The Use of Deep Feature and Surface Feature Information in the Performance of Complex Audit Tasks

Christine E. Earley
Assistant Professor
University of Connecticut
SBA/Accounting, U-41A
Storrs, CT 06269
cearley@sba.uconn.edu
(860) 486-1278

September 2000

I wish to express my appreciation to Vicky Hoffman (dissertation chair), and the rest of my dissertation committee including, Micki Chi, Jake Birnberg and Donald Moser for their help in making this research possible. I would also like to acknowledge individuals from an international public accounting firm for assistance in providing subjects and help in designing the experimental materials, and the financial support of the Deloitte and Touche Foundation. Finally, I would like to acknowledge the helpful comments of Noah Barsky, Jennifer Joe, Rick Tubbs, and workshop participants at the University of Colorado, University of Connecticut, Indiana University, the University of Maryland, and the University of Washington.

The Use of Deep Feature and Surface Feature Information in the Performance of Complex Audit Tasks

This study examines auditor performance with respect to complex tasks. Previous research in both psychology and auditing suggests that expert/novice differences are likely to occur during the problem representation phase of problem solving. In particular, when representing problems, experts focus on “deep structures” or patterns in data, and novices focus on “surface features” (Chi et al. 1981). Few studies in applied domains such as auditing have examined how differences in problem representations affect performance in an experimental setting. The research question addressed in this study is: “To what extent is the superior performance of auditors with a high level of task-specific knowledge and experience attributable to their focusing on deep structures or patterns in data while solving problems?” The task studied is the audit of real estate valuations. This task was selected because it represents an ill-structured problem for which knowledgeable and experienced auditors would be predicted to have a comparative advantage.

The design of the experiment was a 2 X 2 X 2 full-factorial design with one between-subjects factor (experience level), and two within-subjects factors (deep features and surface features). The purpose of the experiment was to demonstrate that more experienced auditors exhibit superior performance with respect to a real estate valuation task, and that this superior performance was due to the recognition of underlying patterns of information about each property.

The results indicate that the experienced participants outperformed novice participants when determining the reasonableness of discount rates used in property valuations. Predictions regarding the use of underlying patterns of individuating information by experienced participants were partially supported. It was also predicted that novices would perform worse than experienced auditors because of their tendency to focus on information other than individuating patterns, and this prediction was also partially supported. An unexpected result was that experienced auditors exhibited a tendency to commit Type II errors in performing the task. An explanation for this result and implications of these findings are presented.

Key Words: *Auditor Expertise, Problem Representation, Cognition*.

Data Availability: *The data is available from the author upon request.*

INTRODUCTION

Past studies of expertise in psychology and auditing have contributed greatly to what is known about experts' processing, or what Bédard and Chi (1993) refer to as "invariants of expertise." Many auditing studies are based on models of expertise which focus on how various combinations of knowledge, ability, motivation, and environment impact expert performance (Libby 1995; Libby and Luft 1993). Many studies in psychology, on the other hand, have adopted more of an information processing approach, studying experts and novices as they solve both well-structured and ill-structured problems (e.g. Newell and Simon 1973). These psychology studies have identified the development of the problem representation as a critical phase in problem solving for both experts and novices, and it is during this phase of problem solving that many expert/novice differences become apparent.

Despite the importance of the problem representation phase in problem solving, there have been very few auditing studies that have examined problem representations directly. Studies such as Bedard and Biggs (1991a;1991b) and Christ (1993), have examined differences in expert and novice representations, but there have been mixed results with respect to the link between problem representation and performance on audit tasks. In addition the Bedard and Biggs (1991a;1991b) studies utilized a process-tracing approach and a relatively small number of subjects, and the Christ (1993) study involved an examination of auditors' free recalls. To date, no auditing studies have examined the link between problem representation and performance in an experimental setting.

The purpose of this study is to experimentally address the differences between auditors with a high degree of task-specific knowledge and experience and auditors with a low degree of task-specific knowledge and experience in their initial representation of problems, and to examine how these differential representations impact performance with respect to complex auditing tasks. Specifically, it is expected that experienced auditors will focus on patterns or deep structures in information when forming their representations, and novices will focus more on surface features or individual cues (consistent with the findings of Chi et al. 1981). The research questions to be addressed is, “To what extent is the superior performance of auditors with a high level of task-specific knowledge and experience attributable to their focusing on deep structures or patterns in data while solving problems?” The task used in this study is real estate appraisal review.[1] It was selected because it represents a highly complex and ill-structured task which normally requires substantial knowledge and experience to perform.[2]

In addition to drawing upon the results of past expertise studies in psychology and auditing, this research draws upon the findings of Earley (1998), an observational field study in which one experienced auditor and several novice auditors were observed while they performed appraisal review and real estate valuation testwork on an actual audit engagement. As a result of the field study, several features of the experienced auditor’s processing were identified as being important in performing the real estate valuation task. Specifically, the experienced auditor’s ability to recognize patterns appeared to contribute to his superior performance. Because Earley (1998) was only an observational study performed with very few participants, the results are not generalizeable; however, they do provide a basis from which to make predictions that can be tested in an experimental setting.

An experiment was designed which examines the extent to which deep features and surface features are utilized in judgments of more and less experienced auditors. The hypotheses were developed based on both existing literature and the observational field study of auditors discussed previously. By studying auditor behavior in an applied setting, and then relating it to what is generally known from research in psychology and auditing, this study attempts to avoid the pitfalls associated with what Solomon and Shields (1995) refer to as a "borrow and test" approach to expertise studies in auditing[3].

The remainder of the paper is organized as follows: relevant observations from Earley (1998) and theory from cognitive psychology and auditing is presented in the next section, and hypotheses are developed. The third section contains a description of the experimental method, and the fourth section contains a description of the experimental results. Discussion and conclusions are provided in the fifth section.

