Department of Accountancy
University of Illinois at Urbana-Champaign presents

NINTH SYMPOSIUM
ON AUDITING RESEARCH

October 4-6, 1990

Research Papers

Co-sponsored by the Auditing Section of the
American Accounting Association

The financial support of the
KPMG Peat Marwick Foundation
is gratefully acknowledged
Towards An Explanation of Auditor Failure
to Modify the Audit Opinions of Bankrupt Companies

by

James C. McKeown*
Distinguished Professor of Accounting
Penn State University

Jane F. Mutchler
Associate Professor
University of Arizona

and

William Hopwood
Arthur Andersen Alumni Professor of Accounting
University of Houston at University Park

September 7, 1990

The support of the KPMG Peat Marwick Foundation's Research Opportunities in Auditing Program is gratefully acknowledged.

*The names are listed in random order
Towards An Explanation of Auditor Failure
to Modify the Audit Opinions of Bankrupt Companies

Abstract

This research examines auditors' failures to modify their opinions for companies that go bankrupt. First, a "hidden fraud" variable is added to a bankruptcy estimation model and is significant when the model is applied to nonstressed companies, but not when applied to stressed companies. Second, auditors' opinion decisions for stressed companies are analyzed by testing four hypotheses: auditors are less likely to qualify the opinions of (1) bankrupt companies with ambiguous probabilities of bankruptcy, (2) larger bankrupt companies, and (3) companies for which there is a shorter time period between their fiscal year-ends and their audit opinion dates. Each of these hypotheses is supported. A fourth hypothesis, that auditors do not qualify bankrupt companies that have hidden fraud, is not supported; but the likelihood of undiscovered fraud is much higher for the nonstressed companies, suggesting that additional audit effort may produce gains. The inverse relationship between client size and the going-concern qualification, after controlling for the relationship between size and bankruptcy, indicates that pressures related to client size may influence the opinion decision.
Towards An Explanation of Auditor Failure
to Modify the Audit Opinions of Bankrupt Companies

Cases in which a company has gone bankrupt without having received a qualified opinion are of particular interest to the public and investigative bodies. Previous studies [Altman and McGough, 1974; Altman, 1982; Menon and Schwartz, 1986; and Hopwood, McKeown and Mutchler, 1989] have found less than 50 percent of bankrupt companies receiving going-concern opinions for the last financial statements issued prior to the bankruptcy. There has been little evidence from previous studies indicating whether this low rate is a result of auditor failure or whether it simply reflects a situation for which the auditor is asked to do the impossible. This suggests a need for theories to explain why auditors often do not qualify the opinions of failing companies. The purpose of this research, therefore, is to develop and test formal models that explain, at least in part, why auditors often fail to qualify the opinions of companies that go bankrupt.¹

On one hand, it may be that fraudulent financial statements mislead auditors by producing a false appearance of financial viability. Therefore, to the extent that auditors fail to detect fraud and use incorrectly reported financial statement numbers in analyses such as those related to typical ratio-based bankruptcy models, they may incorrectly assess the bankruptcy potential for a given company.² Accordingly, this research develops a model to test the incremental explanatory power—relative to the estimation of a firm’s probability of bankruptcy—of a hidden-fraud variable that indicates fraud reported after, but not before, the audit opinion date.³ A necessary condition for this variable to improve the audit-qualification decision is that it must provide information that is incremental to the information that the auditor already has available and uses. If this condition is met, it is likely that auditors can, by searching harder for fraud, more often qualify the opinions of companies that go bankrupt. This also is an open question,
since no previous studies have investigated the usefulness of fraud variables, even though many
bankruptcy models have been tested and proposed as useful to auditors [Zavgren, 1983; Jones,
1987].

On the other hand, however, there could be some factors in the opinion decision process
itself that causes auditors not to qualify the audit opinions of companies that go bankrupt. This
research, therefore, also develops a model of auditors' qualification decisions for companies that
go bankrupt. This model incorporates financial stress, auditor factors, and client factors, and
in terms of these variables, it explains the differences between those bankrupt companies that
receive the going-concern qualification and those bankrupt companies that do not receive the
going-concern qualification. Since only bankrupt companies are considered, the model explains
why auditors qualify the opinions of some bankrupt companies and not others. In addition to
auditor and client variables, the model also includes the fraud indicator variable, which is used
to test for the actual effect of fraud on auditors correctly qualifying the opinions of companies
that go bankrupt. This is necessary, since in the bankruptcy model the fraud variable measures
only the potential, not the actual, effect of fraud on the chance that auditors will correctly
qualify the opinions of companies that go bankrupt.

The next section of this paper develops the hypotheses, describes the method used,
develops the models, and describes the sample. The third section reports the results. The final
section provides a summary, conclusions, and suggestions for further research.
HYPOTHESIS DEVELOPMENT AND RESEARCH METHOD

This section is divided into four parts. The first part deals with general research design considerations in developing the models. The second part develops the hypothesis and the model that tests the importance of a hidden-fraud indicator variable in explaining financial stress and bankruptcy. The third section develops the hypotheses and the model that explains auditors' opinion decisions for bankrupt companies. The fourth section describes the sample.

General Considerations

The first model considers the usefulness of a hidden-fraud indicator variable in the estimation of bankruptcy probabilities. The second model explains the opinion decision for bankrupt companies in terms of financial-stress scores, a hidden-fraud indicator variable and other variables indicative of auditor and client factors. Since both models are developed to provide explanations and tests of hypotheses regarding auditors' opinion decisions, it is necessary to develop them in a context broader than the standard ratio-based financial stress estimation model, which, as discussed here, does not adequately deal with many practical situations. Such a context for the bankruptcy prediction models is provided by Argenti [1976], who performed an in-depth analysis and identified three types of corporate failures. The first type includes fledgling companies who never move beyond a financially-stressed state and fail before they become established. The second type includes companies whose failure is precipitated by a gradual slide into insolvency and portended by signs of financial stress in the financial ratios. The third type includes companies whose failure is sudden but with no apparent signs of financial distress.
In the context of the audit opinion decision, the Argenti results imply that the auditor faces two fundamentally different situations, one in which financial stress is important and another in which it is not. Therefore, both the audit-opinion and extended stress estimation models developed below are applied separately to stressed versus nonstressed companies, and this dichotomization is assumed in the discussion that follows. Given the suggestions in Kida [1980] and Mutchler [1985] that auditors will not consider issuing the going-concern opinion unless the company is exhibiting financial stress of some sort, the dichotomization into stressed and nonstressed companies seems to reflect auditors' decision behavior. To the extent that the economic circumstances of the stressed and nonstressed groups do differ, it should enhance statistical estimation of the models by increasing their within-group sample homogeneity and, consequently, their explanatory power.

Factors in the Stress Estimation Process that Explain Auditor Failure to Modify the Opinions of Bankrupt Companies

There is little information in the literature on the degree to which bankruptcy and fraud are related. Palmrose [1987] in her study on litigation found that only 21 percent of a sample of bankrupt companies were involved in auditor litigation. However, 56 percent of the bankrupt companies in litigation involved management fraud related to misstated financial statements. Although fraud may sometimes be discovered before the date of the last audit report issued before the bankruptcy, it often is not discovered until after the bankruptcy has been declared. In before-bankruptcy-discovery cases, one would expect that the misstated financial statements would be corrected.
In post-bankruptcy-discovery cases, however, it may be difficult, because of the presence of hidden fraud, for the auditor to accurately assess the failure potential through the use of traditional ratio-based bankruptcy models. This leads to the first hypothesis:

\[ H_1: \text{Fraud reported after the audit report date adds explanatory power, beyond that contained in traditional financial ratios and size, to a bankruptcy-probability estimation model.} \]

This hypothesis may be supported, for example, if information about post-audit-report fraud affects the contamination of the reported financial information or if this fraud contributes to the bankruptcy itself.

This hypothesis is tested in an extended bankruptcy-probability estimation model. This model incorporates a hidden-fraud indicator variable in addition to traditional financial ratios and a control variable for size. The model follows the general form:

\[ \text{BR} = f(\text{traditional ratios, fraud indicator variable, size}), \]

where BR is an indicator variable for bankruptcy and the traditional ratios are the six financial ratios that previous research [Hopwood, McKeown and Mutchler, 1989] has identified as being a good set of predictors of bankruptcy. These ratios are net income/total assets, current assets/total assets, current assets/current liabilities, cash/total assets, current assets/sales and long-term debt/total assets. Size is the log of sales.\(^4\) The fraud indicator variable (FRAUDA, for FRAUD first reported After the audit report date) was defined in terms of news items published in the Wall Street Journal Index during the five-year period starting with January 1 of the second calendar year prior to the bankruptcy date and running through the end of the second calendar year following the bankruptcy date.\(^5\) All news items indicating the presence of fraud were noted and classified into one of nine categories.\(^6\) In measuring FRAUDA, an
observation for a company was first coded with a 1 if that company had one or more occurrences in a given category after, but not before, the audit report date. The binary variables were then summed across the nine categories. These variables were tested using logit analysis on both the nonstressed and stressed samples.

Factors in the Opinion Decision Process That Explain Auditor Failure to Modify the Opinions of Bankrupt Companies

This section develops an audit-opinion model to explain why many companies go bankrupt without first receiving a going-concern qualification. The model is predicated on possibly overlapping hypotheses that explain why some bankrupt companies receive going-concern qualifications and others do not. Therefore, it and the hypotheses that follow apply strictly to bankrupt companies. The first hypothesis is that auditors simply fail to qualify their opinions for companies with lower stress scores. Under this hypothesis the opinions of many bankrupt companies may not be qualified because the companies are on the borderline between survival and failure. This hypothesis is therefore subsequently referred to as the "Borderline-Failure Hypothesis":

\[ H_2: \text{Ceteris Paribus, auditors are more likely to qualify client companies with higher financial stress.} \]

In testing this hypothesis, the probability of bankruptcy, \( P(B) \), was used as a measure of financial stress. \( P(B) \) was generated from a logit model containing the financial ratio and size variables as described above.

The variable \( P(B) \) summarizes the company's financial stress and provides a very straightforward test of the borderline hypothesis. The use of the individual ratios in the opinion-prediction model may induce confounding effects. More specifically, the ratios may be related
to the opinion decision for reasons unrelated to a company's financial stress level. Thus, relative to the individual financial ratio and size variables themselves, the single measure P(B) has the advantage of being objective, a priori, and uncontaminated by the auditor's judgment model. Moreover its estimation is invariant with respect to varying cost weights and priors. If the Borderline Hypothesis \( H_2 \) is true, then P(B) should be positively related to qualification. Specifically, companies with a high P(B) should receive the qualification while those with a lower P(B), i.e., those close to the borderline, should not.

