DETERMINANTS OF JUDGMENT PERFORMANCE IN ACCOUNTING SETTINGS: ABILITY, KNOWLEDGE, MOTIVATION, AND ENVIRONMENT

ROBERT LIBBY
Cornell University

and

JOAN LUFT
Michigan State University

Abstract

The goal of this paper is to trace the development of our conceptions of the roles of ability, knowledge, motivation, and environment as determinants of decision performance in accounting settings, and provide a synthesis of the basic constructs, conceptual relations, and methodological guidelines that can be inferred from this diverse literature. We first outline the key characteristics of accounting settings and research principles appropriate for examining the characteristics. The primary emphasis is on identifying interactions between determinants of performance, specifying underlying cognitive processes, and abstraction based on theory and task analysis. We then demonstrate how selected studies which follow these basic principles have greatly enhanced our understanding of accounting-related decisions. Finally, suggestions for future research are presented.

Over the past twenty years, a significant literature has developed aimed at understanding individual and small group decision making in accounting settings. This literature is usually referred to as Human Information Processing, Judgment and Decision Making, or Behavioral Decision Research in Accounting. While this literature is quite diverse in the settings examined, the theories relied upon, and the methodologies applied, its central themes have remained unchanged: assessing decision performance, establishing the factors which determine decision performance, and testing theories of the cognitive processes which produce the decisions. Addressing these themes helps accountants understand why particular decision-making inputs and organizational arrangements are applied to particular accounting settings. For example, multiple combinations of inputs and organizational arrangements are available which will produce one GAAS audit, yet substantial consistencies exist across firms in the manner in which audits are organized. These consistencies in organizational design result from the interaction of the demands of the audit decision-making environment with human decision-making strengths and weaknesses (cf. Simon, 1947; Libby & Trotman, 1992). These themes also provide the information necessary for suggesting remedies for any discovered deficiencies, since we must understand a decision process to improve it (cf. Libby, 1976; Hogarth, 1991a). In a sense, we are establishing the factors that enter into accounting decision

* An earlier version of this paper was presented at the Conference on Decision Making, Cognitive Science and Accounting, Pittsburgh, PA, U.S.A., July 1991. We are grateful to Jacob Birnberg, Rob Bloomfield, Sarah Bonner, Vicky Heiman-Hoffman, Marlys Lipe, Don Moser, and Mark Nelson for helpful comments on earlier versions.
production functions and the forms of those functions.

Behavioral decision research in general is highly interdisciplinary, both in the underlying disciplines from which it developed (economics, statistics, and cognitive and social psychology) and the fields of decision making to which it has been applied (medicine, clinical psychology, law, public policy, etc.). Behavioral decision research in accounting both draws from and contributes to this interdisciplinary field.1 At the same time, its contributions to the practice of accounting are well recognized (see Elliott, 1983; Ashton & Willingham, 1988). These contributions to basic knowledge and to practice result to a great extent from the continuing emphasis on examining the impact of important attributes of accounting settings on decision performance. As is appropriate in a progressing field, our conceptions of these settings have become increasingly rich and representative.

The first decade of Behavioral Decision Research in Accounting led to sentiments similar to those expressed in Toda’s (1962) famous quote about the psychology literature:

Man and rat are both incredibly stupid in an experimental room. On the other hand, psychology has paid little attention to the things they do in their normal habitat; man drives a car, plays complicated games, and organizes society, and rat is troublesome cunning in the kitchen (Toda, 1962, p. 165).

In 1983, it was suggested that if the quote was modified for accounting research, it might read:

Accountants and students are incredibly stupid in an experimental room. On the other hand, accounting research has paid little attention to the things they do in their natural habitat; accountants audit the financial statements of large multinationals, interpret a multitude of complex tax regulations, and develop pension accounting standards, and students are troublesome cunning in the classroom. (My apologies to Toda.) (Libby, 1983, p. 79).

This quote leads to an important, yet until recently ignored question: what do accountants do well? or more precisely: what allows experienced accountants or financial statement users to perform their tasks efficiently and effectively?

The following conceptual equation (1) suggests the determinants of decision performance (Einhorn & Hogarth, 1980; Libby, 1983):

\[ \text{Performance} = f(\text{Ability, Knowledge, Environment, Motivation}). \]  

In 1977 and again in 1982, Libby & Lewis (1977, 1982) provided broadly based reviews of the research to date. Categorizing the early work in light of the above equation suggests that, following the basic psychology literature, the great majority of early studies focused on the impact of decision makers’ basic cognitive abilities and deficiencies on performance. This literature demonstrated that information processing limitations such as limited short-term memory, selective attention, lack of cognitive control, and others (which one might say determine the costs of thinking) could result in systematic and persistent decision errors in some important accounting situations. Clearly, cognitive abilities are a key input into decision performance, and people do not always possess or apply them in a degree sufficient to guarantee error-free judgment.

Libby (1983) suggested that future accounting research should turn its attention to the impact of the latter three categories of determinants and their interactions with human cognitive abilities and limits, and presented detailed examples of research in progress and potential studies addressing these issues. Since that time, a major portion of accounting research effort has shifted to these issues. This more recent literature has dramatically broadened our conception of decision making in accounting settings and has focused attention on the attributes of the accounting settings that distinguish them from those normally examined by psychologists (Gibbins & Jamal, 1991; Libby,

---

1 See Payne et al.'s (1992) recent Annual Review of Psychology chapter which references nine papers by accountants.
The goal of this paper is to trace the development of our conception of the role of knowledge, motivation, and environment as determinants of decision performance in accounting settings, and provide a synthesis of the basic constructs, conceptual relations, and methodological principles that can be inferred from this diverse literature.

The performance equation serves mainly as a heuristic for classifying the main concerns of the literature. Before it can serve as a basis for more formal modeling, we will need to know more about the complex interrelations among the four determinants of performance. Each can affect performance not only directly but also indirectly by affecting other determinants. As this paper will show, substantial progress has been made in discovering such interactions; but at present much about the nature of these relationships remains unknown. While the focus of this paper, like the focus of much recent research, is on the latter three determinants of performance (knowledge, motivation, and environment), ability considerations will sometimes appear when ability interacts with the other determinants.

Unlike Libby & Lewis (1977, 1982), the current paper does not provide an exhaustive review of the literature; it is more selective both as to subject matter and particular studies relied upon. These restrictions reflect the specific purpose of the current paper, our own biases, and the growth in volume and spectrum of issues addressed in the literature.² Both as a reflection of the current emphasis in the literature, and to limit the scope of the paper, we rely most heavily on experimental research in audit contexts as the basis for discussion. The remainder of the paper is organized as follows. In the next section, we outline the research principles and characterization of accounting settings which underlie the discussion. We then illustrate the benefits of applying these principles to the study of each determinant of judgment performance. This is followed by suggestions for future research.

RESEARCH PRINCIPLES AND ATTRIBUTES OF ACCOUNTING SETTINGS

Understanding the determinants of performance in accounting settings requires specifying which elements of those settings are the defining characteristics which can serve as a focus of our research attention. This includes the characteristics of the wide variety of accounting-related decision making tasks and the decision makers that perform them. Recognizing the existence of multiple tasks is critical because they differ in the abilities, knowledge, and effort required for their successful completion. Recognizing the variety among decision makers is necessary because performance will be determined by the fit of the task demands with the decision maker's ability, knowledge, and willingness to expend effort. Salient characteristics of many accounting-related tasks which we will address include:

1. **Learning opportunities.** Education, training, and experience provide opportunities to accumulate a wealth of task-related knowledge.

2. **Guidance and aid.** The environment normally provides a variety of types of judgment guidance and technological aid which we refer to generally as "technology".

3. **Hierarchical group settings.** These tasks are normally performed in hierarchical group settings where the individuals have different status, abilities, knowledge, accountability relations, and incentives.

4. **Sequential, multiperiod tasks.** These tasks involve multiple, sequential steps, and are often repeated in future periods (there is prior involvement).

The first attribute and its implications are recognized in the discussion of knowledge and ability-related issues. While recognizing that

² Growth of the field has necessitated that more recent exhaustive reviews focus on a narrower range of issues than did Libby and Lewis. See e.g. Solomon (1987), Bédard (1989), Klersey & Mock (1989), and Bonner & Pennington (1991).
they also affect learning, we examine the second, third, and fourth attributes in the discussion of environment and motivation.