THEORY AND HYPOTHESES

Problem Representation

In Earley (1998) it was noted that before determining the reasonableness of the dollar amounts in property portfolios, both the experienced and novice auditors spent a great deal of time developing an understanding of each property to be valued. This process of understanding is known in the problem solving literature as developing a problem representation, and has been shown in other studies to be a key component of the problem solving process (see discussions in Bédard and Chi, 1993; Chi et al., 1982, and Newell and Simon, 1972). As described in Newell and Simon's (1972) General Problem Solver (GPS) model, the representation phase is when the problem solver defines the goal, the constraints, and the operations to be performed in order to arrive at a solution. In addition, previous studies in psychology have shown that the quality of the experts' problem representation (including the processes employed and the knowledge structures utilized to represent the problem) is what enables them to solve problems more efficiently and effectively than novices (Voss et al. 1983; Voss 1987; Frensch and Sternberg 1989; Hayes and Simon 1976).

Several auditing studies have examined differences in how experts and novices represent problems (see, for example, Bedard and Biggs 1991a; 1991b; Bierstaker et al. 1999; Christ 1993). In Bierstaker et al. (1999), it was shown that auditors who were prompted toward the problem representation that contained the correct solution experienced enhanced performance in an analytical procedures task. Christ (1993) provided evidence of differences in problem representations among more and less experienced auditors, but the link to performance in the audit planning context was unclear. The current study attempts to clarify the link between problem representation and performance by focusing on certain features of experts' and novices' problem representations, and showing that these features have a direct impact on performance in the real estate valuation task. One of these features is the experts' tendency to focus on patterns of information in representing problems, which is discussed below.

Pattern Recognition

Earley (1998) observed that one of the components of the experienced auditor's processing that enabled him to perform the real estate valuation task more effectively than the novices, was the ability to notice relationships between items of information. The experienced auditor spent a great deal of time developing a plausible "story" about each property to support his conclusions of reasonableness. He often drew upon examples of other properties and compared the current property to previously encountered examples in developing this understanding. This finding is consistent with findings in other observational studies which model experts' problem-solving in complex domains. For example, Klein et al. (1989; 1993) and Klein and Calderwood (1991) performed extensive studies of expert firefighters and aviators on the job, and found that once experts represent the problem, a number of solution paths follow readily. The previously encountered examples allow the expert to abstract enough knowledge to quickly generate a number of possible solutions. Each solution path is mentally evaluated in turn until the most viable solution is determined. Klein (1999) refers to this as the Recognition Primed Decision (RPD) model of decision making. Hershey et al. (1990) performed a study in a financial setting and found support for a model of decision making that is consistent with RPD.

Gick and Holyoak (1980) refer to the process whereby a current example triggers a previous example as "schema activation". Gick (1986, p.102) defines a schema (plural *schemata*) as "a cluster of knowledge related to a problem type. It contains information about the typical problem goal, constraints and solution procedures useful for that type of problem". Schemata are often thought of as being structures that individuals possess, which are *internal* to the problem-solver. In the context of analytical procedures in auditing, Bedard and Biggs (1991a) studied pattern recognition processes among auditors which can be viewed as instances of schema activation. They describe pattern recognition as "recognizing relationships among pieces of financial information, where a concept or causal agent underlies the relationship" (p. 624). Patterns in information are *external* to the problem-solver, but schemata *within* memory allow one to recognize the relationships between the pieces of information making up the pattern. Additional research in psychology has demonstrated a link between experts' use of schemata and superior performance (e.g. Gick 1986; Chi et al. 1981).

Structure of Knowledge

It is clear from previous research that the ability to recognize patterns is largely a function of the schemata that experts possess. Experts are expected to perform better than novices due to the fact that experts possess more and richer schemata. Research on the nature of expert versus novice knowledge structures has provided a number of insights into differences in the schemata of experts and novices.

Chi et al. (1981; 1982) focused on the structure of expert and novice knowledge. Specifically, they examined how experts and novices categorized problems when they initially encountered them. This is an important step in representing problems, because often the type of problem encountered will determine the appropriate solution to be applied. In these studies regarding the structure of knowledge, the manner in which individuals categorized problems provided insight into how schemata were organized in memory. For example in the physics domain, Chi et al. (1981) found that when asked to categorize problems based on how one would go about solving them, experts classified the problems according to underlying laws or principles, (such as conservation of energy), which Chi et al. refer to as “deep structures” (or “deep features”), while novices categorized on the basis of “surface features” of the problem, such as whether it was an inclined plane or a rotational problem. The categorization of problems according to deep features enables experts to arrive at a solution more quickly because the underlying principle activates the appropriate schema. This research regarding physics experts is consistent with the results of Chase and Simon (1973) in which the superior performance of expert chess grandmasters was attributed to their greater understanding of meaningful chess moves, which were arranged in the form of patterns on the chessboard (novices, on the other hand, viewed patterns of chess moves as meaningless or as random combinations of pieces).

Studies in auditing which have examined the organization of knowledge and its impact on judgments have resulted in findings that are consistent with those in psychology. Frederick (1991) found that there were qualitative differences in the way that experts and novices represented knowledge, with expert performance being more sensitive to the organization of presented information. Expert auditors performed better on a recall task when information was presented schematically as opposed to taxonomically, whereas novices were insensitive to the type of representation provided (presumably due to their inability to focus on deep features). Similar to Frederick (1991), other studies such as Frederick et al. (1994) and Nelson et al. (1995) demonstrated that when classifying financial statement errors, experts focused more on deep features and novices focused more on surface features.