As stated above, auditors may find it difficult, if not impossible, to assess potential bankruptcy if the financial statements are misstated due to financial fraud. The next hypothesis is that the causes of some failures do not become evident until after the audit report date because of fraud not reported until that time. This hypothesis is subsequently referred to as the "Hidden-Fraud Hypothesis":

\[ H_3: \text{ Ceteris paribus, auditors are less likely to qualify companies which report fraud for the first time after the audit report date.} \]

The fourth hypothesis is based on previous research [Mutchler, 1986] which suggests that auditors will more often issue going-concern opinions to smaller companies. It may be that the size of a company will help explain the failure to issue the going-concern opinion. The driving force behind the opinion-size relationship is unclear; previous research, however, provides some direction. DeAngelo [1981] suggests that as the size of a particular client increases, relative to the total client base, the auditor will find it harder to maintain independence. It may be that auditors with clients large relative to the total client base find it difficult to issue the going-concern opinion because, for example, of client or audit fee pressures. This suggestion is given credence by the results of Palmrose [1986], who found that client size is a major explanatory
variable in the pricing of audit services. This hypothesis is subsequently referred to as the "Company-Size Hypothesis":

\[ H_4: \text{Ceteris Paribus, auditors are more likely to issue the going-concern opinions to smaller companies.} \]

To test for this client-size effect, the natural log of sales is included as an independent variable. Since size also is included in the logit model used to estimate the probability of bankruptcy, any relationship between size and bankruptcy should already be impounded in the probability of bankruptcy. Thus, if the going-concern opinion is related to size in the current model, then in their opinion decision auditors consider as a factor information that is related to size, but is independent of the relationship between size and bankruptcy.

The last hypothesis is that there is an audit-delay factor that is inversely related to the failure to qualify. Ashton, Willingham and Elliott [1987], Dodd et al. [1984], and Elliott [1982] all found that companies receiving qualified opinions report later than companies receiving unqualified opinions. It is expected that, on average, companies that receive going-concern opinions have a longer audit delay than companies that do not receive the qualification. This may be true because auditors may be required to spend more time on troubled companies, especially when they exhibit signs of not being able to meet their obligations as they become due. It may also be true because auditors spend more time dealing with the management of a company when a going-concern opinion is issued. Conversely, given that the audit delay is shorter for unqualified bankrupt companies, it could be construed as evidence of under auditing.7

This hypothesis is subsequently referred to as the "Audit-Delay Hypothesis."

\[ H_5: \text{Ceteris paribus, the longer the time period of the audit, the greater the likelihood of the auditor issuing the going-concern opinion.} \]
The ARLAG variable is used to test the audit-delay hypothesis. This variable represents the length of time between the fiscal year-end and the date of the audit opinion.

Any of these hypotheses, if true, provide a reasonable explanation for auditors' failure to qualify the opinions of companies that go bankrupt: the companies are on the borderline, there is hidden fraud, the auditors are influenced by companies' sizes, or the opinions are less likely to be modified for companies which have shorter periods between their fiscal year-ends and their audit-report dates.

In summary, these hypotheses suggest a model that contains the four groups of explanatory variables described above: a measure of financial stress as estimated by a bankruptcy model; FRAUDA, an indicator of fraud first reported after the audit report date; a client size measure, the natural log of sales; and ARLAG, a measure of audit delay. Specifically, the model is of the general form:

\[ GC_{br} = f(\text{financial-stress variable, hidden fraud, client size, audit delay}), \]

where \( GC_{br} \) is an indicator variable for going-concern opinions for bankrupt companies and the explanatory variables are as defined above. Note that the model is defined for only bankrupt companies since its purpose is limited to explaining why some bankrupt companies receive going-concern qualifications and other bankrupt companies do not. These variables were tested using logit analysis.
Sample

To identify a sample of bankrupt companies for the study, the *Wall Street Journal Index* was searched for companies listed on either the New York or American stock exchanges and appearing under the heading "Bankruptcies" for the years 1974 through 1985. One-hundred and thirty-four companies were found which had full financial statement data available on the COMPSTAT Annual Industrial file. A nonbankrupt sample of 160 control companies was randomly selected from the same time period.

Data on audit opinions were gathered from 10-K reports available in various university libraries, from copies of audit opinions received from Disclosure Inc, and from the NAARS database. Opinions were classified as going-concern opinions in cases in which the auditor either disclaimed an opinion or expressed doubt about the company’s ability to survive as a going concern. Opinions also were classified as going-concern opinions if the auditor expressed doubt about either the company’s ability to finance future operations or, regardless of the company’s ability to continue in existence, its "recoverability and classification of assets and the amounts and classification of liabilities."

The companies in the sample were partitioned into several categories, which included all possible combinations of companies that were (were not) experiencing financial stress, had (had not) filed for bankruptcy, and had (had not) received the going-concern audit opinion. Measures of stress based on previous research [Kida, 1980; Mutchler, 1985] were used, and a company was classified in the stressed category if it exhibited at least one of the following financial distress signals: (1) negative working capital in the current year, (2) a loss from operations in any of the last three years prior to bankruptcy, (3) a retained earnings deficit in Y-3 (where Y-1 is the last financial statement date preceding bankruptcy), or (4) a bottom-line loss in any of the
last three years before bankruptcy. Any company that did not meet at least one of the criteria was considered to be a nonstressed company.

EMPIRICAL RESULTS

Incidence of Stress and Qualifications in the Bankruptcy and Control Samples

Table 1 presents an analysis of the sample composition. A few items are noteworthy. First, as has been found in previous research, less than 50 percent (i.e., 46 percent) of the stressed bankrupt companies received a going-concern qualification. In addition, it is noteworthy that 12 percent of the bankrupt companies had no signs of financial distress as defined in this paper and that none of the nonstressed companies received the going-concern qualification, giving credence to previous research suggesting that auditors will only consider companies with some sign of financial stress as candidates for the going-concern qualification [Kida, 1980; Mutchler, 1985]. Finally, note that only four of the stressed nonbankrupt companies received the going-concern qualification. This is indicative of a 5.0 percent Type I error rate (qualifying the opinion of a nonbankrupt company) versus a 54 percent Type II error rate (not qualifying the opinion of a bankrupt company).\textsuperscript{11} This goes against the conservative stance expected from auditors and suggests that more work needs to be done to understand auditors' going-concern opinion decisions.

[INSERT Table 1 ABOUT HERE]
Descriptive Statistics on Fraud

Table 2 presents sample statistics on occurrence of fraud within the stressed and nonstressed groups. Note that the nonstressed-nonbankrupt-nonqualified group has the highest incidence of fraud reported before the audit report date, while the nonstressed-bankrupt-nonqualified group has the highest incidence of fraud reported after the audit report date. Also note that the qualified group of stressed-bankrupt companies has a higher incidence of hidden fraud than the unqualified group of stressed-bankrupt companies. Taken together these facts suggest that the auditor may actually have some audit procedures in place for discovering fraud but that these are used or useful for only the stressed companies. Alternatively, the presumably more extensive audit procedures applied to the stressed companies may give the auditor other information which leads to the going-concern qualification even if the fraud is not discovered.

[INSERT Table 2 ABOUT HERE]

Tests of Factors in the Stress Estimation Process That Explain Auditor Failure to Modify the Opinions of Bankrupt Companies

Stressed Bankrupt Companies

Table 3 presents descriptive statistics for the stressed-bankrupt companies compared to the stressed-nonbankrupt companies. The differences in the sample means are generally in the direction predicted by the hypotheses: the bankrupt companies are more likely to be smaller, have bigger losses, greater debt and a lower current ratio. They also are more likely to have higher incidence of hidden fraud.

[INSERT Table 3 ABOUT HERE]
Table 4 presents the results of the logit analysis on the stressed-bankrupt and stressed-nonbankrupt companies. Four of the ratios are individually significant in distinguishing between the bankrupt and nonbankrupt companies. The significant ratios include net income/total assets, the current ratio, current assets/total assets, and long-term debt/total assets. The FRAUDA variable is in the hypothesized direction but is not significant. Thus, the hypothesis that hidden fraud adds explanatory power beyond the traditional financial ratios to a bankruptcy model is not supported for the stressed group of companies. The coefficient of the size variable is in the opposite direction of that hypothesized, indicating no evidence that size is inversely related to bankruptcy for the stressed group.

[INSERT TABLE 4 ABOUT HERE]

Nonstressed Bankrupt Companies

Table 5 reports descriptive statistics for the nonstressed-bankrupt group compared to the nonstressed-nonbankrupt group. The differences in the sample means are generally in the hypothesized direction. The nonstressed-bankrupt group has lower earnings, more debt, and is smaller than the nonstressed-nonbankrupt group. Interestingly enough, the bankrupt group has a higher current ratio. As predicted by hypothesis 1, the bankrupt group has a much higher incidence of hidden fraud than the nonbankrupt group.

[INSERT Table 5 ABOUT HERE]

Table 6 shows the results of the logit analysis of the nonstressed-bankrupt group versus the nonstressed-nonbankrupt group. Only the long-term debt/total assets ratio is individually significant in distinguishing between the two groups. Size is inversely related to bankruptcy for the sample nonstressed group. It is, however, insignificant in distinguishing between the bankrupt- and nonbankrupt-nonstressed companies. Auditors may be interested in the fact that
the hidden-fraud variable is significant \((p = .04)\) in distinguishing between the nonstressed-bankrupt and nonstressed-nonbankrupt observations. This implies that discovery of the hidden fraud may improve the audit results.\(^{12}\)

[INSERT Table 6 ABOUT HERE]

Tests of Factors in the Opinion Decision Process that Explain Auditor Failure to Modify the Opinions of Bankrupt Companies

Table 7 presents the sample means of the variables for the stressed-bankrupt-qualified and stressed-bankrupt-nonqualified groups. The differences in the sample means of the probability of bankruptcy, size, and ARLAG (days between statement date and audit report date) variables are in the direction predicted by hypotheses 2, 4, and 5, respectively. That is, the qualified-stressed-bankrupt companies are likely to have a higher probability of bankruptcy, are likely to be smaller, and are likely to have a longer period of time between the fiscal year-end and the audit report date than are the companies in the nonqualified-stressed-bankrupt group. Note that the qualified group is more likely to experience hidden fraud.