The determinants of performance in equation (1) can be described and related to the attributes of accounting settings as follows. Abilities — here defined as the capacity to complete information encoding, retrieval, and analysis tasks — tend not to be specific to accounting settings. In contrast, the knowledge investigated in the studies discussed below is highly task-specific. Both the content and the organization of knowledge can be changed by decision makers' learning opportunities; and both content and organization can independently affect judgment performance. Both the content and the organization of this knowledge vary across individuals. The environment of accounting (especially auditing) decisions includes such features as judgment guidance and technological aid; hierarchical group settings; accountability relationships; sequential, multiperiod tasks; and substantial monetary incentives for good performance. Decision makers' motivation, or willingness to exert effort, is determined jointly by features of the environment such as monetary incentives and accountability relations, and by characteristics of individuals such as their utility functions and abilities.³

Three common principles underlie the approach taken here to disentangle the multiple interrelations of these factors.

(1) The determinants of performance interact (there are dependencies). As a result, we must simultaneously consider more than one factor to understand the whole picture.

(2) The key to understanding these dependencies is understanding the mechanisms or underlying cognitive processes through which they affect performance. In turn, the processes employed to produce a judgment are determined to a great extent by the task demands. Since accountants perform a variety of decision-making tasks which vary in the degree to which they depend on different cognitive processes, the mechanisms by which performance is influenced are often task-specific. Adequate recognition of these facts requires a more detailed representation of the decision process and the decision task than those employed in earlier research.

(3) The experimentalist's comparative advantage lies in the ability to abstract fundamental components of real world settings and control other potentially influential variables. This allows one to disentangle interrelated elements of the accounting setting which affect behavior and test the effects of conditions which do not yet exist in that setting. Separating fundamental components from mundane realistic detail is critical to the experimental research process. This abstraction process sometimes even requires "research that might seem, at first blush, to be quite far from accounting" (Hogarth, 1991b).

These principles will serve as recurring themes in our discussion of the literature.

The role of knowledge has been more thoroughly investigated in recent years than the roles of ability, environment, and motivation. The next section will discuss the conceptual and methodological tools that have been developed to capture knowledge and ability effects, and the progress that has been made using these tools. This section also provides a foundation for the discussion of other attributes of accounting settings in the subsequent section. Progress in understanding these attributes has arisen from the use of conceptual and methodological tools similar to those developed for knowledge studies, and from a recognition of the importance of interactions among knowledge and the other determinants of performance.

KNOWLEDGE AND ABILITY

In practice, more experienced decision makers are often paid more and are assigned to what

³ Even for people who place a very high value on money, a large sum of money provides no motivation to run a three-minute mile if one lacks the ability to do so.
appear to be more demanding tasks: presumably, their performance is expected to be better than that of their less experienced counterparts. Basic research in cognitive psychology suggests that much of experienced decision makers' advantage lies in their larger knowledge store and, more importantly, the manner in which they organize their knowledge so that it can be effectively brought to bear on the problem.

A great deal of recent attention has been focused on memory and knowledge issues in accounting by the recognition that experience is an important determinant of task assignment in accounting firms, and that experienced accountants rely heavily on the wealth of job-related knowledge that they bring to their tasks (cf. Libby, 1981, 1983; Birnberg & Shields, 1984; Gibbins, 1984; Waller & Felix, 1984). Yet before 1985, few experience-based performance differences had been uncovered and only one study of auditors' memory had been published (Weber, 1980). A major roadblock to successfully examining these issues was a lack of understanding of the necessary conceptual and operational approach.

The "expertise paradigm"

Frederick & Libby (1986) suggested a series of common-sense guidelines that experimental studies of knowledge issues should follow (based on Fiedler, 1982). These guidelines, which will be referred to here as the "expertise paradigm", reflect the influence of our three research principles emphasizing interactions, cognitive processes, and abstraction.

Conceptual perspective: specifying the knowledge and cognitive processes. Given the variety of audit tasks and the differences in the amount and nature of the knowledge they require, the first guideline is that hypotheses should be developed in advance about the effects of specific knowledge elements or their organization on observable behavior. This requires specification of the knowledge necessary to complete a particular task; when, how, and how well it will be acquired; and the process(es) through which it will be brought to bear on the task. Developing hypotheses on the basis of such analysis not only increases the chance for successful demonstrations of important effects, but provides a systematic basis to determine the reasons for failure to demonstrate predicted effects.

This common-sense requirement recognizes that specific training programs and experiences provide a basis for acquiring specific knowledge which will influence performance of only certain tasks (Libby, 1989, p. 134). Not all experiences are alike, and not all knowledge is equally relevant to each task. As a consequence, one should not expect all more experienced auditors to show superior performance at all tasks. This guideline further recognizes that, even assuming homogeneous career paths among auditors, different knowledge is acquired at different career stages (Abdoimohammadi & A. Wright, 1987). When combined with the diminishing effect of exposure on learning, continuous improvement on any one task is unlikely. In fact, performance on some tasks (for example where exposure rates decline) should decrease with time. Finally, this guideline recognizes that the effect of knowledge on decision performance will be a function of the sensitivity to knowledge differences of the cognitive processes on which performance depends. For example, Einhorn (1976) and Libby (1985) suggest that hypothesis generation will be very sensitive to relevant knowledge differences. Bonner (1990) demonstrates the sensitivity of cue selection and cue weighting to knowledge differences and Bonner (1991) demonstrates the relative lack of sensitivity of cue measurement to such differences.4

In general, the diversity in both the tasks performed by auditors and the knowledge required by each task requires disaggregation of the audit opinion formulation process and

4 Bonner & Pennington (1991) analyze the processes involved in a wide variety of audit judgment tasks and their sensitivity to knowledge differences.
detailed analyses of the underlying component processes to form the basis for predictions (Frederick & Libby, 1986).

Operational perspective: knowledge–task interactions. Knowledge is an internal mental state which cannot be directly observed. As a consequence, the second guideline is that demonstrating a hypothesized knowledge difference and/or its effects on performance requires constructing an experimental task where the observable implications of using and not using knowledge (or using different knowledge) are different. The failure to uncover experience effects in many prior studies was the result of having employed tasks where more and less experienced auditors would be expected to have the same knowledge or where different knowledge would be expected to result in the same behavior (cf. Bonner, 1990). Such tasks will be referred to as novice tasks.

The third guideline indicates that the existence of a knowledge effect can best be established by manipulating stimuli and/or context factors and comparing individuals with different experiences. If the knowledge element of interest relates to the stimuli, the stimuli must be manipulated. Alternatively, if it relates to context, then a single stimulus must be evaluated in multiple contexts. (The idea here is that context activates a knowledge element which interacts with the stimulus being evaluated). Finally, if the hypothesized knowledge relates to the interaction of the two (the interpretation of the stimuli is contingent on the context), then both must be manipulated. The power of a design based on the third guideline lies in its ability to eliminate alternative explanations for knowledge or performance differences.

Use of the expertise paradigm. A comparison of early studies of experience effects on audit judgment and more recent studies of knowledge-related effects illustrates the importance of these guidelines. Although the role of training and experience had been recognized in the literature, existing studies in the early 1980s were little more than exploratory analyses which either correlated judgmental performance statistics with years of audit experience or compared students and experienced auditors on these statistics. Knowledge was always used as an ex post explanation of any discovered performance differences. Little attempt was made to determine the knowledge necessary to perform different tasks, the stage at which such knowledge was normally acquired, or the process by which it was brought to bear on decision problems (guideline 1 above). Not surprisingly, the results of the early empirical tests showed little consistency.

The typical study of that era used policy capturing methodology. The theory section was usually quite general, relying on the judgment and decision-making literature where experience effects were usually a secondary issue. This approach did not provide a sufficient basis for developing directional hypotheses about the effects of experience. The typical task required that subjects weight binary internal control cues and record their response on an internal control strength scale. In order to predict the effects of experience on this task, it would be necessary to specify what knowledge and abilities are necessary to complete such a task and when they are acquired. Performance of the internal control task requires knowledge of the direction of the effect of each cue on overall strength, knowledge of the relative importance of each cue to overall strength (separation of duties is more important), and the ability to apply a rule consistently without lapses of attention or memory (cognitive control). The first two elements are acquired in introductory auditing, and the timing of the development of the third is uncertain. As a result, there is little basis to expect any experience-related differences. In fact, most studies found only marginal differences in self-insight and consensus and a slightly greater weighting of the separation-of-duties cues (see W. Wright, 1988; Bédard, 1989).