To summarize, findings from cognitive psychology and auditing have underscored the importance of the problem representation phase of problem solving. One mechanism that occurs during the problem representation phase is the activation of a relevant schema, which relies on the recognition of relationships between cues (a process which is referred to as pattern recognition). Because experts have more examples of previously-encountered patterns (and therefore more schemata) in memory, and because they are able to categorize newly-encountered problems on the basis of such schemata, they are often more efficient and effective problem-solvers. Based on this literature, the following hypotheses are proposed:

H1: Auditors who are experienced with respect to real estate valuation will perform better than novices in making complex real estate valuation judgments.

H2: Auditors who are experienced with respect to real estate valuation are expected to focus on the underlying pattern of cues in information presented to a greater extent than novice auditors.

H3: Auditors who are novices with respect to real estate valuation are expected to focus on surface features of real estate valuation problems rather than the underlying pattern of cues in information presented.

Note from the above hypotheses that differences both in the use of deep feature (pattern) information and surface feature information is expected to result on performance differences between experts and novices. For example, in Earley (1998) it was noted that the expert auditor first focused on comparable industry rates (surface features) to establish an expectation about the reasonableness of the discount rate used in the appraisal, but it was the individuating pattern of information (deep features) about each property that provided the basis for the judgment of reasonableness. The experienced auditor combined the individuating cues about each property into a “story” that guided his further judgments. Novices, on the other hand, examined property cues in a more haphazard manner and were often unable to make meaningful links between different items of information. They tended to focus on surface features individually, such as comparable industry rates or other individual cues, and evaluated each of these in isolation, rather than linking the cues into meaningful patterns. For novices, it is possible that judgments of reasonableness may be driven by over-reliance on particular surface cues.

In order to determine if the experienced-novice differences for processing information found in Earley (1998) hold in an experimental context with a larger number of subjects and cases, both patterns and individual cues are manipulated in this study. The patterns consist of a number of cues of individuating information about each property as described above. The auditors are required to evaluate a given property discount rate, and the pattern of cues is either designed to be consistent (termed the “match” condition in this study), or inconsistent (“no match”) with the given rate. The manipulated “surface feature” cue is a range of comparable industry discount rates, and again the given discount rate either falls within (“matches”) the industry average or outside (“doesn’t match”) this range. The patterns for each property (deep features) and the industry average ranges of discount rates (surface features) were manipulated in a 2 X 2 design as depicted in Figure 1. Predicted performance in each of the four cells for experienced and novice participants in the experiment is also presented in Figure 1.

Hypotheses 1, 2, and 3 taken together predict that although experienced auditors will outperform novices in *total* (H1 predicts a significant main effect of experience level), the performance differences are not expected to occur in all cells. Specifically, for experienced participants only, it is predicted that they will perform equally well in all cells (indicating no significant main effects or interactions predicted within the experienced condition). Novices, on the other hand, are expected to perform well in those cells where surface features and deep features are in agreement (cells 1 and 4), and poorly in those cells where deep and surface features produce competing signals (cells 2 and 3), due to their tendency to focus on surface over deep features. This indicates no predicted main effects for novices, but a significant deep feature by surface feature interaction (a positive crossover interaction). Because this interaction is predicted differentially for experienced versus novice auditors, H2 and H3 taken together indicate a significant experience level by deep feature by surface feature interaction (a 3-way negative interaction). In addition, these predictions also result in an overall significant deep feature by surface feature interaction (a 2-way positive interaction). All predictions depicted in Figure 1 are tested in the experiment, which is described in detail below.

METHOD

Participants

Participants were 39 auditors from an international public accounting firm, who were attending a training session for the firm's real estate industry group. The majority were audit seniors (37 of the 39 participants), with a mean of 2.9 years experience. Thirty-eight of the 39 auditors reported spending at least 10-25% of their time on real estate engagements (one participant had not participated on a real estate engagement before), and 28 had performed some sort of valuation-related testwork in the past, but as discussed in more detail below, most participants had very little task-specific knowledge and experience performing the types of valuation judgments in this study.

The participants were divided into experienced and novice groups based on their experience with real estate appraisal review.[4] They were asked to indicate the number of individual property valuations they had performed in the past year by selecting one of 5 categories. Participants' responses generally fell at the extreme ends of the scale; on one end were the responses "Zero appraisal reviews" and "1 to 5" properties, which were selected by 22 of the 39 participants, who were classified as "novices". On the other end were the categories "11 to 20" and "Greater than 20" properties valued, which were selected by 13 participants,[5] who were classified as "experienced." The remaining four participants selected the category "6 to 10" properties. Because it was impossible to determine whether the number of properties valued was closer to 6 or to 10, the four participants in this group were classified as novices (which works against finding differences between experienced and novice participants). Thus, the category of "11 to 20" properties was selected as the cut-off; this classification scheme resulted in a total of

13 experienced and 26 novice participants. Other task-specific knowledge and experience measures confirmed that this classification was appropriate.[6]

Task Materials

As mentioned previously, the task used in this study was the determination of whether the discount rate given in a property appraisal was reasonable given the background information about the property. Property cases were developed based on actual audited properties held by a large national real estate investment company. The names, exact locations, and dollar amounts associated with each of the properties were changed to disguise them, but the pattern of information about each was left intact. An example of one property case is presented in the Appendix.

A real estate audit manager from an international public accounting firm (the same firm that provided the experimental participants) helped to design each of the property cases from the actual property information given. This provided some assurance that the cases were realistic and that the information provided was typical of that encountered on an actual audit. Property cases were developed to be consistent with the four cells in the experimental design (described further in the next section). For example, underlying patterns of information were designed to be either consistent or inconsistent with the discount rates provided in the property appraisal, and the discount rates were either consistent or inconsistent with industry averages. Thus, because each case had one correct answer (Reasonable or Not Reasonable) based on the underlying pattern of information in each case, there was an objective criterion measure available for evaluating each judgment.