[INSERT Table 7 ABOUT HERE]

Table 8 presents the results of the logit analysis. The probability of bankruptcy, audit delay (ARLAG), and size variables are significant at the .05 level. There is strong support for the Borderline-Failure Hypothesis, the Audit-Delay Hypothesis and the Client-Size Hypothesis. Auditors are less likely to qualify the opinions of companies whose stress scores indicate a stress level which presents an ambiguous probability of bankruptcy. In addition, bankrupt companies that do not receive going-concern audit opinions are more likely to have shorter time periods between their fiscal year-ends and the dates their opinions are signed and are more likely to be larger companies than companies that receive going-concern opinions.
It is particularly important to note that, as stated earlier, the probability of bankruptcy measure incorporates the relationship between bankruptcy and size. Thus, the fact that the size variable is significant indicates that size is a factor in the auditor’s opinion decision above and beyond the fact that there may be a relationship between bankruptcy and size. These results provide evidence consistent with the DeAngelo [1981] work, discussed earlier, which indicates that the larger the client is relative to the total client base, the more likely it is that auditor independence will be compromised.

The coefficient of the FRAUDA variable is not significant, which is not surprising given the above results which provide no support for the hypothesis that FRAUDA has incremental explanatory power (over financial ratios and size) in estimating bankruptcy probabilities. Furthermore, as was discussed above, in stressed companies fraud is much more likely to be discovered before rather than after the audit-report date. In any event, the data do not support the hypothesis that the auditors' failure to discover hidden fraud accounts for their failure to render qualified opinions for stressed companies that go bankrupt.

As can be seen from Table 1, there were only 16 bankrupt companies that fell in the nonstressed category, and none of those received the going-concern qualification. Thus, it is impossible, or at least trivial, to estimate a logit model for the nonstressed-bankrupt-qualified companies versus the nonstressed-bankrupt-unqualified companies. The model is simple and has a 100 percent rate of classification accuracy: all nonstressed companies receive an unqualified opinion. Therefore, auditors appear to ignore all of the explanatory variables in the above bankruptcy model, despite their association with bankruptcy. Also, there is no evidence that auditors are more likely to qualify the opinions of the nonstressed companies’ financial
statements that contain hidden fraud. This implies that the auditors may gain by extending the audit to discover hidden fraud. Descriptive statistics on the nonstressed bankrupt companies are presented in Table 5.

SUMMARY, DISCUSSION AND DIRECTIONS FOR FUTURE RESEARCH

This research investigated reasons for auditors’ failures to qualify the opinions of companies which go bankrupt, a problem that occurs in over 50 percent of the cases of bankruptcy. Two models were investigated for separate samples of stressed and nonstressed companies. The first model, which classifies companies as bankrupt and nonbankrupt, contains financial ratios, company size, and hidden fraud (fraud reported after the audit-report date) as explanatory variables. In this model hidden fraud was found to have incremental explanatory power for the nonstressed companies but not for stressed companies. This difference in incremental explanatory power was reflective of the differences between the two samples in the relative incidence of hidden versus nonhidden fraud. In the stressed sample, fraud was no more likely to be reported after the audit opinion date than before. However, in the nonstressed sample, fraud was over four times more likely to be reported after the audit opinion date than before. The incremental explanatory power of hidden fraud in the nonstressed sample is consistent with the notion that a more extensive auditing process, one designed to discover fraud, will make it possible for auditors to qualify their opinions more often for those nonstressed companies that later go bankrupt.

The second model focused on factors in the opinion decision having the potential to explain auditors’ failure to modify the audit opinions of companies that go bankrupt. In terms of these factors, several hypotheses were tested and found to be supported by the data: the
Borderline-Failure Hypothesis, the Client-Size Hypothesis, and the Audit-Delay Hypothesis. Companies that do not receive qualified opinions are more likely to have ambiguous bankruptcy probabilities, are more likely to be larger, and are more likely to have shorter time periods between their fiscal year-ends and audit report dates than those companies that do receive going-concern qualifications. The fact that the Borderline Hypothesis is supported suggests that the auditor may be setting the probability-of-bankruptcy qualification threshold too high. That is, the auditor may need to modify the opinions of companies that have lower bankruptcy probabilities than those that they now modify. The empirical support for the Client-Size Hypothesis is noteworthy, given the fact that the relationship between bankruptcy and size was controlled. This raises the issue as to whether auditors' independence is influenced by the size of the client, e.g., that the auditor may be bowing to client pressure or, even in the absence of client pressure, that the auditor does not want to face loss of the audit fee. The fact that the Audit-Delay Hypothesis is supported suggests that there are other auditor or client factors that influence the opinion decision. It may be, for example, that some auditors do not have adequate resources to do an extensive audit of a troubled company, and, thus, there may be some underauditing. Client pressures for lower fees or a faster audit also could lead to underauditing. It also may be that different industries require different levels of auditing or that different auditing firms will, ceteris paribus, have differing audit levels.\footnote{13}

The importance of hidden fraud also was considered in the audit-opinion model. This variable does not appear to explain the auditors' failure to qualify the opinions of stressed companies that go bankrupt. Moreover, the lack of incremental significance of this variable in the bankruptcy model means that for stressed companies there is no evidence the auditor's opinion as an indicator of bankruptcy would be improved by more extensive auditing procedures
that lead to the discovery or prediction of fraud. The opposite is true, however, for nonstressed companies. For those companies, there were no cases observed in which auditors qualified their opinions, and this resulted in their failing to qualify 16 companies in the sample. But nonstressed companies with hidden fraud are more likely to go bankrupt. This implies that auditors stand to gain by focusing on the discovery of hidden fraud in nonstressed companies.

There are many avenues of research still open in the area of the auditor’s going-concern opinion decision. The fact that the Client-Size and Audit-Delay Hypotheses are supported indicates that there are other client or auditor factors (or client-auditor interaction factors) that may explain the opinion decision. This research has considered only New York and American Stock Exchange companies; therefore, future research can determine whether the relationships found in this research exist in the over-the-counter companies. Furthermore, although this research has laid the groundwork and has focused on financial and nonfinancial characteristics of the company, other variables may be important. For example, audit-firm, market, industry, and other nonfinancial variables may increase the explanatory power of the model.

The stress-estimation models used in this research can be extended to include other nonfinancial events. The division of the bankrupt companies into stressed and nonstressed categories indicates the need for a new approach and new models for stress estimation, particularly in relation to the auditor’s going-concern opinion decision.


<table>
<thead>
<tr>
<th>Sample Composition</th>
<th>Stressed</th>
<th>Nonstressed</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bankrupt Qualified</td>
<td>54 (46%)</td>
<td>0</td>
<td>54 (40%)</td>
</tr>
<tr>
<td>Bankrupt Nonqualified</td>
<td>64 (54%)</td>
<td>16</td>
<td>80 (60%)</td>
</tr>
<tr>
<td>Total Bankrupt</td>
<td>118 (88%)</td>
<td>16 (12%)</td>
<td>134</td>
</tr>
<tr>
<td>Nonbankrupt Qualified</td>
<td>4 (5%)</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Nonbankrupt Nonqualified</td>
<td>76 (95%)</td>
<td>80</td>
<td>156</td>
</tr>
<tr>
<td>Total Nonbankrupt</td>
<td>80</td>
<td>80</td>
<td>160</td>
</tr>
<tr>
<td>Totals</td>
<td>198</td>
<td>96</td>
<td>294</td>
</tr>
</tbody>
</table>
Table 2
Average Incidence of Fraud News Events

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>FRAUDB</th>
<th>FRAUDA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stressed</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bankrupt</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qualified</td>
<td>54</td>
<td>.204</td>
<td>.204</td>
</tr>
<tr>
<td>Nonqualified</td>
<td>64</td>
<td>.219</td>
<td>.125</td>
</tr>
<tr>
<td>Nonbankrupt</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qualified</td>
<td>4</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>Nonqualified</td>
<td>76</td>
<td>.132</td>
<td>.145</td>
</tr>
<tr>
<td><strong>Nonstressed</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bankrupt</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qualified</td>
<td>0</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Nonqualified</td>
<td>16</td>
<td>.188</td>
<td>.875</td>
</tr>
<tr>
<td>Nonbankrupt</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qualified</td>
<td>0</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Nonqualified</td>
<td>80</td>
<td>.288</td>
<td>.175</td>
</tr>
</tbody>
</table>

FRAUDB = number of fraud categories with news items before the audit report date

FRAUDA = number of fraud categories with news items only after the audit report date

N/A = no observations in this cell
Table 3
Mean Variable Values for Stressed Bankrupt (SB)
Versus Stressed Nonbankrupt (SNB)

<table>
<thead>
<tr>
<th>Variable</th>
<th>SNB (N=118)</th>
<th>SB (N=80)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Net Income/Total Assets</td>
<td>-.0002</td>
<td>-.1126</td>
</tr>
<tr>
<td>Current Assets/Sales</td>
<td>.4272</td>
<td>.4653</td>
</tr>
<tr>
<td>Current Assets/Current Liabilities</td>
<td>1.6638</td>
<td>1.2917</td>
</tr>
<tr>
<td>Current Assets/Total Assets</td>
<td>.4487</td>
<td>.5433</td>
</tr>
<tr>
<td>Cash/Total Assets</td>
<td>.0528</td>
<td>.0405</td>
</tr>
<tr>
<td>Long-Term Debt/Total Assets</td>
<td>.2682</td>
<td>.2904</td>
</tr>
<tr>
<td>Natural Log of Sales</td>
<td>4.6418</td>
<td>4.4774</td>
</tr>
<tr>
<td>Fraud reported after audit report date</td>
<td>.1375</td>
<td>.1610</td>
</tr>
</tbody>
</table>
Table 4
Logit Results For Stressed Bankrupt (SB) Versus Stressed Nonbankrupt (SNB)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>$\chi^2$</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-2.295</td>
<td>5.46</td>
<td>.019</td>
</tr>
<tr>
<td>Net Income/Total Assets</td>
<td>-14.737</td>
<td>25.74</td>
<td>.000*</td>
</tr>
<tr>
<td>Current Assets/Sales</td>
<td>.130</td>
<td>.06</td>
<td>.813</td>
</tr>
<tr>
<td>Current Assets/Current Liabilities</td>
<td>-1.678</td>
<td>22.37</td>
<td>.000*</td>
</tr>
<tr>
<td>Current Assets/Total Assets</td>
<td>5.826</td>
<td>23.82</td>
<td>.000*</td>
</tr>
<tr>
<td>Cash/Total Assets</td>
<td>-5.611</td>
<td>1.58</td>
<td>.209</td>
</tr>
<tr>
<td>Long-Term Debt/Total Assets</td>
<td>3.099</td>
<td>6.12</td>
<td>.013*</td>
</tr>
<tr>
<td>Natural Log of Sales</td>
<td>.206</td>
<td>2.18</td>
<td>.930</td>
</tr>
<tr>
<td>Fraud reported after audit report date</td>
<td>.213</td>
<td>.25</td>
<td>.310</td>
</tr>
</tbody>
</table>

Full Model $R^2 = .303$

SB = Stressed Bankrupt (coded as 1)
SNB = Stressed Nonbankrupt (coded as 0)

P values are for two-tailed hypotheses for the intercept and ratios, but one-tailed for sales and fraud. Hypotheses rejected at an alpha level of .05 are indicated by *. 