Bonner (1990) illustrates the power of a design which employs the same policy-capturing methodology as previous studies but follows the three guidelines. Bonner suggested that, in early studies of experience effects in
internal control evaluation, both the experienced and inexperienced auditors may have had the requisite knowledge (the researchers had examined novice tasks). Following guideline 1, Bonner specified that knowledge of relevant cue weights would normally be acquired earlier for internal control risk assessments than for analytical risk assessments; and therefore, more experienced auditors' advantage in cue weighting would be greater in the analytical risk task. She also recognized the need to control for the possibility that any discovered experience-related difference may have resulted from the overall superiority of senior people at all such tasks (an ability difference) and not the effects of task-specific knowledge. Following guideline 3, she separated out these effects by assessing cue weighting by experienced and inexperienced auditors in novice (internal control) and expert (analytical risk) tasks with the same structures. If senior auditors' advantage came from superior ability rather than task-specific knowledge, they should have outperformed novices equally in both tasks. She found, however, that the difference in consensus in cue weighting between experienced and inexperienced auditors was much greater for analytical risk assessments than control risk assessments. She concluded that prior results seem to be due in part to choice of novice tasks (e.g. control risk assessments). In other experiments (Bonner, 1990, 1991) she examined experience effects on cue selection and cue measurement.

This is a particularly good example of the application of the expertise paradigm and the research principles described above. It demonstrates an important knowledge–task interaction. The representation of the judgment process is more detailed than in prior studies. Abstraction is employed to disentangle the effects of experience on the various component processes producing judgments.

Use of the expertise paradigm is not, however, restricted to policy-capturing studies of judgment. Our understanding of knowledge effects has also been enlarged by new methodologies: studies that examine knowledge more directly, through memory tests, and studies that link memory phenomena to judgment.

Frederick (1991) investigated differences in the organization of general internal control knowledge between experienced and inexperienced auditors. Following guideline 1, he hypothesized that training exposes both inexperienced and experienced auditors to a taxonomic structure for internal control knowledge (organized by internal control objectives), whereas experiencing actual systems allows only experienced auditors to acquire a schematic structure (organized by flow of transactions). He operationalized these hypotheses based on the finding that the temporal links, which exist only in the schematic structure, act as retrieval cues aiding recall from memory. These temporal links are the mechanism through which the knowledge difference should result in a difference in behavior. Since only experienced auditors have the schematic structure, only they should exhibit superior recall of schematically organized controls. As a consequence, the implications of using and not using the knowledge are different (guideline 2).

He tested these predictions by presenting a group of practising auditors having three to five years of experience and a group of students having no experience with a listing of the same 33 controls over purchases and disbursements listed either taxonomically (by objective) or schematically (by flow of transactions). Since his hypothesis related to the stimuli, he varied the stimuli (organization of the information) and the subjects' experiences (guideline 3) in a $2 \times 2$ between-subjects design. Participants were told to study the controls and then to recall and record them as they come to mind. He then counted the number of controls recalled in each condition. As predicted and illustrated in Fig. 1, the experienced auditors exhibited superior recall in the schematic versus the taxonomic condition, but the inexperienced exhibited no significant difference in recall across organizations of the stimuli.

As in Bonner's study, the power of the design lies in the high degree of assurance it provides that the result was due to the specified
knowledge difference (here, a difference in organization), and not other subject differences or knowledge content differences. The relative weakness of simpler designs is made evident by considering the two simpler alternatives. First, if only the schematically structured materials had been employed, and the experienced auditors outperformed the inexperienced, two important alternative explanations for the results would not have been eliminated: (1) experienced auditors having superior memory for all stimuli; and (2) experienced auditors having greater knowledge content but the same organization (since familiarity with stimuli eases recall). In the second simple alternative design, if only experienced auditors had been asked to learn the taxonomically and schematically organized materials, superior memory for the schematically organized material could have resulted from the inherent ease of remembering the schematic ordering independent of internal control knowledge (for example, because of its relation to a simpler mnemonic). The $2 \times 2$ design used by Frederick eliminates each of these alternative explanations because they would have produced parallel lines in Fig. 1.

Frederick (1991) is a good illustration of the power of the expertise paradigm because internal control evaluation is an area where prior policy capturing research did not detect reliable experience effects. The paper also well illustrates the three research principles by: (1) demonstrating an interaction of knowledge with environment (the organization of internal control data); (2) applying theory related to basic cognitive processes (the functional properties of schematic organization); and (3) employing an abstract task (list recall), while still capturing an important attribute of audit knowledge, and providing results that are relevant to practical issues such as the organization of internal control workpapers. The experimental task need not look like a typical accounting task for the results to have significant implications for practice.

Experience, knowledge, ability, and performance: dimensions of expertise

The studies described above look at "one piece of the puzzle" at a time: the relation of a general measure of experience either to performance or to knowledge. These relations are represented in Fig. 2 panels (a) and (b). The need to examine directly the relations between knowledge and judgment performance has also been recognized (see e.g. Libby, 1983; Bédard, 1989; Choo, 1989; Davis & Solomon, 1989; Libby, 1989). Bonner & Lewis (1990) made three major innovations in this area by jointly demonstrating the relations between components of audit knowledge and performance, how these relations will vary across four audit tasks (internal control evaluation, ratio analysis, earnings manipulation, and interest rate swap), and how ability affects performance in these tasks. The conceptual relations examined are represented in panel (c) of Fig. 2.5 Bonner and Lewis constructed knowledge and ability tests and built regression models of the relations of knowledge and ability to performance for each task. While the psychometric approach taken is quite different on the surface from the experimental approach taken in the two studies described above, Bonner & Lewis (1990) still follow the basic research principles presented

---

5 Bonner and Lewis also included experience as an indirect measure or proxy for knowledge.
above by recognizing task related dependencies in the relations, analyzing differences in the reasoning processes involved in the tasks, and abstracting and disentangling elements of real world settings. The study also employs the expertise paradigm.

We suggest here that a more complete model than that examined in Bonner & Lewis (1990) relates ability, experience, knowledge, and performance in the manner presented in Fig. 3. The model recognizes: (1) that there are only two classes of inputs in the model (abilities and experiences), (2) that these two inputs (along with effort⁶) cause the internal state of knowledge which is an intermediate output variable, and (3) along with the direct effects of abilities, knowledge affects performance (an output variable). As noted earlier, our environment is one where accountants complete a variety of judgment tasks which differ in the opportunities to learn relevant knowledge and in the nature of the abilities, knowledge, and effort required for successful task completion. Performance will be determined by the fit of the accountant's ability and knowledge to those required by the task. As a consequence, a separate version of Fig. 3 could be constructed for each specific audit task, and the strength of the relations among the four variables should differ across tasks. The emphasis in the model is on interactions with task, differing underlying cognitive processes, and the disentangling of related elements of the environment.

Bonner and Lewis' knowledge measures are primarily measures of knowledge content, and their judgment tasks require subjects to generate this knowledge content in appropriate contexts (e.g. to list errors that could occur, given certain internal control weaknesses; or to determine income effects, given a certain pattern of errors). If the appropriate knowledge is generated, the task is performed correctly; no further judgments, such as probability

---

⁶ Effort determines the degree to which available knowledge and abilities are brought to bear on the task and (along with abilities) determines the degree to which people acquire knowledge from experiences. Effort issues will be considered at greater length in the following section on environment and motivation.
judgments, are required. A model like that in Fig. 3 also has potential for clarifying determinants of performance for these latter types of judgments. But before the full potential is realized, we need to learn more about the relationship of knowledge in memory to judgment, which remains a problematic issue in cognitive psychology. For example, Moser (1992) suggests that the manner in which one elicits or measures the knowledge in memory can determine whether a memory–judgment relationship will be found. He demonstrates that a request to recall presented information, rather than to self-generate all of the relevant information a subject can think of may not provide an accurate measure of the knowledge people are actually bringing to bear on a judgment. Heiman-Hoffman (1992) uses the self-generation measure of knowledge in memory and path analysis to demonstrate that experience-related knowledge differences (link 1 in Fig. 3) cause performance differences (link 3 in Fig. 3) in hypothesis evaluation in analytical review. A few other accounting studies (Moser, 1989; Heiman, 1990) have examined the relation of generation from memory and probability-judgment performance, but more remains to be learned in this area.