Once the cases were designed, the instrument was pilot tested with four auditors from the same firm (two partners, one manager, and one senior), who had significant experience performing real estate valuations (all would have been classified as “experienced” for purposes of this study). The pilot subjects provided feedback about the amount of time needed to complete the experiment and the realism of the task. Based on their comments, the cases were modified as necessary to make them more realistic.

Design and Procedure

The experiment was a 2 X 2 X 2 mixed factorial design, with one between-subjects factor (experience level), and two within-subjects factors (deep features and surface features). As mentioned previously, experience level was a measured variable with groups split on the basis of experience with the task. Deep features, or the individuating pattern of information about each property, was a manipulated variable with two levels: “match” or “no match.” That is, the individuating pattern of information about the property was either consistent with the discount rate per the appraisal review (match) or inconsistent (no match). Surface features, or the range of rates provided by an outside industry report, was also a manipulated variable with two levels, again “match” or “no match.” The rate per the appraisal was either within (match) or not within (no match) the range of rates per the industry report. Each participant completed two cases per cell, resulting in a total of eight [(2 X 2) X 2] individual property cases.

Each participant was presented with a booklet containing instructions and the eight cases, followed by a debriefing questionnaire. The order of the eight cases was counterbalanced to control for possible order effects.[7] The participants were instructed to assume that they were reviewing summary appraisal information about each property on an actual audit engagement. Their task was to determine whether or not the discount rate was reasonable based on the information given. If the participants answered “Not Reasonable,” they were asked to provide a reasonable range for the correct rate. In addition, they were asked to provide information regarding reasons for their answers; however, this section was marked "Optional". The participants evaluated each case one at a time, giving their judgment for each case before proceeding to the next case.

After completing the judgments for all eight cases, each participant answered a debriefing questionnaire providing demographic information such as number of years with the firm, amount of time spent on real estate engagements, etc. In addition, there were items designed to test the participants' general real estate and valuation knowledge, such as a list of real estate terms they were asked to define. There were also items designed to assess task-specific experience, such as the number of valuations performed, the nature of the valuation experience, property types valued, etc.[8]

RESULTS

Data from the set of eight cases were analyzed to test the three hypotheses. A categorical modeling procedure was employed in performing the analyses. Because the dependent variable is dichotomous (either correct or incorrect), and involved repeated measures (each subject saw all eight cases), a procedure was employed which provided weighted-least-squares estimates of the model parameters.[9] Once the parameters were estimated, Chi-square test statistics were calculated to test the significance of the main effects (experience level, deep features, and surface features) and interactions in the model. The summary results, including the Chi-square values and significance levels, are presented in Table 1.

Recall that Hypothesis 1 predicted that experienced auditors would perform better than novices on the overall judgments of reasonableness. The mean number of correct judgments was 6.4 (s.d. = .77) for experienced participants, and 4.8 (s.d. =1.50) for novices. The difference in means was statistically significant ($t = 3.63$, p(1-tailed)$< .001$). In addition, the main effect of

Experience Level was significant as indicated in Table 1 ($\chi^2 = 22.18$, $p < .001$). Thus, H1 was supported.

Hypothesis 2 predicted that reliance on deep structures accounts for the superior performance of experienced auditors, and H3 predicted that reliance on surface features over deep features accounts for the poor performance of novice auditors. As noted previously, these two hypotheses combined predict a deep features by surface features by experience level interaction (a negative three-way interaction), and a deep feature by surface feature interaction (positive two-way interaction). As noted in Table 1, the three-way interaction is negative, but not significant ($\chi^2 = .09$, $p = .954$), and the two-way interaction is positive and significant ($\chi^2 = 8.51$, $p = .01$) which would suggest mixed support for H2 and H3. However, these hypotheses led to specific predictions for each cell of the contingency table in Figure 1. A further analysis of the results in terms of these predictions is presented in Figure 2.

Hypothesis 2 predicted that pattern recognition would account for the superior performance of experienced auditors. It was predicted that experienced auditors would have a comparative advantage over novices when the deep features, but not the surface feature (industry average) was consistent with the correct answer (cells 2 and 3). As noted in Table 1, there was a significant main effect of deep features overall ($\chi^2 = 6.81$, $p = .03$), which was not as predicted. This indicates that for cases in which the pattern of information was consistent with the discount rate per the appraisal, both experienced and novice participants performed better than for cases in which the pattern was not consistent with the rate. In cell 2, the fact that the pattern was consistent with the rate given was more helpful to experienced than novice participants as evidenced by the fact that experienced participants performed significantly better than novices (t = 2.39, p(1-tailed) = .01), as predicted. However, in cell 3, the fact that the pattern was not consistent with the rate seemed to impair the performance of experienced as well as novice participants. In this cell, although the mean number of correct judgments was higher for experienced participants than for novices, the difference was not statistically significant (t = 1.02, p(1-tailed) = .16). Furthermore, experienced participants performed worse in cell 3 than in cells 1, 2 and 4 (contrary to prediction).[10] These results indicate some support for H2 in cell 2 (when the pattern matches the rate), but not in cell 3, when the pattern does not match the rate.

Hypothesis 3 predicted that novice auditors would perform worse than experienced auditors because they would focus on industry rates rather than patterns. The results in cells 1, 2 and 3 are consistent with this prediction. Notice that in cell 1, focusing on surface features would lead to the correct solution because the industry rates are consistent with the pattern of information, and in this cell there was no statistically significant difference between experienced and novice participants' performance (t = .50, p = .62), as predicted. In cells 2 and 3, focusing on surface features would lead to the incorrect solution because industry rates were inconsistent with patterns. As reported above, novices performed poorly in these two cells in comparison to the more experienced participants, as predicted (but the difference was not statistically significant in cell 3). In cell 4, if novices focused on surface features, they would have gotten the correct answer because industry rates were consistent with the pattern, and would have performed as well as experienced participants. But in this cell novices did not perform as predicted, getting only 48% of the judgments correct, and performing significantly worse than experienced participants (t = 3.3, p(1-tailed) < .001). Thus, the results indicate partial support for H3, in that novices performed as predicted in cells 1, 2, and 3, but not in cell 4. Further discussion of these results is provided in the next section.