Table 5
Mean Variable Values for Nonstressed Bankrupt (NSB) Versus Nonstressed Nonbankrupt (NSNB)

<table>
<thead>
<tr>
<th>Variable</th>
<th>NSNB (N=80)</th>
<th>NSB (N=16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Net Income/Total Assets</td>
<td>.0673</td>
<td>.0395</td>
</tr>
<tr>
<td>Current Assets/Sales</td>
<td>.4291</td>
<td>.6619</td>
</tr>
<tr>
<td>Current Assets/Current Liabilities</td>
<td>2.5888</td>
<td>3.0411</td>
</tr>
<tr>
<td>Current Assets/Total Assets</td>
<td>.5863</td>
<td>.5982</td>
</tr>
<tr>
<td>Cash/Total Assets</td>
<td>.0774</td>
<td>.0356</td>
</tr>
<tr>
<td>Long-Term Debt/Total Assets</td>
<td>.1561</td>
<td>.3116</td>
</tr>
<tr>
<td>Natural Log of Sales</td>
<td>5.8338</td>
<td>5.3310</td>
</tr>
<tr>
<td>Fraud reported after audit report date</td>
<td>.1750</td>
<td>.8750</td>
</tr>
</tbody>
</table>
Table 6
Logit Results For Nonstressed Bankrupt (NSB) Versus
Nonstressed Nonbankrupt (NSNB)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>$\chi^2$</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-3.9686</td>
<td>1.43</td>
<td>.232</td>
</tr>
<tr>
<td>Net Income/Total Assets</td>
<td>-16.1886</td>
<td>1.11</td>
<td>.293</td>
</tr>
<tr>
<td>Current Assets/Sales</td>
<td>1.3760</td>
<td>.45</td>
<td>.504</td>
</tr>
<tr>
<td>Current Assets/Current Liabilities</td>
<td>-.2246</td>
<td>1.55</td>
<td>.213</td>
</tr>
<tr>
<td>Current Assets/Total Assets</td>
<td>4.5921</td>
<td>2.72</td>
<td>.099</td>
</tr>
<tr>
<td>Cash/Total Assets</td>
<td>-19.5476</td>
<td>2.72</td>
<td>.099</td>
</tr>
<tr>
<td>Long-Term Debt/Total Assets</td>
<td>8.5606</td>
<td>4.79</td>
<td>.029*</td>
</tr>
<tr>
<td>Natural Log of Sales</td>
<td>-.1813</td>
<td>.38</td>
<td>.270</td>
</tr>
<tr>
<td>Fraud reported after the audit report date</td>
<td>1.0116</td>
<td>3.06</td>
<td>.040*</td>
</tr>
</tbody>
</table>

Full Model $R^2 = .226$

$\text{NSB} = \text{Nonstressed Bankrupt (coded as 1)}$
$\text{NSNB} = \text{Nonstressed Nonbankrupt (coded as 0)}$

P values are for two-tailed hypotheses for the intercept and ratios, but one-tailed for sales and fraud. Hypotheses rejected at an alpha level of .05 are indicated by *.
Table 7
Mean Variable Values for the Stressed Bankrupt Qualified (BQ) and Stressed Bankrupt Unqualified (BNQ) Groups

<table>
<thead>
<tr>
<th>Variable</th>
<th>BNQ (64 firms)</th>
<th>BQ (54 firms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability of bankruptcy based on ratios</td>
<td>.128</td>
<td>.280</td>
</tr>
<tr>
<td>Days between statement date and audit report date</td>
<td>74.266</td>
<td>98.963</td>
</tr>
<tr>
<td>Fraud reported after audit report date</td>
<td>.125</td>
<td>.204</td>
</tr>
<tr>
<td>Size (natural log of sales)</td>
<td>4.751</td>
<td>4.153</td>
</tr>
</tbody>
</table>
Table 8
Logit Analysis Results for the Bankrupt/Qualified (BQ) Versus the Bankrupt/Nonqualified (BNQ) Stressed Firms

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>( \chi^2 )</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.001</td>
<td>1.05</td>
<td>.306</td>
</tr>
<tr>
<td>Probability of bankruptcy based on ratios</td>
<td>2.288</td>
<td>7.19</td>
<td>.004*</td>
</tr>
<tr>
<td>Days between statement date and audit report date</td>
<td>.018</td>
<td>4.92</td>
<td>.013*</td>
</tr>
<tr>
<td>Fraud reported after audit report date</td>
<td>.747</td>
<td>2.56</td>
<td>.945</td>
</tr>
<tr>
<td>Size (natural logarithm of sales)</td>
<td>-.272</td>
<td>3.36</td>
<td>.033*</td>
</tr>
</tbody>
</table>

Full Model \( R^2 = .105 \)

\[ \chi^2 = 25.11 \quad .000 \]

BQ = Bankrupt/Qualified (coded as 1)
BNQ = Bankrupt/Nonqualified (coded as 0)

P values are for one-tailed hypotheses (as specified in the text) except for the intercept. Hypotheses rejected at an alpha level of .05 are indicated by *.
NOTES

1. The newly issued auditing standard, SAS No. 59 [AICPA, 1988], indicates that the auditor will not qualify the opinion when there is evidence that an entity will not be able to meet its obligations when they become due. Rather, the auditor will issue an unqualified opinion with an explanatory paragraph. SAS No. 34 [AICPA, 1981] was in effect during the time period of our sample. Under SAS No. 34, the auditor would issue a qualified opinion in similar circumstances. Thus, in this paper the term "qualification" is used.

2. SAS No. 16 [AICPA, 1977], the standard in effect during the time period covered by this research, required the auditor to plan the audit to search for material misstatements. SAS No. 53 [AICPA, 1988] increases responsibility by obligating the auditor to design the audit to provide reasonable assurance of detection.

3. It is acknowledged that, since this fraud was not reported until after the audit report date, it may not have been possible for the auditor to discover it during the audit. However, for expository purposes, this type of fraud is referred to as "hidden fraud" in the remainder of the paper.

4. Previous studies have not found size to be an important variable in predicting bankruptcy. Those studies, however, did not separately examine stressed and nonstressed companies. Therefore, size also is included as a control variable.

5. For the nonbankrupt companies the "bankruptcy" date was randomly generated. This was necessary so that news items could be appropriately gathered for the nonbankrupt sample.

6. The categories were fraud against the company by an employee, fraud against another company, fraud against a lender, fraud against stockholders, fraud against consumers, fraud against the government, misrepresented financial statements, auditor fraud, and other fraud.

7. It is acknowledged that the audit delay measure as used in this research cannot be construed as a perfect proxy for audit extensiveness.

8. For purposes of this research, any opinion other than the going-concern opinion was considered to be unqualified.

9. Since this period covers a variety of economic conditions, the bankruptcy model reported below was separately estimated for two subperiods of the sample period: 1974-1981 and 1982-1985. The results of the estimations were essentially the same, indicating some degree of stability of the relationships over the 1974-1985 period.

10. There is no need for the sample to be proportional to the population proportions as long as proper adjustments are made for the fact that they are not equal. See Maddala [1983] pp. 90-91. These adjustments were made where appropriate in the analyses in this study.
11. Note that in calculating these error rates the stressed sample was used as the base. This is because the auditor will apparently only consider stressed companies as potential recipients of the going-concern opinion. If the entire control sample is used as a base for the Type I error rate, that error rate would be 2.5 percent. If the entire bankrupt sample is used as a base for the Type II error rate, that error rate would be 60 percent.

12. The audit procedures may result in the fraud being either uncovered or prevented. For convenience sake, throughout the remainder of this paper the term "discovery" is used to include both of these concepts.

13. Results are mixed on the relation between auditor and audit extensiveness. Newton and Ashton [1989] found that the more structured auditing firms had a greater audit delay while Williams and Dirsmith [1988] found an inverse relationship between audit structure and audit delay.
A Laboratory Market Investigation of the Demand for Strategic Auditing

by

Steven J. Kachelmeier

Discussant: David E. Wallin

Section 2
A LABORATORY MARKET INVESTIGATION
OF THE DEMAND FOR STRATEGIC AUDITING

Steven J. Kachelmeier
Assistant Professor
University of Texas at Austin
Department of Accounting, CBA 4M.202
Austin, Texas 78712
Ph. (512) 471-3517

September 1990

This paper is based on my doctoral dissertation at the University of Florida. I am indebted to the members of my Ph.D. committee: Rashad Abdel-khalik (chairman), Michael Bamber, John Lynch, and Bill Messier. Helpful comments and suggestions were received from Stan Baiman, Doug DeJong, John Dickhaut, Nick Dopuch, Paul Newman, Michael Schadewald, and participants at the 1989 meeting of the American Accounting Association. I am grateful to Bruce Nearing and Sundararaman Thiagarajan, who assisted me in conducting the laboratory markets. Financial support was provided by the Deloitte, Haskins & Sells Foundation and by the University of Florida.
A LABORATORY MARKET INVESTIGATION
OF THE DEMAND FOR STRATEGIC AUDITING

ABSTRACT

Past laboratory market research exploring the demand for auditing has generally assumed an exogenous audit function, where reliable investigations can be obtained for a fee. This study advances the structure of these settings by incorporating human, strategic auditors. Laboratory markets were conducted where the profits earned by buyers were conditional upon the reports of sellers. Buyers were able to investigate these reports by purchasing audit services from a third group of subjects. However, auditors made private investigation decisions in certain markets, and thus were themselves subject to risks of moral hazard. Findings indicated that despite the complexity of the setting with its multiple sources of moral hazard, both sellers and auditors expended the cost of diligent effort most of the time. There was evidence of significant associations between actions taken and negotiated prices, indicating the presence of multiperiod reputations. However, the highest prices paid to auditors and the lowest frequencies of seller misrepresentations occurred in the control markets without auditor moral hazard.
A LABORATORY MARKET INVESTIGATION
OF THE DEMAND FOR STRATEGIC AUDITING

One frequently cited role for auditing is the mitigation of information asymmetries between owners and managers of firms [Scott, 1984]. This view recognizes that typical agency theoretic solutions to moral hazard problems between owners and managers involve compensation schemes contingent upon managerial output. However, if neither effort nor output is observable by owners, such solutions may not be attainable. Auditing, therefore, arises to provide certification of reported outcomes, making contingent compensation schemes feasible.