Conclusion

The experimental literature examining relations among ability, experience, knowledge, and performance is the fastest growing segment of judgment and decision research in accounting. Studies have been conducted related to knowledge of financial statement errors (Butt, 1988; Libby & Frederick, 1990; A. Ashton, 1991; Tubbs, 1991; Nelson, 1992), internal controls (Weber, 1980; Libby et al., 1985; Brown & Solomon, 1990; Spires, 1991), substantive procedures (Brown & Solomon, 1991), going concern indicators (Choo, 1991; Choo & Trotman, 1991), and other areas of audit judgment. Most of the research has concentrated on documenting the nature of knowledge and learning processes related to common audit decision problems that auditors must solve in the conduct of nearly every audit. These studies necessarily employ more general measures of experience and knowledge. Fewer studies have examined specialized decision problems where those who have relevant training and experience are few in number, and where more specific measures of experience and knowledge must be employed. The research is also currently being extended beyond the audit domain to other areas of accounting-related judgment (e.g. Moser, 1989, 1992; Bonner et al., 1992).

Significant contributions have been made by research employing a wide variety of approaches including policy capturing, probabilistic judgment, learning and memory, problem solving, psychometrics, and archival data analysis. This diversity in method is an important strength of the literature. Broader approaches (e.g. Bonner & Lewis, 1990) and deeper, more narrowly focused studies which examine individual links or pairs of links in the Fig. 3 model (e.g. Frederick, 1991) have different strengths and weaknesses and are both making important contributions to the study of the nature of audit expertise. Similarly, studies which examine learning based on experience in the lab (e.g. Butt, 1988; Nelson, 1992) and those which assess the effects of experiences accumulated in practice (e.g. Libby & Frederick, 1990) also make important contributions.

It is critical to recognize that an understanding of all of the links in Fig. 3 is necessary to understanding the nature of expertise in auditing. This point is best illustrated by considering the fact that, even where experience–performance relations have been adequately demonstrated, we cannot understand their implications for decision improvement without understanding the nature and source of the knowledge differences which produce the performance differences. We can help novice decision makers perform like experts only when we understand how experts learn to perform the way they do and the best ways to impart that knowledge. Hogarth (1991a) makes a similar point when he suggests that one cannot determine how to improve performance without understanding the process.
Analyzing the existing literature in the light of the above discussion suggests two keys to successful study of the nature of audit expertise. The first is precision in definition of research purpose, conceptual variables of interest, and their relations. Employing the conceptual model in Fig. 3 will hopefully aid in developing such precision. Once the purpose and relations of interest have been adequately defined, a design must be employed which allows detection of the proposed effects while eliminating alternative explanations for results. The guidelines in the "expertise paradigm" provide this needed second key. Successful studies of knowledge-related issues must develop hypotheses in advance based on specification of the knowledge necessary to complete a particular task; when, how, and how well the knowledge will be acquired, and the process(es) through which it will be brought to bear on the task. They must then test these hypotheses using subjects who can reasonably be expected to have different knowledge and experimental tasks where the knowledge differences have observable implications.

As discussed above, a principal motivation for accountants' interest in knowledge-related issues is the prevalence of hierarchical group settings where individuals with different knowledge, as well as different status, abilities, accountability relations and incentives, perform different subtasks. Similarly, the audit technology available for performance of different audit tasks helps determine the abilities, knowledge and effort necessary for task completion. The direct effects of these and other environmental attributes, and additional interactions with knowledge are discussed in the next section.

ENVIRONMENT AND MOTIVATION

The typical study of judgment and decision making examines individual decision makers, relatively unconstrained in their decision strategy, making a single decision or judgment. Many judgment tasks in accounting settings, in contrast, are performed by hierarchically structured groups whose choice of strategy is closely guided or constrained by professional standards, firm policies and procedures, and decision support systems. Moreover, a particular judgment in an accounting setting (such as a control reliance judgment) forms part of a more complex sequence of judgments (the formation of an audit opinion); and this sequence may be reiterated period after period, offering opportunities for various kinds of interperiod dependencies.

Clearly these characteristics of accounting environments have the capacity to influence judgment performance. We believe that examining their effects requires essentially the same conceptual and methodological approach as examining knowledge effects. Environmental characteristics affect performance by interacting with elements of the model shown in Fig. 3. To specify the exact nature of the interaction, one must understand the underlying cognitive processes involved in a particular task, the knowledge and ability they require, and the effort needed to access and apply that knowledge and ability.

Two principal classes of interactions can be distinguished. In one class, the environmental factor of interest changes the task requirements; that is, it changes the ability, knowledge, and/or effort that the decision maker must bring to the task to achieve a given level of performance. For example, without computer technology, correct performance of extensive numerical calculations requires either considerable ability, considerable effort or some combination of the two; with the technology, these requirements are reduced (although the knowledge requirement, e.g. how to use a spreadsheet, may be increased). In the second class of interaction, the environmental factor does not alter the requirements of the task but changes the amount or allocation of effort that decision makers are willing to employ to fulfill those requirements: that is, it changes motivation. As will be suggested below, many environmental factors characteristic of accounting settings influence performance both by changing input requirements for a task and by changing decision makers' motivation.
The recent literature has focused on four characteristics of the accounting environment: technology, group processes, prior involvement, and accountability. The rest of this section will discuss each of these characteristics in turn. We intend to show how each characteristic can affect both task requirements and motivation, and to illustrate the progress that has been made by employment of the three research principles: attention to interactions, understanding underlying cognitive processes, and experimental abstraction.

Audit technology

"Audit technology" is defined here to include the whole apparatus of guidance and support that is made available to auditors with the purpose of aiding and controlling their judgment. It includes not only formal decision aids but also a wide range of other guides to collecting, evaluating, and combining information, such as professional standards, firm policies, and documentation procedures. The quantity of such guidance that is available has increased in recent years; and correspondingly, research has advanced in this area, both in the design of experiments that can unambiguously explain the effects of such technology, and in the range and complexity of the effects that have been hypothesized.

Audit technology as a substitute for deficiencies. As early audit judgment research focused on deficiencies in human abilities that limited performance, so early audit technology research focused on decision aids as substitutes for these deficient abilities. Jiambalvo & Waller (1984), Butler (1985), Daniel (1988), Libby & Libby (1989), R. Ashton (1990), Kachelmeier & Messier (1990), and Emby (1990) investigated the capacity of aids to reduce the errors and inconsistencies in cue combination and the systematic biases in probability estimation that had been found in previous accounting judgment research.

Capturing the effect of this type of decision aid requires researchers to specify what cognitive processes are involved in the task with and without the aid (a requirement similar to guideline 1 of the "expertise paradigm"), and to create experimental tasks for which the observable implications of the aided and unaided processes are different (similar to guideline 2). Kachelmeier & Messier (1990) provide an example of the value of breaking down a judgment into distinct subtasks with different knowledge and ability requirements in order to explain the effects of an aid. They proposed that a quasi-statistical decision aid for audit sampling suggested by the AICPA would counteract the widespread tendency, demonstrated in basic judgment research, to overestimate the informativeness of small samples and thus undersample. This type of aid substitutes a formula for human combination of relevant cues. It is intended to change the task requirements by eliminating the need for knowledge of the appropriate combination rule and the ability to apply it consistently. However, Kachelmeier & Messier (1990) demonstrate that the aid can also have unintended effects on the task requirements.

In their experiment, subjects either estimated the sample size intuitively, supplied parameter estimates without making a sample-size decision (the researchers then used the AICPA aid to combine the parameter estimates), or estimated it themselves using the aid. As the authors predicted, sample sizes were larger in the second condition than the first. Moreover, sample sizes were larger in the second condition than in the third, indicating that subjects in the third condition tended to "work backwards", supplying parameter estimates that would result in a sample size estimate that more closely matched their intuition. Thus Kachelmeier and Messier were able to show that performing the combination subtask affected performance of the parameter estimation subtask in ways which would not have been apparent in an experimental design that did not separate these two subtasks.