DISCUSSION AND CONCLUSION

The purpose of this study was to show that auditors with task-specific knowledge and experience subjects are able to perform better than novices in an ill-structured task setting due to their ability to utilize patterns in data. As predicted, experienced participants outperformed novices with respect to the real estate valuation task. However, it was not the case that the experienced participants *always* relied on the pattern of data (deep features) when making their judgments, and that novices *always* overweighted surface features.

While experienced participants performed well for those cases represented by cells 1, 2 and 4 as noted in Figure 2, they performed relatively poorly for the cases represented by cell 3. This result was not predicted *a priori*, but it is possible that the predictions as depicted in Figure 1 were too simplistic. For example, H2 predicted that the experienced auditors would utilize patterns in their judgments, but there was no mention of how patterns would be used differently when the surface features matched the discount rate per the appraisal versus when they did not match. As noted above, the results regarding the experienced participants in cell 3 indicate that the role of pattern information is different when discount rates given in the appraisal are within the industry average range versus when they are outside of the range. A possible explanation for the relatively poorer performance in cell 3 is that experienced auditors process patterns differently depending on the nature of the signal provided by the surface feature manipulation (i.e. whether the discount rate is within the industry average range or outside of the range). It is reasonable to assume that the industry average rate range may elicit an expectation on the part of the auditor regarding an appropriate value for the property discount rate. Because the industry rate ranges provided in the experiment reflect the ranges of typical properties in specific geographical areas, it is likely that the experienced auditors expect to find the discount rate in the appraisal to be within the industry average range. If they encounter a discount rate outside of the industry average range, this might be considered “unexpected” and cause them to investigate further. Thus, in this study, cells 2 and 4 were cases where the rate would be “unexpected,” and cells 1 and 3 were cases where the rate would be “expected.”

A robust finding in the psychology literature is that unexpected events elicit more causal reasoning than expected events (e.g. Hastie 1984).[11] Much of the research in this area has been performed in the social psychology domain, and these studies have examined the circumstances under which individuals engage in more effortful processing to determine the cause of an event (i.e. to explain another individual’s actions). The general finding from studies such as Pyszczynski and Greenberg (1981) and Wong and Weiner (1981) is that individuals expend additional effort seeking “explanation-relevant” information when unexpected events occur. This tendency is weaker when expected events occur (Hastie 1984). It is possible that in cells 2 and 4 of this study, auditors who encountered an unexpected discount rate were alerted to investigate further (and engage in more causal reasoning) to see whether the pattern of information supported the atypical discount rate. Consequently, experienced participants in cells 2 and 4 utilized the pattern of information and performed well. In contrast, in cells 1 and 3, experienced participants encountered an *expected* discount rate, which might have led them to exert less effort in investigating the pattern of information. In cell 1, because the pattern of information supported the typical rate, this strategy would allow experienced participants to correctly respond that the discount rate was reasonable. However, in cell 3, the pattern of information indicated that the rate was unreasonable. In this case, insufficient attention to the pattern would cause auditors to incorrectly accept the discount rate as being reasonable when it was actually unreasonable.

Note that when auditors in cell 3 were incorrect (38% of the time), they committed a Type II error of incorrect acceptance (by stating that the discount rate was “Reasonable” when the deep features indicated it was “Not Reasonable”.) Type II errors have an adverse impact on audit effectiveness (see Kida 1984; Kinney and Uecker 1982; and Rosman et al. 1996). For example, earnings management represents a case where a company’s operations seem to conform to expectations (generally analysts’ forecasts), when in reality the pattern of company operations indicates otherwise. An interesting extension of this study could examine the differential processing of auditors with respect to expected versus unexpected events, and the relationship of this processing to the generation of Type II errors.

In addition to providing insight into what factors lead to better performance in ill-structured task settings among knowledgeable and experienced auditors, this experiment was also designed to investigate what factors lead to poor performance on the part of novices. In contrast to findings from psychology, the results indicate that the poor performance of novices may not be due to over-reliance on surface features alone. It is possible that heuristics used by novices (i.e. over-reliance on the industry rate information) are only applied in simple cases, where all information is consistent (as in cell 1). In contrast, when the task is more complex, as in cells 2, 3 and 4, novices may resort to random guessing. Alternatively, novices may try to process more deeply or reason causally in these cells, but their lack of knowledge or relevant schemata may not allow them to arrive at the correct judgment. Unfortunately, the design of this study does not allow for more than conjecture as to whether the novices’ incorrect judgments were based on systematic factors or were the result of random guessing; however, future auditing research could examine which type of cues novices focus on and which processes they employ during problem solving.

Due to the limited focus of this study on pattern recognition process only, there were many contextual features that were not incorporated in order to keep the study tractable. Future research could look at the effects of other contextual factors, such as the review process (specifically how reviewers may be influenced by information that is expected versus unexpected, and how this interacts with effort during their reviews), and source credibility (how much the auditor relies on information provided by management, the appraisal, and outside industry reports). It would also be interesting to determine if auditors with more expertise than the experienced participants in this study would be subject to the same effects as that found in cell 3 (or if an increased amount of expertise mitigates the effect). It is interesting to note that the more experienced participants in this study were audit seniors, often the in-charge auditor on their real estate audit engagements, and they indicated that they performed a number of real estate valuations in the past year[12]. It may be that firms should take a closer look at whether auditors at this level have the requisite training and proficiency to perform the task, and/or whether they should be reviewed more closely.