Recent experimental work has explored the stewardship role of auditing in laboratory market settings.¹ DeJong, et al. [1985] conducted laboratory markets in which investigations lowered the frequency of suboptimal private cost choices (analogous to managerial effort choices) by sellers. However, the investigations in DeJong et al. [1985] were of the actual cost (effort) choices of sellers, not of outcome representations. As Scott [1984] points out, auditors generally do not observe or attest to managerial effort, but rather attest to the financial outcomes which managers report. Dopuch, King, and Wallin [1989] studied laboratory markets with both cost and outcome reporting decisions, demonstrating that credibility mechanisms (auditing) could mitigate welfare losses even when only outcome reports could be investigated.

The present study accepts as a maintained hypothesis the improved efficiency from auditing which these studies demonstrated. However, past studies have assumed an exogenous audit mechanism, where reliable investigations could be obtained simply by paying a fixed cost. The next logical step is to relax this assumption by providing for human, strategic auditors.² Thus,
there are three distinct types of participants in the laboratory markets of this study: buyers, sellers, and investigators, analogous to the roles of owner, manager, and auditor. When both managers and auditors are subject to moral hazard risks, it is unclear whether auditing will continue to mitigate welfare losses due to asymmetric information. In brief, the auditor may choose to attest to managerial reports without paying the cost of a diligent investigation, and the manager’s best response to this strategy may be to misrepresent outcomes and expend suboptimal effort.

There are two other important differences between the laboratory markets of this and prior research. First, owners and managers in this study negotiate contingent payment contracts conditional upon outcomes reported by managers after the negotiations take place. In this manner, a closer consistency is achieved with the agency theoretic underpinnings of this research. Prior research has generally used single-price negotiations, where sellers’ representations are revealed as part of the negotiation process. Second, there is no outcome feedback after each period other than that provided by the audit function. As noted by Dopuch, King, and Wallin [1989], King and Wallin [1990], and Baiman [1979], nonaudit outcome feedback can itself serve a monitoring role similar to auditing. By preventing such feedback, this study isolates the role of auditing for monitoring strategic behavior.

There are two levels of research objectives. The first level is exploratory, documenting the extent to which moral hazard risks affect observed behavior in the richer setting considered here. In the parlance of experimental economics, this research can thus be viewed as a "boundary" experiment (see Smith [1982b]), testing whether the welfare loss mitigation observed in prior studies will continue in a more complex environment with strategic auditors, contingent compensation, and no outcome feedback. Specifically, this
research asks 1) whether managers and auditors will shirk on their choices of effort level, 2) whether managers will misrepresent outcomes, given their effort choices, and 3) whether owners will consider perceived shirking incentives in their price negotiations for managerial and auditor services. The second level of objectives addresses the factors associated with observed behavior. In particular, the incremental effect of auditor moral hazard is examined by conducting laboratory markets both with and without auditor effort choices. The importance of an independent public signal of auditor effort is also examined by conducting markets with and without such a signal.

The remainder of this paper consists of four sections. The next section develops the laboratory market model and hypotheses to be tested, followed by a second section describing the specific design and methods used. Laboratory market results are presented and discussed in the third section, and the final section of the paper offers concluding comments.

MODEL AND HYPOTHESES

Consider a market with three types of individuals: owners (O players), managers (M players), and auditors (A players). Only M players may take productive actions in this market. However, M players are unable to keep the outputs from their decisions, and hence must sell their productive services to O players. For simplicity, assume that each M player exerts either low or high effort. As a surrogate for effort, assume that the M player pays either a low cost (C_L) or a high cost (C_H). These costs lead to valuable outcomes, which can also be either low (V_L) or high (V_H). Assume that p(V_L | C_L) = 1 and that p(V_L | C_H) = p(V_H | C_H) = .5. These conditional probabilities capture an environment in which outcome depends both on effort and on an independent state
variable. Given these probabilities, a choice of \( C_H \) maximizes the total expected surplus when the following inequality is imposed:  

\[
.5(V_H - V_L) > C_H - C_L
\]  

(1)

Assuming that (1) is satisfied and players are risk neutral, it is in the M player's best interest to pay the high cost \( C_H \) if s/he is remunerated for the resultant increase in expected output. But if cost (effort) choices are unobservable to O players, contracts contingent on managerial cost choices are not possible, and M players face a single-period shirking incentive to pay the lower \( C_L \). The excess of the left over the right side of (1) would then be the expected welfare loss due to unobservable effort.

The common solution to this moral hazard problem is to make managerial remuneration contracts contingent upon outcomes as opposed to effort. However, this solution assumes that outcomes are publicly observable. If instead M players privately observe outcomes and O players can only observe M player outcome reports, there exists a second level of moral hazard: M players may misrepresent the outcome to be \( V_H \) even when the true outcome is \( V_L \). Let \( \hat{V} \) be the M player's outcome representation. The O player may negotiate a contingent compensation contract \( (R | \hat{V}) \), where the M player receives \( R_H \) (\( R_L \)) when s/he reports \( \hat{V}_H \) (\( \hat{V}_L \)). In this setting, O players may demand audits to ensure themselves that a report of \( \hat{V}_H \) means that the true outcome is in fact \( V_H \).

Suppose, however, that auditors (A players) are themselves subject to risks of moral hazard. Specifically, assume that an A player can only observe an M player's true outcome if the A player privately pays an investigation cost \( T \). For simplicity and to focus on strategic incentives, assume that the A player observes the true outcome with certainty by paying \( T \), whereas failure to pay \( T \) provides no information to the A player. The A player's audit report can take only two forms: (1) the M player misrepresented the true outcome (leading
to a penalty against that M player) or (2) no evidence of a report misrepresentation was found. As summarized in Figure 1, the first type of report (analogous to a qualified audit opinion) occurs only if the A player pays the investigation cost T and finds the M player's report to be false. By contrast, the second type of report (analogous to an unqualified audit opinion) can occur either if the A player pays the cost T but finds the M player's report to be truthful or if the A player fails to pay T.

[INSERT FIGURE 1]

The following time line summarizes the game:

- O and M players negotiate contracts R contingent upon reported outcomes V to be announced later
- M players choose C_H or C_L, observe V, and commit to a report of V_H or V_L
- O players contract with A players for audits
- V is publicly announced for each M player
- A players choose whether or not to pay T, and announce audit findings

At the end of the time line, any audit penalty transfers from M players to O players are imposed. At the end of a game consisting of several iterations of this time line, O players receive the actual values V.^5

Single-Period Incentives

The structure modeled above was operationalized in laboratory markets consisting of several iterated periods. However, as a benchmark for the consideration of predicted behavior, consider the incentives players would face if the above time line was played only once. It will prove helpful to start at the end of the time line and use backward induction to infer Nash strategies. The final strategic decision is the A player's choice of whether or not to pay the investigation cost T. Importantly, O players cannot infer this choice from a favorable A player's report, since an audit report that the M player's
representation is truthful could occur whether or not T is paid. Since A player fees cannot be made contingent upon the payment of T, there is no single-period incentive for A players to pay T.6

If M players reason that A players have no incentive to pay T, there is no reason for M players to report truthfully. Hence, in the single-period game, the Nash response of the M player is to always report $\hat{v}_H$, even under the threat of audit penalty. But if the M player always reports $\hat{v}_H$, s/he has no incentive to pay the higher cost. Thus, the unique single-period Nash equilibrium is for M players to pay $c_L$, report $\hat{v}_H$, and for auditors (if employed) to not pay T, thereby not discovering M player misrepresentations. However, rational 0 players would be aware of these incentives, and would not be willing to pay M players any more than $v_L$ irrespective of M players' reports. Further, 0 players would realize that auditor moral hazard risks render audit reports worthless, and would not be willing to engage A players for any positive fee.

It is important to note that the single-period prediction of no demand for auditing exists even in the absence of opportunities for explicit collusion or side-payments between A players and M players. In essence, the model forces auditors to be independent, but this independence is only of the weak form as defined by Antle [1984]. That is, while A players are not allowed to accept side-payments from M players, they are allowed to issue opinions without exerting diligent audit effort. The cost of this effort leads to the collapse of the audit market in the single-period structure described above.7

Multiperiod Incentives

This subsection explores the possibility that a multiperiod setting with repeated contracting may reduce the adverse effects of moral hazard present in the single-period Nash equilibrium. That is, if the above game is repeated
several times, there may be reputation-building incentives which counter the single-period incentives for M and A players to shirk.8

First consider A players. If an A player pays the cost of investigation T and finds an M player's report to be false, that A player would conclusively demonstrate that s/he did indeed pay T. This in turn would presumably enhance the willingness of O players to negotiate high investigation fees for that A player in a competitive market, possibly providing A players with a multi-period incentive to pay T. More importantly, if M players believe that A players will elect to pay the investigation cost T with some positive probability, then M players should in turn report low outcomes truthfully if the following inequality holds, as shown in the Appendix:

\[ \gamma \Theta U > R_H - R_L \]  

(2)

where: \( \gamma \) = the perceived probability that an engaged auditor will choose to pay T,

\( \Theta \) = the probability that an auditor is engaged,

\( U \) = the exogenous penalty imposed on an M player for a detected false report,

and \( R_H \) (\( R_L \)) = negotiated M player fees contingent on a high (low) reported outcome, as defined earlier.

The penalty parameter U is exogenous in this model, and the probability of audit \( \Theta \) can be reasonably approximated by the ratio of A players to M players, assuming that each A player will be engaged for some positive fee and restricting A players to only one audit contract each period.