Little research has been done thus far on decision aids as substitutes for task-specific knowledge, rather than abilities, although developers of expert systems stress the role of these systems in providing knowledge that the
Auditor has never acquired, or knowledge that
the auditor may fail to access at a given moment
(Graham et al., 1991; Willingham & Ribar,
1988). It is interesting to observe, however, that
the knowledge-providing functions of decision
support systems recently deployed in major
accounting firms correspond with some of the
needs demonstrated in the expertise literature.
DS (Broderick, 1988), a general support system
for auditors in the field, highlights the links
between control weaknesses, errors, and their
effects on accounts — links which Frederick &
Libby (1986) and Bonner & Lewis (1990) show
are weak or absent in novices’ (though not
experienced auditors’) memory structures.
Thus an unaided junior auditor may find it
difficult to think of what errors a given control
weakness might give rise to; but DS performs
this generation task and thus allows the junior
auditor to behave more like an experienced
auditor. It also links potential errors to relevant
substantive tests, another task where novices
have been shown inferior to experienced
auditors (Bonner & Lewis, 1990; Biggs et al.,
1988). Another system currently employed in
audits, Risk Advisor (Graham et al., 1991), warns
of inconsistencies in data gathered — thus,
again, enabling novices to behave more like
experienced auditors, whose better-developed
knowledge structures enable them to detect
inconsistencies in evidence more readily than
novices (Moeckel, 1990).

Audit technology as a complement to human
abilities and knowledge. In the cases cited
above, interactions between technology and
knowledge or abilities are in the form of
substitution relationships. That is, a given ability
or knowledge element affects performance in
the absence of the substituting technology, not
in its presence. With the technology, novices
and experienced subjects perform alike. Inter-
actions in the form of complementary relation-
ships are also possible, however, and have begun
to be addressed in more recent papers. In a
complementary interaction, an ability or know-
ledge characteristic affects performance in the
presence of the technology, not in its absence:
only experienced subjects perform better with
the complementary technology. Thus, as
noted above, Frederick (1991) showed that
experienced auditors had better memory for
internal controls when the order of presentation
was schematic than when it was taxonomic.
Here the technology characteristic (e.g. informa-
tion order in working papers) did not
substitute for knowledge the auditors lack: it did
not help novices. Rather, it enabled experienced
auditors to take more advantage of the know-
ledge they have. To the degree that better
memory improves judgment performance, we
may suppose that experienced auditors' per-
formance will be better when a firm's docu-
mentation technology presents internal control
information in schematic order than when it
does not.

An effect of this complementarity of know-
ledge structure and documentation technology
on judgment performance has been directly
demonstrated by Ricchiute (1992), who found
that, in making a going-concern judgment, audit
partners weighted information more heavily
when it was presented in schematic (causal)
order than in the order usually encountered in
the workpapers of the subjects' own firm.
Ricchiute proposed that presentation in a causal
order changed the task requirements by dimin-
ishing the cognitive effort required to evoke a
failing-firm or viable-firm schema and match the
evidence to the schema, and that (other things
equal) auditors would tend to make "the more
easily constructed decision". Thus the auditors
tended to conclude that there was substantial
doubt of continued viability when information
indicating failure was presented in causal
order and information indicating viability was
presented in working-paper order; but to
conclude that there was no substantial doubt
when the order conditions were reversed.

Ricchiute's experiment closely controlled the
presentation of information and the amount of
time subjects could spend on the task: informa-
tion was presented one item at a time on a
computer screen, and subjects could not return
to previous items. While this was clearly unlike
the way in which information is examined in
actual audits, it allowed the experiment to
capture the predicted effects of the ease of accessing the schema and matching it to audit evidence. Under such tightly constrained conditions, subjects could not offset the disadvantage of the workpaper-order presentation by spending extra time and effort rereading and reorganizing the information presented. Thus the experiment suggests that these time and effort costs could be avoided (an efficiency gain), while still allowing auditors to take evidence fully into account, if the documentation technology was changed to match their memory organization more closely.

**Audit technology and motivation.** Relatively little work has been done on the role of audit technology in altering motivation. One exception is McDaniel's (1990) examination of the effects of audit structure and time pressure on an audit sampling task, which considered both task-requirement and motivation effects. Structure clearly affected the cognitive requirements of the task, which was to obtain enough evidence to achieve 95% confidence that the error in inventory did not exceed the tolerable error rate. Subjects in the unstructured condition were given audit objectives and had to decide on procedures, selection methods, and sample sizes for themselves and then carry out the selected procedures; subjects in the structured condition only had to carry out listed procedures. Increasing time pressure (allowing a shorter period of time) limited the duration of effort that could be applied to the tasks and thus increased the effort-intensity (or knowledge or ability) requirements for achieving a given level of performance.

McDaniel suggested, however, that structure and time pressure could also affect motivation. The higher standards implied in either shorter allowances of time or the high sample-size requirements of the structured program induced subjects to exert more effort (sample more items correctly in a given time) although no extrinsic rewards were offered for doing so. McDaniel further suggested that the combination of both structure and very short time allowances would create so much stress that subjects' motivation would actually be reduced and performance would suffer in consequence. While this prediction was only weakly supported, additional investigation may be in order.

The potential for interaction of technology with other environmental factors, and thus the need for studying their joint as well as individual effects, was further demonstrated by R. Ashton (1990), who examined the effects of a decision aid, a monetary incentive, feedback, and a justification requirement on auditors' prediction of bond ratings. The incentive, the feedback, and the justification requirement each improved accuracy of predictions in the absence of a decision aid, but had no significant effect in the presence of the aid. Moreover, the three other factors reduced the variability of subjects' judgment only in the absence of the aid. Ashton proposed both cognitive and motivational grounds for these performance effects. He also raised important questions about other environmental factors' impact on how the aid was actually used. In the absence of feedback, justification, or a monetary incentive, subjects followed the aid's recommendations to a considerable degree. In the presence of these factors the aid played an important role in setting a minimum performance standard; but subjects followed its decision recommendations significantly less often.

**Group processes**

Full use of the conceptual and methodological tools presented here may also help to clarify and extend the results of the literature on group processes in auditing. Solomon (1987), reviewing the multi-auditor judgment literature, observed that results were contradictory, and that the reasons group performance exceeded individual performance in some instances but not others were unclear. Nor was it clear whether existing results could be generalized beyond the narrow range of audit tasks used in the group decision-making studies. As in the comparable situation in the early audit expertise literature, these difficulties may be resolved — and to some extent have begun to be resolved — by specifying the knowledge, abilities, and effort required for particular tasks, and the way
particular group processes are expected to affect the inputs available for these tasks.

Just as all knowledge is not equally relevant to every task, and all decision aids do not affect performance in the same ways, so all groups cannot be treated as conceptually identical. Solomon (1987) provided a list of group characteristics that must be specified (size, group structure, decision rule, type of interaction, practiced versus ad hoc groups, etc.), since groups that differ on these dimensions are likely to perform at least some tasks differently.

Groups as determinants of cue-combination rules. Like the technology research, early group-process research focused on abilities: assessing whether multi-auditor judgment reduced errors common to individual judgment. Trotman & Yetton (1985) were the first to investigate the mechanisms by which differently structured groups could reduce judgment error. They compared interacting seniors, non-interacting review of a senior by a manager, and mathematical composites of seniors in an internal control reliance judgment. While none of the group processes used in this study was identical to the group process actually encountered in audit firms, it was only by creating such artificially distinctive groups that the authors could begin to demonstrate the source of the advantages that “real world” group processes might possess. They reasoned that if the source of the advantage was interaction — the opportunity to discuss conflicting judgments — then the interacting group would be superior to the other two. If the source of the advantage was heavier weighting of the better judgment, the review and interacting groups would be superior to the equal-weighted mathematical composite. If there was an advantage in the sequential nature of review, the review group would outperform the others. If any “second opinion” improved individual judgment, but the manner of combining it did not matter, then all the groups would outperform individuals but would not differ among themselves. In Trotman and Yetton’s study, where the task was an internal control evaluation similar (but not identical) to that in early audit judgment studies, the last possibility proved to be the case: any of the group processes improved performance, but none of the groups outperformed others.