Another interesting extension to this work would be to observe pattern recognition processes aid auditors who do not have a high amount of real estate knowledge, but do have a high degree of experience as auditors. It could be that certain schemata developed in other industry domains would transfer to real estate, resulting in expert-level performance despite a lack of domain-specific knowledge.[13]

Finally, the results of this study may have implications for practicing auditors, particularly with respect to determining how to identify the causes of Type II errors in practice, and for improving the quality of on-the-job training. With respect to training, it may be helpful for more knowledgeable auditors to determine what patterns they are using in the performance of various ill-structured tasks, and try to explain these patterns to novice auditors in order to improve the efficiency and effectiveness of knowledge acquisition.

Endnotes

[1] This task was selected because it represents a departure from previous studies in auditing, and it is relevant in today's auditing environment, particularly with respect to the implementation of reporting requirements under SFAS 121, "Accounting for the Impairment of Long-lived Assets and for Assets to be Disposed Of" (FASB, 1995), and the increase in the presence of REIT investments in companies' investment portfolios (Woolley et al. 1997).

[2] In addressing expertise differences in auditing, the selection of the task to study becomes extremely important. One must select a task for which experts have a comparative advantage. Such tasks generally fall into the category of what cognitive psychologists term "ill-structured" tasks (Newell and Simon 1972; Voss and Post 1988), those for which there is no obvious correct solution, and which require a great deal of knowledge and skill to perform.

[3] Solomon and Shields (1995) note that several past studies of judgment and decision making in accounting contexts have been replications of studies in psychology using accountants as subjects, rather than psychology students. They note that this approach may result in misapplication of the underlying psychology theories, and point out that it is necessary to understand the characteristics of accountants and accounting contexts that make them unique when developing theory and hypotheses in accounting studies.

[4] There has been some lack of consistency in the auditing literature about how to categorize participants as novice or expert. In a review of auditing studies examining the effect of experience on judgments, Colbert (1989) noted that researchers need to examine *task-specific* experience and that for unstructured and semi-structured tasks, experience has a greater effect on performance than for structured tasks (see further discussion regarding the nature of task-specific performance in Gibbins and Jamal 1993 and Bédard 1991, as well as studies by Bonner 1990, and Abdolmohammadi and Wright 1987). In this study, a strictly task-related criterion for expertise was considered appropriate, given the ill-structured nature of the task.

[5] Six of the 13 experienced subjects reported performing 11 to 20 appraisal reviews in the past year, and 7 reported performing more than 20 appraisal reviews.

[6] Several other analyses were performed to evaluate this classification of groups into experienced and novice participants. In order to determine the level of task-specific domain knowledge (as opposed to experience with the task), each subject was asked to provide definitions of six common real estate valuation terms. The mean number of correct definitions was 3.8 for experienced particpants, and 2.1 for novices. This difference was statistically significant ($t=8.12$, $p < .001$), indicating that experienced participants had a higher knowledge of relevant real estate terms than novices. In addition, experience level was significantly positively correlated at the $\alpha = .01$ level with the amount of time spent

on real estate engagements, and how often the participant was consulted regarding valuation issues on other audit engagements, indicating that the experienced group tended to have more experience with real estate audits in general and was more likely to be sought after as a valuation expert by other members of the firm.

[7] The order of the eight cases was counterbalanced in the form of a randomly selected Latin square using the procedure described in Myers and Well (1995, p. 346 - 348). Five 8 X 8 squares were designed, resulting in each case being included in each position only once for each Latin square.

[8] Of the novice auditors, a total of 9 out of 26 (or 35%) of the participants reported having some experience evaluating offices, apartments, or both. By contrast, all 13 experienced subjects had experience evaluating offices and/or apartments with 100% having office experience, and 77% also having apartment experience.

[9] The procedure used was procedure CATMOD in SAS. It was determined that the CATMOD procedure was the most appropriate given the nature of the data in this study. For more information regarding this procedure, please consult the SAS User's Guide (1985).

[10] Paired t-tests were performed to determine whether the mean number correct in cells 1, 2 and 4 were significantly different from one another. The results indicated that none of the differences were significant a $\alpha = .05$, therefore the three means were combined and compared to the mean number correct in cell 3. A paired t-test comparing the mean in cell 3 (62% correct) to the mean in cells 1, 2 and 4 combined (86% correct) indicated that performance in cell 3 was significantly worse than in the other cells ($t = 1.75$, p(1-tailed) = .05).

[11] The idea that information which conforms to expectations is processed differently from information which does not conform to expectations, is related to previous research in auditing regarding the role of confirmatory versus disconfirmatory information in complex judgments (i.e. Kida 1984; Choo 1991) and generation of alternative hypotheses to explain why observed events might *not* have occurred (Heiman 1990).

[12] Indeed in a self-reported measure, 7 out of the 13 "experienced" subjects labeled themselves as "competent," or better with respect to their performance of the valuation task on-the-job.

[13] Using a going concern task, Thibodeau (1997) demonstrated that knowledge of patterns can transfer between different audit domains.

References

Abdolmohammadi, M. and A. Wright. 1987. An examination of the effects of experience and task complexity on audit judgments. The Accounting Review, 62 (1): 1 - 13.

Bédard, J. 1989. Expertise in auditing: Myth or reality? Accounting, Organizations and Society 14: 113 - 131.

__________ and M.T.H. Chi. 1993. Expertise in auditing. Auditing: a Journal of Practice and Theory 12 (supplement): 21 - 45.

__________ . 1991. Expertise and its relation to audit decision quality. Contemporary Audit Research (Fall): 198 - 222.

Bedard, J. and S. Biggs. 1991a. Pattern recognition, hypothesis generation, and auditor performance in an analytical task. The Accounting Review 66: 622 - 642.

__________ . 1991b.The effect of domain-specific experience on evaluation of management representations in analytical procedures. Auditing: A Journal of Practice and Theory (Supplement): 77 - 90.