For any perceived probability \( \gamma \), inequality (2) suggests an inverse relationship between the negotiated premium for a high reported value (\( R_H - R_L \)) and the frequency of truthful M player reports of low values. Inequality (2) also suggests that the frequency of M player misrepresentations will be lowest in markets without auditor moral hazard, ceteris paribus, since \( \gamma \) would then
equal unity. As described in the method section, control markets with no
auditor investigation costs are conducted to examine this prediction.

If (2) holds, managers stand to lose more in expected penalties by
misrepresenting low outcomes than the negotiated premium for a high outcome
report. An M player who elects to report low outcomes truthfully for this
reason also benefits by paying the high cost \( C_H \), as long as the following
inequality also holds (see Appendix for derivation):

\[
.5(R_H - R_L) > C_H - C_L
\]

(3)

where all terms are as previously defined. Note that (3) is independent of the
perceived A player strategic parameter \( \gamma \), as long as (2) is satisfied.

Suppose, however, that (2) does not hold. In such cases, M players face
an incentive to misrepresent low outcomes, securing a premium \( (R_H - R_L) \) in
excess of the expected detection penalty \( \gamma \theta U \). However, even when the
M player's strategy is to misrepresent low outcomes, a choice of high effort
\( (C_H) \) is still optimal in cases where

\[
.5 \gamma \theta U > C_H - C_L
\]

(4)

as shown in the Appendix. If (4) holds, the expected audit penalty under low
effort and misrepresented outcomes exceeds the cost of high effort, motivating
\( C_H \) as an equilibrium choice.

In sum, if both (2) and (3) hold, it is in the M player's best interest to
pay \( C_H \) and report truthfully. If (2) holds but not (3), the M player has no
motivation to pay \( C_H \), but should still report truthfully. If (2) does not hold
but (4) holds, the M player should pay \( C_H \), but faces an incentive to misrepre-
sent the outcome if it turns out to be low. Finally, if neither (2) nor (4)
holds, there is no monetary incentive to pay \( C_H \) or report truthfully.

It should be emphasized that the above analysis of M player strategies is
based on an M player's best single-period response, given negotiated fees and
perceived A player strategies. However, the same multiperiod reputation reasoning used to motivate the possible incentive of A players to pay the investigation cost \( T \) may also apply to M players. That is, it is possible that M players may attempt to build multiperiod reputations for optimal effort and truthful reporting, even if it is not in their best single-period interests to do so. For example, assume that inequality (2) does not hold. In such cases, the M player can obtain a contract premium \( (R_H - R_L) \) in excess of the expected penalty \( \gamma \theta U \) by misrepresenting a low outcome to be high. However, the M player may anticipate that a detected misrepresentation would lead to a loss in future reputation in addition to the immediate expected pecuniary penalty of \( \gamma \theta U \). That is, M players may reason that 0 players would not continue to negotiate large premiums in future periods for high reported outcomes which are likely to be misrepresented. Thus, incentives on the part of M players to sustain high negotiated prices in future periods may lead them to report truthfully even if (2) does not hold. Inequality (2) (and by implication, (3) and (4)) should therefore be viewed as conservative boundaries on the predicted degree of M player shirking in a multiperiod setting.

The multiperiod reasoning developed in the above paragraphs ignores the backward inductive logic that in a repeated finite game with one single-period Nash strategy, all repetitions are strategically equivalent to a single-period game (e.g., see Selten [1978]; Davis [1985]). Thus, it is not immediately obvious whether multiperiod forces would change the single-period predictions in the present setting. However, there is evidence from other experimental settings that observed actions are often inconsistent with backward inductive logic. Prominent examples are provided by Camerer and Weigelt [1988, §4.5.2] and by McKelvey and Palfrey [1989]. Both of these studies develop arguments that subjects act as if they perceive some positive probability that their
opponents will "irrationally" ignore the implications of backward induction from the final period. If "irrational" players are construed as a separate type, it is possible to reconcile such settings to the reputation model of Kreps and Wilson [1982]. Further, the finite period logic is sensitive to subjects' beliefs about when the experiment will end (see Roth and Murnighan [1978]; Shubik [1982, §9.4]). In this experiment, subjects were not told how many periods there would be.

**Hypotheses**

As noted above, strict single-period reasoning predicts that A players will not pay the investigation cost $T$, and that M players will therefore have no incentive to report truthfully or pay the cost of high effort. However, in a multiperiod setting, possible reputation-building incentives on the part of A players and perceptions of those incentives by M and 0 players may lead to a Pareto-superior outcome. The first set of hypotheses is stated in the alternative form which predicts that such incentives will be sufficient to depart from the single-period Nash equilibrium of consistent shirking:

**H1:** Players will tend to elect nonshirking strategies.

**H1a:** A players will tend to pay the investigation cost $T$.

**H1b:** M players will tend to report $\hat{V}_L$ when they observe $V_L$.

**H1c:** M players will tend to pay the high cost $C_h$.

If nonshirking strategies prevail (supporting H1) due to the desire of M and A players to sustain higher fees with multiperiod reputations, one would expect to see an association between negotiated fees and the strategies elected. Hypothesis set H2 examines the relationship between observed strategies and the negotiated prices which 0 players agree to pay:
H2: The strategies adopted by individual M and A players will be significantly associated with negotiated prices.

H2a: Unfavorable audit findings will lead to downward revisions of subsequent prices contingent upon high M player outcome reports.

H2b: Unfavorable audit findings will lead to upward revisions of subsequent A player audit fees.

H2c: Negotiated premiums for high M player reports will be positively associated with the frequencies of M player cost choices (predicted from inequality (3)).

H2d: Negotiated premiums for high M player reports will be negatively associated with the frequencies of truthful M player reports of low values (predicted from inequality (2)).

Hypotheses H2a and H2b examine the information content of audit reports as a basis for market revisions of subsequent M and A player fees. Hypotheses H2c and H2d examine the relationships between contract prices and M player strategies. Specifically, inequality (3) implies a direct relationship between the premiums negotiated for a high reported value and the frequency of high cost choices. Conversely, inequality (2) predicts an inverse relationship between the price premiums for a high reported value and the likelihood that M players will truthfully report low values.

Hypothesis set H3 directly examines the influence of auditor moral hazard on the results from H1 and H2. To do this, auditor moral hazard is varied, with the investigation cost T present in some markets but not in others. Within the auditor moral hazard markets, certain markets also include a public signal analogous to litigation against auditors. This second variation is intended to shed further insight on the influence of auditor reputations.

H3: Market variations affecting the degree of auditor moral hazard will significantly influence strategies and prices.

H3a: M player cost and reporting strategies will significantly vary between the different market types.
H3b: A player investigation frequencies will be significantly higher when a public signal is present.

H3c: Prices received by both M and A players will significantly vary between the different market types.

METHOD

Six laboratory markets were conducted with ninety graduate and senior undergraduate business student volunteers. There were two market runs for each of three types: 1) control markets where auditors could costlessly observe M player reports (markets 1A and 1B), 2) auditor moral hazard markets with no public signal (markets 2A and 2B), and 3) auditor moral hazard markets with a public signal which revealed auditor choices with probability .25 (markets 3A and 3B). Details on the market variations are explained later.

Each market required fifteen participants. Seven of these were randomly designated as O players, five as M players, and three as A players. Subjects were seated by type at separate tables. There are two reasons why a different number of subjects were assigned to each type. First, in competitive laboratory markets where demand exceeds supply, prices converge upward to the expected redemption values of buyers (e.g., see Smith [1982a]). If inequality (1) holds, these expected values are higher when M players pay $C_N$. Thus, a structure of excess demand may give players the incentive to attempt to build multiperiod reputations for electing the high cost and reporting truthfully. Second, this structure allows prices to be roughly interpreted as the expected values perceived by O players. As previously indicated, if shirking is predominant, one would expect these values (and prices) to be no higher than the lowest possible outcome $V_L$ for M players and to be zero for A players.

Experimental instructions (available upon request) were read aloud to subjects. The instructions avoided possible role constructs inherent in the terms "owner," "manager," and "auditor" by using the neutral labels of X, Y,
and Z players, respectively. (Discussion in this paper will continue with the more descriptive labels of O, M, and A players in order to simplify exposition.) All subjects knew that both costs and redemption values could be either low or high, and that in the auditor moral hazard markets, A players could only observe true outcomes by paying a positive cost. A pre-experimental questionnaire was administered to ensure that subjects understood the instructions.

Subjects were given profit calculation sheets which informed them of their specific costs or values. All M players were told that they could pay either a high cost \( C_H \) of $0.90 or a low cost \( C_L \) of $0.30 for each experimental unit sold to an O player. The instructions common to all subjects emphasized that the high cost would give a 50% chance that the O player's redemption value would be high, whereas the low cost always meant that the O player's redemption value would be low. All O players were informed that the high redemption value \( V_H \) was $4.00 and that the low redemption value \( V_L \) was $1.00. These parameters satisfy inequality (1). In market structures 2 and 3 (auditor moral hazard), A players were informed that they would have to pay an investigation cost \( T \) of $0.35 to observe the true value of an M player's unit.

**Market Sequence**

**Stage 1: O-M Negotiations.** Each market period consisted of five stages. During the first stage, O and M players negotiated prices for M player units, the surrogate for managerial services, in an oral double auction. Since these units could be of either high or low value (as yet undetermined), prices were in the form of contract pairs. For example, a typical offer from an M player would be "I will sell a unit to any O player for $3.00 if high and $1.00 if low." A typical bid from an O player would be "I will buy a unit from player M2 or player M5 for $2.50 if high and 70 cents if low." Players were required
to specify both the high value and the low value contract price, although these amounts could be the same. Further, because the possibility of M player reputations is an important consideration, players were allowed to restrict their bids to certain player(s), at the bidder’s option. Without this feature, M players might have been less likely to consider costly reputation-building strategies, since there would have been nothing to stop other M players from "free riding" on one M player's reputation.

Both M and O players were free to propose or accept contract proposals during these negotiations, with the players allowed to form only one contract each. The negotiation period continued until each M player was able to secure a contract, or until the time limit of five minutes elapsed. There were no instances in any of the six markets where an M player failed to secure a contract with an O player for lack of time.