As suggested above, these results can be interpreted in terms of the interaction of particular group processes with the cognitive requirements of the task. We have noted that performance of the internal control task requires knowledge of the direction of the effect of each cue on overall strength, the relative importance of each cue to overall strength, and the ability to apply the cue weighting rule consistently. Since this is a “novice task”, for which virtually all subjects have the requisite knowledge, we should not expect interindividual knowledge differences that would provide an advantage to interacting or review groups, nor is there any evident characteristic of the task requirements that would provide an advantage to sequential review. A principal weakness of individual judgment is in random error arising from inconsistency; and any combination of two judgments helps to reduce random error.

As we might expect, these effects were task-specific. Trotman (1985) pointed out that heavier weighting of the better judgment would affect performance — and thus equal-weighted mathematical composites would perform less well — only on tasks for which there were substantial, and detectable, differences in expertise. In a more complex task (estimating dollar error), where subjects had better opportunities to identify the best judge, interacting groups did outperform mathematical composites. Here again, however, the performance of pairs of seniors working together was equal to the performance of managers reviewing seniors’ work and subsequently interacting with them (Trotman, 1985). This suggests that accuracy gains may not be the principal target of the

---

7 Solomon distinguished between groups and teams; for the sake of convenience we will use the term “group” here to mean multi-auditor combinations of any kind.
hierarchical nature of review in this type of task. The specific hierarchy employed is more likely aimed at insuring that the firm’s utility for Type I and Type II errors is reflected in those judgments, at establishing accountability relations through the performance evaluation system, and at training less experienced auditors. Thus, the existence of the hierarchy may have its main effect on the pre-review actions of the subordinate. This motivational possibility is discussed further in the next section on accountability.

Very few studies of group processes in auditing have been published since Solomon’s (1987) review, although there is general agreement on the importance of understanding multiperson processes in audit judgment, and many unsolved problems remain. For example, the very widespread use of review in auditing suggests that it has advantages not completely captured in research up to the mid-1980s.

Groups as memory aids. The broadening of focus of accounting judgment research to include a range of knowledge as well as ability issues has, however, begun to improve our understanding of the mechanisms of group judgment. These studies have again focused on the sources of gain from review; that is, how review can improve judgment performance by changing the cognitive requirements of the task or the knowledge, ability, or effort available to fulfill these demands. Johnson (1991) showed that when auditors worked together in three-person consulting groups, their memory for information from working papers was more complete and accurate than the memory of auditors working alone, and their recall deteriorated less over time than individual auditors’ recall did. Performance improvement from this source will depend on the degree to which retrieving knowledge in memory matters to performance on a particular task. Presumably, this improvement results because recall tasks are “eureka tasks” where if any one person in the group recalls an item, all will recognize its validity. This is probably a number-of-heads phenomenon versus a group-structure phenomenon. However, since only one structure was investigated, one cannot unequivocally rule out additional effects of group interaction.

Knowledge and memory effects of group processes include not only changes in the overall quantity of evidence retained in memory but also changes in the selection of evidence that is encoded in memory and subsequently retrieved. Libby & Trotman (1992) focused on a structural effect; the sequential nature of the review process and its effects on auditors’ attention to evidence in a going-concern decision. They suggest that this particular structural element results in offsetting biases which insure that adequate attention is given to decision-inconsistent information. While initial decision makers in their experiment were more likely to recall evidence consistent with their decision than inconsistent with it, reviewers with the same rank as the decision makers were more likely to attend to and recall evidence inconsistent with the initial decision, thus offsetting the original decision makers’ bias. These effects were dependent on the reviewers’ knowing the initial decision maker’s opinion, since it was this opinion that, combined with the reviewers’ knowledge of common features of failed and non-failed firms, made inconsistent information salient.

The purely sequential and non-interactive review in this study, and the absence of rank differences between reviewer and reviewee, are not normal characteristics of review in the natural environment. However, by controlling for subject interaction and experience differences, the study was able to capture the pure effect of sequential presentation, and eliminate the superior knowledge of reviewers or the value of reviewer–reviewee interaction as alternative explanations for the findings.

In general, it may not be practicable in the laboratory to create an experimental setting that bears a close superficial resemblance to actual audit teamwork in its totality. It is, however, both possible and methodologically desirable to break the process down into manageable parts, i.e. subgroups and subprocesses that can be operationalized and analyzed separately. Separate consideration of these subparts should be the
result not of ignoring the nature of the "real world" but of examining it so closely that individual components can be clearly identified and distinguished and their interactions specified. As the research principles discussed above suggest, an emphasis on cognitive processes and abstraction in the studies described here allowed some important task-group structure interactions to be uncovered. Experiments in more complex environments can then follow.

As noted earlier, studies of group processes raise questions of motivation as well as the questions of cognitive requirements that have been discussed above. Members of the hierarchical groups common to accounting settings may have different motivations as well as different knowledge or abilities, and different group structures may affect the motivation of members differently. For example, audit partners may not face the same incentives as junior auditors; and the prospect of a discussion with one's supervisor may result in different motivation than the prospect of a discussion with one's peers. While these differences in motivation may also influence the performance of groups, they remain as yet largely uninvestigated.

Prior involvement and accountability

Two further characteristics of accounting settings have attracted the attention of behavioral researchers: the fact that decision makers are accountable to supervisors, clients, or others, and the fact that they often have a prior involvement in the decision, e.g. through participation in a previous year's audit. Both motivation and cognitive-process explanations for the effects of these environmental features are possible.

Research on these features has begun more recently than research on technology and group processes, and is correspondingly less extensive. Gibbins & Emby (1985) were instrumental in noting the importance of accountability in audit settings, observing that "much of the information-gathering during the judgment process is in the service of providing the justification rather than determining a preferred decision or action". The nature of accountability relations differs depending on the party to whom the decision maker is accountable and the way in which those accountability requirements are manifested (continued employment by the client, legal liability to third parties, producing a written explanation for superiors, etc.). Most studies to date have focused on accountability among members of the audit team.

For example, Kennedy (1992) noted that accountability is a potential source of gain in the audit review process, since it may affect the subordinate auditor's judgment before a group process combines it with the superior's judgment. Accountability can serve as an effort-motivating device in various ways. While accounting convincingly for one's decisions to a superior in the natural environment may affect performance evaluations and thus, eventually, monetary rewards, accountability can also induce increased effort through non-monetary means: social pressure to comply with task demands, or the desire to preserve self-esteem by not appearing incompetent. R. Ashton (1990), Johnson & Kaplan (1991), Tan (1991) and Anderson et al. (1992) have all shown performance effects of accountability or justification manipulations, some in situations where no monetary rewards were at stake.

The limits of these motivation effects depend on the effort-sensitivity of the tasks involved and have begun to be clarified. Kennedy (1992) hypothesized that recency biases in audit judgment [demonstrated by Ashton & Ashton (1988), Asare (1989), Tubbs et al. (1990), Messier & Tubbs (1990), and Messier (1990)] were due to subjects' choice of a relatively low-effort information processing strategy, and could therefore be mitigated if subjects exerted more effort in examining the information given them. She also hypothesized, in contrast, that the curse-of-knowledge bias (cf. Camerer et al., 1989) was the result of cognitive mechanisms not under conscious control, and therefore not sensitive to changes in effort. Making subjects accountable at the beginning of a going-concern judgment task was expected to eliminate the
former bias but not the latter. Subjects who were not accountable (and, in the recency task, subjects who were made accountable after they had seen the information but before the judgment) were expected to display judgment biases. As predicted, MBA subjects exhibited a recency bias when they were not made accountable, or when they were made accountable only after they had seen the information; but early accountability eliminated the bias. In contrast, accountability manipulations had no effect on the curse-of-knowledge bias for either auditor or MBA subjects.