Bierstaker, J., J. Bedard, and S. Biggs. 1999. The role of problem representation shifts in auditor decision processes in analytical procedures. Auditing: A Journal of Practice and Theory 18 (1): 18-36.

Bonner S. E. and B.L. Lewis. 1990. Determinants of Auditor expertise. Journal of Accounting Research, 28 (supplement): 1 - 28.

__________, 1990. Experience effects in auditing: The role of task-specific knowledge. The Accounting Review, 65 (1): 72 - 92.

Boritz, J. E. 1992. Discussion of expert and novice problem-solving behavior in audit planning. Auditing: A Journal of Practice and Theory, 11 (supplement): 21 - 32.

Chase, W.G. and H.A. Simon. 1973. Perception in chess. Cognitive Psychology 4:55 - 81.

Chi, M.T.H., P. Feltovich, and R. Glaser. 1981. Categorization and representation of physics problems by experts and novices. Cognitive Science 5: 121 - 152.

__________, R. Glaser, and R. Rees. 1982. Expertise in problem solving. In R. Sternberg (ed.) Advances in the Psychology of Human Intelligence (vol. 1), Hillsdale, NJ: Lawrence Erlbaum Associates.

Christ, M. Y. 1993. Evidence on the nature of audit planning problem representations: An examination of auditor free recalls. The Accounting Review, 68 (2): 304 - 322.

Choo, F. 1996. Auditors' knowledge content and judgment performance: A cognitive script approach. Accounting, Organizations, and Society, 21 (4): 339 - 359.

Choo, T. 1991. Research on confirmation bias in accounting: Achievement and prognosis. Working paper. Nanyang Technological University.

Colbert, J. 1989. The effect of experience on auditors' judgments. Journal of Accounting Literature, 8: 137 - 149.

Davis, J.S. and I. Solomon. 1989. Experience, expertise, and expert-performance research in public accounting. Journal of Accounting Literature 8: 150 - 164.

Dreyfus, H.L. and S.E. Dreyfus. 1986. Mind over machine. New York, NY: The Free Press.

Earley, C.E. 1998. Expert versus novice approaches to the solving of ill-structured auditing problems: A case study. Master's thesis. University of Pittsburgh.

Financial Accounting Standards Board. 1995. Accounting for the impairment of long-lived assets and for long-lived assets to be disposed of. Statement of Financial Accounting Concepts No. 121 Stamford, CT: Financial Accounting Standards Board.

Frederick, D.M. 1991. Auditor's representation and retrieval of internal control knowledge. The Accounting Review 66 (2): 240 - 258.

_________, V.B. Heiman-Hoffman, and R. Libby. 1994. The structure of auditors' knowledge of financial statement errors. Auditing: A Journal of Practice and Theory 13: 1 - 21.

Frensch, P.A. and R. Sternberg. 1989. Expertise and intelligent thinking: When is it worse to know better? in R. Sternberg (ed.) Advances in the Psychology of Human Intelligence (vol. 5), Hillsdale, NJ: Lawrence Erlbaum Associates.

Gibbins, M. and K. Jamal. 1993. Problem-centered research and knowledge-based theory in the professional accounting setting. Accounting, Organizations, and Society, 18 (5): 451 - 466.

Gick, M.L. 1986. Problem solving strategies. Educational Psychologist 21: 99 -120.

_________, and K.J. Holyoak. 1980. Analogical problem solving. Cognitive Psychology 12: 306 - 355.

Hastie, R. 1984. Causes and effects of causal attribution. Journal of Personality and Social Psychology 46 (1): 44 - 56.

Hayes, J.R. and H.A. Simon. 1976. The understanding process: Problem isomorphs. Cognitive Psychology, 8: 165 - 190.

Heiman, V. B. 1990. Auditors' assessments of the likelihood of error explanations in analytical review. The Accounting Review 65 (3): 875 - 890.

Hershey, D.A., D.A. Walsh, S.J. Read, and A.S. Chulef. 1990. The effects of expertise on financial problem solving: Evidence for goal-directed, problem-solving scripts. Organizational Behavior and Human Decision Processes, 46: 77 - 101.

Kida, T. 1984. The impact of hypothesis-testing strategies on auditors' use of judgment data. Journal of Accounting Research 22 (1): 332 - 340.

Kinney, W.R. and W.C. Uecker. 1982. Mitigating the consequences of anchoring in auditor judgments. The Accounting Review 57 (1): 55 - 69.

Klein, G. A. 1999. Sources of Power: How People Make Decisions Cambridge, MA: MIT Press.

___________, J. Orsanu, R. Calderwood, and C.E, Zsambok (eds.). 1993. Decision making in action: Models and methods Norwood, NJ: Ablex Publishing Corp.

_________ and R. Calderwood. 1991. Decision models: Some lessons from the field. IEEE Transactions on Systems, Man and Cybernetics 21: 1018 - 1026.

__________, __________, and D. McGregor. 1989. Critical decision method for eliciting knowledge. IEEE Transactions on Systems, Man and Cybernetics 19: 462 - 472.

Libby, R. 1995. The role of knowledge and memory in audit judgment. In R.H. Ashton and A.H. Ashton (eds.) Judgment and Decision Making Research in Accounting and Auditing New York, NY: Cambridge University Press.

__________ and J. Luft. 1993. Determinants of judgment performance in accounting settings: Ability, Knowledge, motivation and environment. Accounting, Organizations and Society 18 (5): 425-450.

Myers, J.L and A. D. Well. 1995. Research Design and Statistical Analysis. Hillsdale, NJ: Lawrence Erlbaum Associates.

Nelson, M.W., R. Libby and S.E. Bonner. 1995. Knowledge structure and the estimation of conditional probabilities in audit planning. The Accounting Review 70 (1): 27 - 47.