Stage 2: M Player Cost and Reporting Decisions. After the O-M negotiation period had ended, M players were asked to indicate whether they wished to pay the low cost of 30 cents or the high cost of 90 cents. A contract sheet was provided for this purpose, where M players circled either HIGH or LOW. The experimenter then dealt one card from a standard 52-card deck (publicly reshuffled each period) face down to each M player. The M players privately observed these cards, which indicated their true unit values. If an M player circled the HIGH cost, a red card signified a high unit value, whereas a black card signified a low unit value. If an M player circled the LOW cost, his/her unit value was always low, irrespective of the color of the card dealt. After the M players observed their true unit values, they again circled HIGH or LOW on their contract sheets to indicate the values they wished to report to the O players who had purchased their units. An M player with a high true value always circled the HIGH report, but an M player with a low true value
could report HIGH or LOW. Since M players made their cost and reporting
decisions before O players negotiated audit contracts, M players could not
condition their strategies on whether or not their reports would be audited.

Stage 3: O-A Negotiations. The O players were aware that an M player
might report a high valued unit even though the true value of the unit was low.
Although M players had committed to their reporting decisions in Stage 2, these
reports had not yet been announced, so all five O players with an M player
contract faced the possibility of a misrepresented high value. In Stage 3,
O players had the opportunity to secure investigations of reported values from
one of the three A players. Both O and A players were free to propose or
accept investigation fees in an oral double auction, with the restriction that
each player could have only one investigation contract each period. Negotia-
tions between O and A players were limited to five minutes.

Since there were five O-M contracts each period but only three A players,
the market structure induced excess demand for A players, for reasons similar
to those in the O-M market. That is, in an excess demand market, negotiated
A player fees are roughly interpretable as the perceived value of auditing. An
additional implication of excess demand in the market for audit services is
that M players would know that there was a limited (3/5) probability that their
representations would be investigated. This limit was present in all six
markets, possibly enticing M players to engage in some shirking behavior even
in the control markets (1A and 1B) with no auditor moral hazard.

Stage 4: M Player Reports Announced. After the O-A negotiation period
had ended, the researcher publicly announced the reported values of each M
player. The rationale for delaying the announcement of M player reports until
this stage was to preserve the desired market structure of excess demand for
auditing in stage 3, such that audit fees would approximate the value of
auditing as perceived by O players. If M player reports were announced earlier, it would be possible that fewer audits would be demanded (due to low reported values) than the number of audits available (three each period). However, in order to allow O players to condition their demand for auditing upon reported values, any audit contract between an O and A player was nullified if the corresponding M player reported low. An audit would have no value if an M player admitted to a low outcome, so O players did not have to pay any audit fees in such cases. Negotiated audit contracts were valid only if an M player reported high, leading to the fifth stage.

Stage 5: A Player Decisions and Reports. The three market designs differed only in this final stage. In markets 1A and 1B (no auditor moral hazard), the researcher simply announced the true value of each M player unit covered by an A player investigation. In markets 2A and 2B (auditor moral hazard but no public signal), A players were first asked to circle YES or NO on their contract sheets to indicate their willingness to pay a $0.35 investigation cost. If an A player circled YES, the researcher announced that the investigation did or did not find the corresponding M player's report to be false, depending upon whether the M player reported truthfully. If an A player circled NO, the researcher always announced that the investigation did not find the M player's report to be false. All players were aware of these rules.

Markets 3A and 3B proceeded identically to markets 2A and 2B, with the one additional step that after A players' reports were announced, another card from the deck (the public signal) was turned face up and shown to all. If the card was a heart, diamond, or club, nothing further was done, and the period ended. But if the card was a spade (a 25% chance), the true values of all units covered by an audit contract were announced. This had the effect of making all
players aware of instances where an M player reported falsely, but the A player did not pay the cost necessary to detect the misrepresentation.

Announcements of false reports led to monetary penalties. In all markets, a publicly announced penalty of $3.00 was transferred to the appropriate O player from the profits of an M player whose report was found by an A player to be false. However, M players were not penalized when discovery of a false report was due to the public signal (markets 3A and 3B only). In such cases, the only penalty was the forfeiture of the corresponding A player's fee for that period. The public signal was intended only to (possibly) affect reputations, not to dissuade shirking with additional monetary penalties. That is, even with the public signal, A players would still face a single-period shirking incentive to not pay the investigation cost.15

Profit Calculations and End-of-Period Procedures. Each subject was provided with profit sheets to tally information regarding each period's outcomes. M and A players profited from the excess of their contract prices over their costs. For M players, revenues depended upon the reported (not necessarily true) values of units sold. O players profited from the excess of redemption values corresponding to the true values of units purchased over the sum of the prices paid to M players corresponding to the reported values of those units and the investigation fees paid to A players, if any. Those O players who purchased a unit with a high reported value were asked to leave their profit calculations blank until the end of the experiment, unless a perfectly reliable audit report (markets 1A and 1B only) or public signal (markets 3A and 3B only) confirmed that the true value was indeed high. It was thus emphasized to O players that they could not be sure of their true redemption values for units reported to be high due to the possibility of M and/or A player shirking. An assistant entered data for each period into a
personal computer which was programmed to calculate profits and produce summary sheets for each player at the end of the session.

Since players faced the possibility of losses from their contracts, each period started with a monetary endowment of $1.00 for 0 and A players and $.25 for M players. In the two markets without investigation costs (IA and IB), the A player endowment was reduced to $.65/period. Profits (losses) each period were added to (subtracted from) these endowments.

The five-stage process described above repeated for thirteen periods, which typically took two to two and one-half hours of a three-hour scheduled block. The number of periods was not announced to subjects. At the termination of the session, subjects filled out post-experimental questionnaires, were paid, asked not to discuss the experiment with other students, and dismissed. Actual payments to subjects generally ranged between $10 and $30, with an average of $18.71. Market variations were run in random chronological order.

RESULTS AND DISCUSSION

Hypothesis Set 1.

Table 1 displays the strategic choice frequencies for each market. While evaluation of this table must rely on qualitative judgment, there seems to be evidence in support of hypothesis set 1. Starting with A player investigation cost choices (hypothesis H1a), it is clear that A players voluntarily chose to pay this cost most of the time (range = 58.8% to 88.5%; average = 73.8%). This was the case even in markets 2A and 2B, where A players faced no public disclosure of their choices. This data can be interpreted as reflecting a desire on the part of A players to build multiperiod reputations for diligence in order to support high negotiated audit fees. Recall that if an A player paid the $.35 investigation cost and if the associated M player was found to
have misrepresented the value of his/her unit, the market would have conclusive evidence of both facts. The post-experimental questionnaires of A players indicated an awareness of this point, as these two comments illustrate:

At one point I was offered $.05 more than my cost but I still paid the cost. I was hoping to raise the fees of the investigation—not to avoid paying the cost. As the investigation fees went up I was very pleased. Additionally, reports of false high values also contributed to this. As the fees were in my mind reasonable, I was firm in my conviction to pay the cost.

-------------------

I initially planned to pay no investigation costs at all. Quickly I discovered the negative impact on fees. [I changed] my strategy to play it straight and only accept higher bids.

[INSERT TABLE 1]

The observed frequencies of M player reporting and cost choices provide indirect evidence that M players were aware of this point as well. As previously discussed, inequality (2) is a lower bound on M players’ equilibrium incentives to report truthfully. With publicly known values of the penalty U ($3.00) and the probability of audit θ (3/5), inequality (2) simplifies to the prediction that M players should report low values truthfully if

$$1.8\gamma > R_H - R_L,$$

(2').

The perceived probability γ that an A player will elect to pay the cost of investigation is of course unobservable in the general case. However, in control markets 1A and 1B it is common knowledge that γ = 1, reducing (2') to the prediction that truthful reporting is the best equilibrium M player strategy whenever $$R_H - R_L < 1.80.$$ Examination of individual contracts revealed premiums less than $1.80 in 51 of the 76 instances in markets 1A and 1B in which there was a low true value. M players reported truthfully in 38 of these 51 cases, or 74.5% of the time. In the 25 instances in markets 1A and 1B of low true values where (2') was not satisfied, M players still reported
truthfully in 15 cases (60.0%), suggesting possible reputation-building incentives incremental to the immediate pecuniary incentives reflected in \(2'\).

Instances of truthful reports were lower in the four treatment markets with auditor moral hazard, as discussed later in the analysis of hypothesis set H3. This is consistent with the notion that perceived values of \(\gamma\) would be lower than unity in these markets. Still, truthful reporting given a low true value occurred over half of the time in all but one market. The exception was market 3B, where low values were truthfully reported only 23.1% of the time. The probable explanation for this anomaly is that negotiated audit fees in market 3B were by far the lowest of the six markets, with an average of only \$43. It is likely that M players inferred that low A player fees implied low incentives for A players to pay the cost of investigation, therefore tempting M players to misrepresent outcomes relatively often.

Stronger evidence against shirking is provided by M player cost choices (hypothesis H1c). In all six markets, M players voluntarily elected to pay the higher initial cost an average of 78.5 percent of the time, with a high degree of consistency between markets (range = 73.8% to 81.5%). This propensity to elect the higher cost is particularly important, since that choice assured the market as a whole of the highest possible expected surplus.\(^{17}\)

In the equilibrium model, cost choices were shown to be governed by inequalities (3) and (4), depending upon reporting strategies. The high incidence of high cost choices in all six markets suggests that both (3) and (4) held most of the time. The data confirm this. Inequality (3) reduces to

\[
.5(R_H - R_L) > .6
\]  

(3')

when the exogenous values of \$0.90 and \$0.30 are substituted for \(C_H\) and \(C_L\), respectively. Examination of individual contracts indicates that (3') held in 361 of the 390 M player contracts across all six markets. Evaluation of (4) is
more problematic, since (4) depends on the perceived (and hence unobservable) value of $\gamma$. However, (4) can be solved for $\gamma$, indicating that high cost choices are optimal whenever

$$\gamma > 2/3$$

(4').

By design, (4') holds trivially in markets 1A and 1B, where $\gamma$ is known to equal unity. In the other markets, M players also acted as if the perceived probability that an A player would choose to investigate exceeded 2/3. This was the case even in market 3B, where M players paid the high cost 73.8 percent of the time. It was argued earlier that the relatively low negotiated audit fees in market 3B probably led to relatively low perceptions of $\gamma$, tempting M players to frequently misrepresent low outcomes in that market. Yet, the high frequency of high cost choices indicates that perceived values of $\gamma$ exceeded 2/3. These two observations are not necessarily inconsistent. For example, a value of exactly 2/3 for $\gamma$ implies (marginal) satisfaction of (4'), inducing high cost choices. However, for all 65 M player contracts in market 3B, a $\gamma$ value of 2/3 would fail to satisfy (2'), giving M players a monetary incentive to misrepresent low outcomes.