She was able to exclude alternative hypotheses because her design closely followed the guidelines of the "expertise paradigm". Two slightly different versions of the going-concern task were used, one designed to activate the recency bias, and the other to activate the curse-of-knowledge bias. By using two versions of the task with two different sets of cognitive requirements (like the two tasks in Bonner, 1990), Kennedy was able to exclude alternative explanations for her results. If she had used only the curse-of-knowledge version, the lack of accountability effects might have been attributed to a failure of the manipulation: subjects might not have been sufficiently convinced by the accountability "cover story" to exert more effort. However, since the same manipulation succeeded in affecting judgments on the recency version for the MBAs, this explanation is excluded for these subjects. The use of a treatment group in which subjects were informed of their accountability after seeing the information but before making their judgment confirmed that the information-examination phase was, as predicted, the effort-sensitive part of the task.

Accountability can also mitigate the judgment biases that arise from prior involvement in a decision. Previous experience with a client affects the auditor's knowledge, both in terms of overall quantity and in terms of systematic attention and memory shifts. A. Wright (1988) showed that acquaintance with prior year's workpapers alerted auditors to potential errors and risk areas in the current year's audit, but found no evidence that it caused them to "anchor" inappropriately on previous procedures. Tan (1991), however, showed that the effect of prior judgments depended on the degree of auditors' involvement in those judgments. Subjects making a going-concern decision tended to notice and recall information inconsistent with the prior period's judgment if they had not made that judgment themselves, but were given it as the result of a prior auditor's work. This tendency to attend to inconsistent information was significantly smaller if the prior year's judgment had been made by the subject. These results are somewhat comparable to Plumlee's (1985) finding that internal auditors reviewing an internal-control program were more likely to notice its strengths if it was one they had designed themselves in the past, and more likely to notice its weaknesses if it was designed by someone else.

Tan (1991) showed that an accountability manipulation reduced the effects of prior-period involvement and moderated the extremeness of auditors' judgments in the going-concern task. Like R. Ashton (1990) and McDaniel (1990), Tan's study illustrates the importance of examining more than one environmental factor at once, because of the potential for significant interactions. In Ashton's and McDaniel's studies, a factor (technology) that improved performance when considered in isolation had no effect when combined with other factors (justification, feedback, monetary incentives, or time pressure). In Tan's study, in contrast, a factor (prior involvement) that had a negative effect on performance was rendered less harmful by conjunction with another environmental factor (accountability). A full understanding of how the environment of

---

8 Auditor subjects did not exhibit recency bias in any condition. Kennedy speculated that the task — a going-concern judgment — was easier for the auditors, and that they may have developed habits of accountability that carried over into the laboratory task even in the absence of a specific accountability manipulation.
accounting decisions is organized and how changes in this environment affect performance will require further multiple-factor studies of this sort that capture interactions among the factors.

The literature on prior involvement has focused more on cognitive requirements and the accountability literature more on motivation as a mechanism for affecting performance; but there is some potential for each factor to work through both of these mechanisms. The question of motivation arises with prior involvement, since self-esteem and self-presentation concerns may induce people to be less critical of their own prior work than the work of others. The question of cognitive requirements arises with accountability, since writing (or anticipating writing) a rationale memo may have different requirements than simply making a judgment. It is not clear whether a given amount of effort employed in whatever way the subject prefers would have the same effect on performance as an identical amount of effort employed in writing a rationale. Some authors have made this suggestion, but the two effects have not been entirely disentangled.

Monetary incentives

Monetary incentives are often substantial for accounting tasks in the natural environment. Like other environmental characteristics, these incentives can affect performance either by motivating additional effort or by altering the cognitive requirements of the task.

Not all cognitive processes are equally sensitive to effort changes, and thus added effort will not improve performance to the same degree in all tasks or by all decision makers. Kennedy (1992) provided one illustration of this basic principle. Libby & Lipe (1992) provided another, showing that recognition memory is less sensitive to effort induced by a monetary incentive than recall memory. The task — memorizing a set of internal controls — was not identical to tasks usually performed in audits, which requires a combination of memory and judgment processes. The relatively artificial task, however, allowed a clear distinction between incentive effects on the two different memory processes; and this in turn has relevance to actual audit practice. Different audit technologies are differentially dependent on recognition and recall; in general, the more structured the methodology, the more recognition and less recall is required. Performance with different audit technologies may therefore be differentially sensitive to the provision of incentives, although this effect remains to be demonstrated.

Incentive-induced effort may also interact with abilities. Awasthi & Pratt (1990) found that monetary incentives increased effort (time spent) on judgment tasks but improved performance only among subjects who had scored high on a test of their fundamental ability to abstract familiar concepts from complex settings. In a similar fashion, Libby & Lipe's (1992) results suggest the possibility that incentive-induced effort may interact with knowledge. A performance-based incentive resulted in effort increases in a memory task, but these effort increases were more effective for the more knowledgeable subjects when memory traces were weak.

Monetary incentives, like other environmental characteristics, may influence performance not only by changing the amount of effort people are willing to exert, but also by affecting the cognitive requirements of the task. Luft (1992) showed that the structure of a monetary incentive attached to the outcome of a task affected the memorability of outcomes and thus affected judgments dependent on memory for outcomes. Subjects received a combination of base pay and a performance-based incentive resulting in identical monetary outcomes in both treatment conditions, but structured as a bonus for meeting a standard in one condition and a penalty for not meeting a standard in the other condition. The bonus condition received $2 ("experiment dollars" scaled down for the final payoff) as base pay in each period plus a $20 bonus for meeting a performance standard; subjects in a penalty condition received $22 base pay minus a $20 penalty if they failed to meet the performance standards.
to the "success" outcome was expected to make successes more easily memorable for subjects in this condition than those in the penalty condition, while the penalty attached to the "failure" outcome was expected to make it more easily memorable for subjects in the penalty condition. Thus, the effort requirements for storing and retrieving information about past outcomes differed under the two different incentive schemes. Subjects who had worked under a penalty scheme for 12 periods believed they had performed less well than equally performing subjects who had worked under a bonus incentive. Correspondingly, the difference in the attractiveness of the incentive contract between bonus and penalty subjects became greater with experience.

**Conclusion**

Accounting decisions are often made in structured institutional settings. Research on environmental effects on judgment has begun to help us understand how these settings function and why they are organized as they are. The effects of differently structured decision-making groups, support technology, compensation schemes, accountability requirements, and multiperiod involvement in judgments have all been explored to varying degrees.

The literature on environmental factors has benefited from the growth of expertise studies. As we have learned more about the role of knowledge in determining performance, it has become possible to develop a fuller picture of the effect of environmental factors, including not only their interactions with ability but also their interactions with knowledge. The expertise literature has also provided an experimental paradigm for capturing influences on performance, which can be used for environmental as well as knowledge studies.

A difficulty in dealing with environmental factors is that what appears as a single factor from the point of view of practice — a compensation scheme, a review process, a decision support system — can, from the point of view of theory, be a bundle of conceptually distinct factors that affect performance by different mechanisms. As the papers discussed above show, an environmental factor may simultaneously affect the cognitive requirements of a task, perhaps in more ways than one, and the decision maker's motivation to fulfill those requirements. Full understanding of the role of a particular environmental factor requires researchers to distinguish the various underlying mechanisms by which it affects performance (cf. research principle 2, above), and this in turn may require considerable abstraction (research principle 3).

Successful abstraction begins with the natural environment but does not end there. An examination of the "real world" of accounting helps us judge what kinds of technology, what group processes, etc., are important to study; it also serves as a basis for predicting what specific aspects of these factors are key to influencing performance. An unambiguous test of these predictions, however, will often require a degree of experimental control that excludes much superficial "realism" from the laboratory.

**DISCUSSION AND FUTURE DIRECTIONS**

In 1976, Einhorn described the typical setting for accounting studies of that era as characterized by well-defined tasks where the subject was given perfectly reliable information of some sort, and the range of hypotheses or actions that could be entertained were highly restricted by the dependent variable. He suggested further that many real-world accounting tasks are ill-defined, information must be searched for, data are rarely perfectly reliable, and hypothesis formation is necessary as well as hypothesis testing. While many studies which followed responded to Einhorn's criticism, accounting settings in the early 1980s were still represented for the most part as single-task, single-person, single-period, single-structure settings, populated by homogeneous limited ability/knowledge individuals. In more recent years, our conceptions of these settings have become increasingly rich and representative, recognizing that accounting settings involve multiple
tasks that vary in important respects: the abilities, knowledge, and effort required; the knowledge, status, abilities, accountability relations, and incentives of the individuals assigned to the tasks; the forms of guidance and aid available; and the number of periods over which the task is performed.

