the determinants of auditor expertise. *Accounting, Organizations, and Society* 19 (November): 701–716.

Lien, Y., and P. Cheng. 1990. *A Bootstrapping Solution in the Problem of Differentiating Spurious from Genuine Causes.* Paper presented at the 31st Annual Meeting of the Psychonomic Society.

Lyness, K. S., and E. T. Cornelius. 1982. A comparison of holistic and decomposed judgment strategies in a performance rating simulation. *Organizational Behavior and Human Performance* 29 (February): 21–38.

MacGregor, D., S. Lichtenstein, and P. Slovic. 1988. Structuring knowledge retrieval: An analysis of decomposed quantitative judgments. *Organizational Behavior and Human Decision Processes* 42 (December): 303–323.

Messier, Jr., W. F. 1995. Research in and development of audit decision aids. In *Judgment and Decision Making in Accounting and Auditing,* edited by R. H. Ashton and A. H. Ashton. Cambridge: Cambridge University Press.

Nelson, M. W. 1993. The effects of error frequency and accounting knowledge on error diagnosis in analytical review. *The Accounting Review* 68 (October): 803–824.

———, R. Libby, and S. E. Bonner. 1995. Knowledge structure and the estimation of conditional probabilities in audit planning. *The Accounting Review* 70 (January): 27–47.

Pei, B. K. W., and J. H. Reneau. 1990. The effects of memory structure on using rule-based expert systems for training: A framework and an empirical test. *Decision Sciences* 21 (Spring): 363–386.

Pincus, K. V. 1989. The efficacy of a red flags questionnaire for assessing the possibility of fraud. *Accounting, Organizations, and Society* 14 (Nos. 1/2): 153–163.

Raiffa, H. 1968. *Decision Analysis.* Reading, MA: Addison-Wesley.

Ravinder, H. V., D. N. Kleinmuntz, and J. S. Dyer. 1988. The reliability of subjective probabilities obtained through decomposition. *Management Science* 34 (February): 186–199.

Reimann, P., and M. T. H. Chi. 1989. Human expertise. In *Human and Machine Problem Solving,* edited by K. J. Gilhooly. New York, NY: Plenum Press.

Ricchiute, D. N. 1992. Working-paper order effects and auditors' going-concern decisions. *The Accounting Review* 67 (January): 46–58.

Sherman, S. J., M. N. McMullen, and I. Gavanski. 1992. Natural sample spaces and the inversion of conditional judgments. *Journal of Experimental Social Psychology* 28 (September): 401–421.

Sneath, P. H., and R. R. Sokal. 1973. *Numerical Taxonomy.* San Francisco, CA: W. H. Freeman and Co.

Stone, D. N., and D. A. Schkade. 1991. Numeric and linguistic information representation in multiattribute choice. *Organizational Behavior and Human Decision Processes* 49 (June): 42–59.

Tubbs, R. M. 1992. The effect of experience on the auditor's organization and amount of knowledge. *The Accounting Review* 67 (October): 783–801.

Vessey, I. 1991. Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision Sciences* 22 (Spring): 219–241.

Weber, R. 1980. Some characteristics of the free recall of computer controls by EDP auditors. *Journal of Accounting Research* 18 (Spring): 214–241.