Newell, A. and Simon, H.A. 1972. Human Problem Solving Englewood Cliffs, NJ: Prentice-Hall.

Pyszczynski, T.A. and J. Greenberg. 1981. Role of disconfirmed expectatncies in the instigation of attributional processing. Journal of Personality and Social Psychology 40 (1): 31-38.

Rosman, A. J., I. Seol, and S.F. Biggs. 1999.The effect of stage development and financial health on auditor decision behavior in the going-concern task. Auditing: A Journal of Practice and Theory 18 (1) : 37 – 54.

Solomon, I. and M.D. Shields. 1995. Judgment and decision research in auditing. In R.H. Ashton and A.H. Ashton (eds.) Judgment and Decision Making Research in Accounting and Auditing New York, NY: Cambridge University Press.

Thibodeau, J. C. 1997. The development of knowledge structures that transfer across audit domains. Working paper. Bentley College.

Voss, J. F. 1987. Learning and transfer in subject-matter learning: A problem-solving model. International Journal of Educational Research 11 (6): 607 - 622.

__________, T.R. Greene, T. Post, and B.C. Penner. 1983. Problem-solving skill in the social sciences. In G.H. Bower (ed.) The Psychology of Learning nd Motivation: advances in Research Theory New York, NY: Academic Press.

__________, and T. Post. 1988. On the solving of ill-structured problems. in M.T.H. Chi, R. Glaser, and M.J. Farr (eds.) The Nature of Expertise Hillsdale NJ: Lawrence Erlbaum Associates.

Wason, P.C. and P.N. Johnson-Laird. 1972. Psychology of Reasoning: Structure and Content Cambridge, MA: Harvard University Press.

Woolley, S., K. Morris, R.A. Melcher, and S. A. Forest. 1997. The new world of real estate. Business Week (September 22): 78 - 87.

Wong, P. T. and B. Weiner. 1981. When people ask “why” questions, and the heuristics of attributional search. Journal of Personality and Social Psychology 40 (4): 650-663.

Table 1 Chi-Square Values for Response Frequencies

Source	df	Chi-Square	P-value
Intercept	1	193.92	.000
Experience Level	1	22.18	.000 [a]
Deep features	1	6.81	.033
Surface features	1	2.40	.302
Deep * Surface	1	8.51	.014 [b]
Experience * Deep	1	4.53	.104
Experience * Surface	1	3.29	.193 [b]
Experience * Deep * Surface	1	0.09	.954

[a] Test of H1.

[b] Test of H2 and H3 combined.

Figure 1 Experimental Predictions

		SURFACE FEATURES	
		RATE MATCHES INDUSTRY AVG.	RATE DOES NOT MATCH IND. AVG.
DEEP FEATURES	PATTERN MATCHES	Cell 1 EQUAL NUMBER OF CORRECT JUDGMENTS (Both groups perform well) *Correct Answer: Reasonable*	Cell 2 EXPERIENCED AUDITORS PERFORM BETTER *Correct Answer: Reasonable*
	PATTERN DOES NOT MATCH	Cell 3 EXPERIENCED AUDITORS PERFORM BETTER *Correct Answer: Not Reasonable*	Cell 4 EQUAL NUMBER OF CORRECT JUDGMENTS (Both groups perform well) *Correct Answer: Not Reasonable*

Figure 2 Experimental Results

SURFACE FEATURES

DEEP FEATURES	RATE MATCHES INDUSTRY AVG.	RATE DOES NOT MATCH IND. AVG.
PATTERN MATCHES	Cell 1 Experienced = 88% correct Novice = 85% correct *Prediction supported.* *Correct Answer: Reasonable*	Cell 2 Experienced = 81% correct Novice = 54% correct *Prediction supported.* *Correct Answer: Reasonable*
PATTERN DOES NOT MATCH	Cell 3 Experienced = 62% correct Novice = 50% correct *Prediction not supported.* *Correct Answer: Not Reasonable*	Cell 4 Experienced = 88% correct Novice = 48% correct *Prediction not supported.* *Correct Answer: Not Reasonable*

APPENDIX

Experimental Instrument - Sample Case

<u>BACKGROUND INFORMATION</u>			
PROPERTY NAME:	Union Place	**TOTAL SQUARE FEET OR UNITS**	349,179 SF
APPRAISED VALUE:	$50, 100,000	**YEAR BUILT:**	1987
LOCATION:	San Francisco (Central Business District)	**RISK FREE RATE OF RETURN**	1997 9% 1998 8% 1999 11%
PROPERTY TYPE:	Office (Class A)		
<u>LEASING INFORMATION:</u>			
OCCUPANCY RATE:	95%	**LEASE EXPIRATIONS: (% of total square footage)**	1997 17% 1998 38% 1999 7%
<u>MARKET INFORMATION</u>			
COMPETITION:	Average vacancy rate in CBD is 10 - 12%. Average market rent per square foot of similar buildings in CBD is $15 to $26. Estimated demand of 800,600 additional square feet per year is expected, with the total vacant inventory in the area to be absorbed in 1.7 years.	**OTHER FACTORS:**	Expected volatility in the market in 1997 as office space leased by companies whose original space was damaged in the 1989 Loma Prieta earthquake is vacated by those tenants, who will move back into their original space.
<u>ASSUMPTIONS</u>			
DISCOUNT RATE: (per appraisal)	11.25%	**DISCOUNT RATE: (per industry reports)**	11.0% to 11.7%

<u>Instructions</u>

Please evaluate the **discount rate per the appraisal** in light of the information about the property given above, and answer the following:

1. The discount rate per the appraisal is: **Reasonable** **Not Reasonable**

2. If you answered “Not Reasonable”, was the discount rate (circle one): **Too High** **Too Low**

Please provide a reasonable range for the discount rate: ____________________________