We have reviewed portions of several studies which we believe successfully examine elements of this more complex environment. While the review is not by any means exhaustive, the studies discussed illustrate how successful studies recognize three research principles: that the determinants of performance interact, that understanding these interactions requires an understanding of the basic cognitive processes required by the task, and that abstraction based on careful examination of task and psychological and accounting theory is necessary to disentangle these dependent relations. Examples of these dependencies include the effects of knowledge organization interacting with organization of workpapers, prior involvement with accountability relations, incentives with knowledge, incentives with task structure, group structure with task structure, and ability with task structure. We further suggest that care in specifying a study's purpose, conceptual variables and relations, and employment of designs sufficiently powerful to uncover the effects of interest and eliminate alternative hypotheses, while always important, take on a larger role in examinations of these more complex issues.

The complexity of these issues and the need to disentangle related effects leads to great diversity in purpose among individual papers. Recognition of this diversity is necessary in evaluating contributions to the literature and the appropriateness of various research approaches. For example, some studies of knowledge attempt to document important aspects of auditors' knowledge (e.g. Brown & Solomon, 1991; Libby et al., 1985), while others focus on how training and experience give rise to knowledge (learning). Of those that examine experience effects, some manipulate experiences in the laboratory (e.g. Butt, 1988; Nelson, 1992), while others observe the effects of experience accumulated in the field (e.g. Libby & Frederick, 1990). Some of these same studies directly assess the effects of knowledge using memory tests (e.g. Frederick, 1991), while others infer knowledge differences based upon systematic judgment differences (e.g. Abdolmohammadi & A. Wright, 1987), and still others simultaneously assess knowledge and judgment differences and their relations (e.g Heiman-Hoffman, 1992). Comparable differences in purpose and approach are evident in studies of environment and motivation. The appropriateness of these choices must be judged by their match to the purpose of the paper. We suggest that the model presented in Fig. 3 will provide a basis for more systematic definition of research purpose which will provide guidance for these choices.

The growing complexity of the issues addressed in the developing literature also provides added impetus for consideration of the interactions of the determinants of performance. Future research must recognize that the determinants of performance have multiple effects including both direct effects on performance and indirect effects on the other determinants. For example, decision aids can substitute for knowledge, place effective limits on the actions decision makers may take, and at the same time determine what is learned from the environment and brought to the next round of decision making. Similarly, as suggested by the model in Fig. 3, ability has direct effects on performance, and indirect effects as a determinant of learning (Libby & Tan, 1992). We recommend that these more complex effects be given serious consideration in future research.

The multiperiod nature of most audit tasks suggests a number of highly important relations that must be examined. For example, while many studies focus on how experience results in knowledge acquisition, a few recent studies (e.g. Moeckel & Plumlee, 1989; Moeckel, 1990) suggest how prior knowledge can significantly affect what can be learned from experience in the current period. Such possibilities deserve further attention.
Consideration must also be given to the fact that any factor that increases motivation to perform well in multiperiod settings can result not only in actions taken to improve the current period's performance, but also actions that have no effect, or a negative effect, on current performance but are intended to improve performance in the future. We do not fully understand the effects of motivation on long-term learning, nor do we fully understand other environmental influences on learning.

In the longer term, environmental factors can influence performance in a variety of ways by affecting the acquisition of knowledge. What accounting decision makers learn on the job may depend in part on whether they work alone or in a particular kind of group. One of the expected benefits of the hierarchical groups common to accounting settings is the opportunity for less-experienced members to learn from more-experienced members (Dirsmith & Covaleski, 1985). What decision makers learn may also depend on the technology they use. Arguments can be made that added structure aids learning by systematizing the knowledge components. Decision support systems can also change the quantity and timing of policy guidance and feedback provided to auditors. At the same time, the possibility exists that routinization of tasks results in mechanical responses which make decision makers insensitive to novel situations or changes in the environment.

No individual study we are aware of looks at interaction of type of group process and type of task: that is, one kind of group may do better at one task, another kind at another task. But we might expect such interactions, given the variety of group structures and processes that are used in audit firms. Task forces for special issues are not necessarily structured the way audit teams are; and within audit teams, the composition and interaction of subgroups vary with the task: for example, the decision to accept a new client does not involve the same people interacting in the same way as the decision to rely on internal control. A fuller specification of the cognitive requirements of different tasks, and the capacity of different groups to fulfill these requirements, would add to our understanding of the way auditing and other accounting work is organized. Trade-offs among organizational control devices such as review processes and decision aids, both as tools to limit actions that can be taken and as training devices, are also worthy of study.

As R. Ashton (1990), Libby & Lipe (1992), and Luft (1992) point out, the effects of monetary incentives depend not only on their magnitude but also on their form. A tournament incentive, for example, may have different effects on performance than a piece-rate incentive with the same expected payoff. Monetary incentives employed in the natural environment are often considerably more complex than those employed in the laboratory, and much remains to be learned about the effects, both for cognition and motivation, of the particular forms these incentives take.

Nearly all of the research discussed here was conducted in the audit context. This reflects the emphasis in the literature since 1980. Recently, there has been growing interest in investigating related issues in financial analysis (e.g. Libby et al., 1987; Moser, 1989; Maines, 1990) and other accounting domains. We believe that there are many important issues related to the effects of knowledge, ability, environment, and motivation in credit and investment analysis judgments. Understanding how aspects of these institutional settings affect learning and performance of these tasks will add greatly to our understanding of accounting-related judgments. We recommend a similar approach to that taken in the audit literature, beginning with an analysis of key attributes of the settings and task requirements. Many may be similar to the attributes of audit settings and tasks described above, but key differences are likely. A similar approach in other areas of accounting could also be taken.

An audit, a cost analysis, and the preparation of a set of financial statements are all complex activities that involve many separate judgments. Experimental research has tended to examine component judgments singly, for reasons of simplicity and control. We have continued to
recommend such an approach here, but a final note of warning is relevant. The question remains: is the optimal audit opinion (for example) achieved by optimal performance on each component judgment? Or, if we wish to improve performance, might it be more efficient to allow relatively poor performance on one difficult (costly) component task if it can be compensated for by a less costly alteration on some other task? Pincus's (1989) finding is suggestive: auditors performed better on a management fraud judgment when they collected smaller and more biased information sets than when a checklist forced them to collect more extensive and balanced information. It is possible that auditors had developed judgment strategies that worked well with the “cheaper” information set. If interdependencies of this sort have developed in the natural environment, it may be hazardous to try to improve a single component judgment without considering “downstream” effects.

Two final recommendations are in order. The types of interdependencies we have discussed above are quite complex. Efficiency demands that we often learn from studies of analogous tasks in other fields. The difficulty of doing so is open to debate. For example, one might suggest that it is harder to infer effectiveness in auditing than in medicine because few audits fail and, as a consequence, there is little outcome feedback. However, we suggest that indeed the situations are similar since most patients get well regardless of the treatment, and many treatments that are effective such as antibiotics are effective against a broad range of diseases, so that the treatment may succeed even if the diagnosis is incorrect. Even autopsies of patients whose treatment has failed rarely tell the physician which alternative treatment would have been more effective. In both auditing and medicine, corrective action requires a signal that an error has been made and the means of correction, which requires knowing why the error occurred. It is clear that in both the medical and many accounting settings, one or both may often not be available. We propose that careful analogies should continue to be a major source of new knowledge concerning our domain.

The final recommendation is a warning. All of the literature discussed here is aimed at adding to our knowledge of the determinants of performance. But it should be recognized that, while experiments are powerful tools for assessing the factors affecting performance and their dependencies, they are generally ill-suited for estimating the magnitude of effects or absolute levels of performance. This results because most experiments do not sample tasks and cases randomly from the environment. As a result, absolute statements about levels of performance or comparisons between accounting and non-accounting tasks are often inappropriate.

BIBLIOGRAPHY


Libby, R., Determinants of Performance in Accounting Decisions, in Accounting Research Convocation (University of Alabama, 1983).