**Hypothesis Set 2.**

If the frequencies reported above reflect attempts to build multiperiod reputations, one would expect to observe an association between the signals observed in the markets and the prices negotiated. First, hypothesis H2a calls for an analysis of the association between M player fees and A player audit findings. The test for this association is performed on the subset of 128 M player reports of a high value that were covered by an audit investigation. These 128 investigated reports are categorized by whether the corresponding A player determined the M player's report to be false. The change in the high conditional price from the market period covered by the audit contract to the
next period is then used as the dependent variable in a conventional t-test for the effects of audit findings on price.

Table 2 reports the results of this analysis. For all markets combined, M player fees contingent on a high reported value increased by 2.50 cents on average following a favorable audit report, but fell by 8.50 cents on average following an audit detection of a report misrepresentation. A t-test shows this difference to be significant (t = 2.44; p = .008), but this test does not take into account differences between markets. Separate analyses are therefore performed at the individual market level, reaching significance at the .05 level in only two of the markets (2A and 3B). However, all six markets are in the expected direction (lower price shifts following an unfavorable audit), which a sign test indicates would occur with probability .016 (i.e. 1/2^6) under the null hypothesis. Therefore, evidence indicates that investigation findings were relevant in subsequent O-M price negotiations.

[INSERT TABLE 2]

To follow up on the above analysis, the incidence of restricted bids by O players is examined. Out of the 968 total O-M price proposals in all six markets, there were 268 offers to sell by M players and 700 purchase bids by O players. In turn, 147 of the 700 purchase bids were restricted to one or more particular M players. To test the possibility that O players would tend to identify and thus restrict their bids to the M players with the most favorable reputations, the correlation between the number of restricted bids and the number of audit penalties for each M player is computed across all six markets. This correlation is negative, as expected, and significant in a one-tailed test (r = -.30; p = .05), corroborating the conclusion that O players were sensitive to M player reputations.
Hypothesis H2b calls for an analysis of whether negotiated A player fees were responsive to audit findings. This analysis is performed on the same subset of 128 audit investigations of a reported high value used in the test of M player fees, with the dependent variable now defined as the change in the A player's negotiated fee in the period subsequent to the audit. Results are shown in Table 3. Contrary to the results for M player fees, Table 3 shows no evidence that A player fees were responsive to audit findings across all markets or in the markets considered separately, thus failing to support hypothesis H2b. Thus, despite post-experimental questionnaires indicating that many A players perceived an association between their choices and investigation fees, such fees did not appear to differentially react to audit findings.

[INSERT TABLE 3]

While the discussion thus far has focused on audit findings as the principal means of communication of cost and reporting choices in a multiperiod setting, hypotheses H2c and H2d call for the analysis of one other important communication available to M players -- the contingent price contract. Specifically, inequality (3) predicts a direct relationship between the high report price premium \( R_H - R_L \) and the frequency of high cost choices (hypothesis H2c), while inequality (2) predicts an inverse relationship between the high report premium and the frequency of truthful reports of low values (hypothesis H2d). Table 4 presents a regression analysis of these hypotheses.

[INSERT TABLE 4]

The regression in Table 4 models each M player's average price spread (the average difference between the high \( R_H \) and low \( R_L \) contract prices) as a function of the percentage frequencies of high cost choices and truthful reporting choices given a low true value. Indicator variables for each separate market are used in lieu of an intercept term to capture systematic
differences between treatments and replications. This regression is run on thirty observations, reflecting five M players in each of six markets.

There is a significant positive association ($t = 3.20; p = .004$) between average price spreads and percentage frequencies of high cost choices, supporting hypothesis H2c. This indicates that those M players who chose the high cost more often were more successful in negotiating high price premiums. A similar relationship between quality and price was observed by DeJong et al. [1985], but in a simpler setting without contingent contracting and strategic auditors. The present finding indicates that participants were able to differentiate the value of private decisions even in a relatively complex setting with several intervening variables and sources of moral hazard.

The regression also shows a significant inverse association ($t = -1.99; two-tailed p = .059$) between M player price spreads and M player reporting choices, as predicted by hypothesis H2d. That is, those M players who truthfully represented outcomes more often had lower negotiated price premiums. Thus, there is evidence that given a low outcome (which could arise with probability .50 even if the cost choice was high), those M players with the higher price spreads were more tempted to misrepresent their outcomes in order to receive the higher contingent price. This is consistent with Lewicki’s [1983] model of deceptive behavior, which posits that the propensity to lie is a function of the potential reward for lying. This finding corroborates and extends the recent work of Baiman and Lewis [1989], which examined the Lewicki model in a setting involving employment contracting. While Baiman and Lewis [1989] conducted experiments in which deceptive representations could not be detected and exposed, this study suggests that a relationship between the propensity to lie and the monetary reward for lying holds even when there is a potential for audit detection.
Hypothesis Set 3.

The final set of hypotheses examines the effects, if any, of the different market structures on strategic choices and prices. First, hypothesis H3a focuses on M player strategic choices. It is evident from Table 1 that the percentage frequency of high M player cost choices does not significantly vary between markets, remaining about 80% in all six markets. A contingency table test for independence confirms the absence of any significant differences, whether conducted across all six markets ($\chi^2_{df=5} = 2.00; \ p > .50$) or across the three market types ($\chi^2_{df=2} = 0.82; \ p > .50$).

By contrast, there are significant differences between the markets in the frequencies of M player reporting choices. A test for differences between the six individual markets is highly significant ($\chi^2_{df=5} = 25.36; \ p < .001$). The test remains significant when markets are grouped by the presence (markets 2A, 2B, 3A, and 3B) or absence (1A and 1B) of auditor moral hazard ($\chi^2_{df=1} = 6.09; \ p = .014$). Statistically, these tests should be interpreted with caution, since the underlying assumption of independent trials is clearly violated in a laboratory market setting. But it is noteworthy that of the six markets, the two highest frequencies of truthful reporting are in the control markets without auditor moral hazard. The hypergeometric probability of drawing two specific markets from a sample of six is $(2!)(4!)/(6!)$, or $0.067$. Therefore, there is evidence that auditor moral hazard led M players to be more aggressive in their reporting choices, as predicted by inequality (2).

Hypothesis H3b calls for a test of whether the investigation cost decisions made by A players depended upon the presence of a public signal of such choices. Recall that in markets 3A and 3B, each market period ended with a 25 percent chance that all true values covered by an audit contract would be announced, thereby exposing instances of auditor shirking. If A players
reacted to this possibility, one would expect higher frequencies of investigation cost choices in the two markets with a public signal. However, the frequencies reported in Table 1 do not support this conjecture. A contingency table test does not reach significance either for the four auditor moral hazard markets considered separately ($\chi^2_{df=3} = 5.87; p = .118$) or for groupings based on the presence or absence of the public signal ($\chi^2_{df=1} = 0.18; p > .50$). Thus, there is no evidence that the public signal affected A player incentives.

If O players’ perceptions regarding M and A player incentives were different in the three market types, one would expect to see systematically different prices, which is the focus of hypothesis H3c. Table 5 indicates the average prices negotiated in each of the six markets. These prices are plotted by market period in Figures 2-4 for low and high M player fees and for A player investigation fees, respectively. The following repeated-measures ANOVA model was used to test for price differences between the three market structures:

$$\text{PRICE} = \text{TRMT} + \text{REP(TRMT)} + \text{SELLER(REP(TRMT))} + \text{PERIOD} + \text{PERIOD \times TRMT}$$
$$+ \text{PERIOD \times REP(TRMT)} + \text{PERIOD \times SELLER(REP(TRMT))}$$

where:

- **PRICE** = Negotiated market price,
- **TRMT** = Market treatment, a fixed between-subjects factor reflecting the three different market types,
- **REP(TRMT)** = Replication nested within treatment, a random between-subjects factor reflecting the two identical runs of each market treatment,
- **SELLER(REP(TRMT))** = Seller code, a random factor, and
- **PERIOD** = a fixed repeated-measures factor, from 1 to 13.

This ANOVA model measures treatment differences incremental to differences between replications.¹⁸ Trend effects are controlled by incorporating market periods as a repeated-measures factor with thirteen levels.

[INSERT TABLE 5 AND FIGURES 2, 3, AND 4]
A different ANOVA is run for the prices of M player units conditional on both a low and a high reported value and for A player investigation fees. Sources of variation are shown in Table 6. For the ANOVA on M players' prices conditional on a low reported value, the market treatment factor is clearly insignificant, as expected. The sources of moral hazard which varied across market structures did not affect the value of units admitted to be low, and it is therefore not surprising that low conditional prices did not differ between the three market environments.

[INSERT TABLE 6]

For M player units conditional on a high reported value, there is both a significant market treatment effect and a replication effect, indicating that high prices vary between replications, but also that there are incremental differences between the three market structures. Specifically, prices are the lowest in the two control markets without auditor moral hazard (1A and 1B).

To follow up on these differences, orthogonal pairwise comparisons are performed on the treatment means, shown in Table 7. Since market structure 1 represents a control condition, its mean is compared to the average of the means for market structures 2 and 3. The second comparison is a direct test of the difference between structures 2 and 3, intended to capture any incremental effects of the public signal. Table 7 shows that prices conditional on a high reported value in market structure 1 are significantly lower than in market structures 2 and 3 (p = .025). It appears that in market structure 1, where A players were known to be reputable, the resultant high demand for audits led O players to bid up A players' fees at the expense of lower bids for M player units. This explanation is consistent with the ANOVA results for investigation fees, reported below. Table 7 also provides weak evidence (p = .083) that M players' high conditional prices are higher in market structure 3 (public
signal) than in structure 2. There is no apparent explanation for this
difference.

[INSERT TABLE 7]

The ANOVA for A player investigation fees (Table 6) shows only a
marginally significant treatment effect ($F = 4.77; \ p = .117$). However, the
pairwise comparisons of Table 7 indicate that the reason the significance level
is not more pronounced is the lack of any price effects of the public signal,
as evidenced by the clearly insignificant comparison between market structures
2 and 3 ($F = 0.21; \ p > .50$). When the four markets with auditor moral hazard
are aggregated and compared to the two markets without auditor moral hazard,
the difference is significant ($F = 9.30; \ p = .055$). The highest investigation
fees were in control market 1A (average = $.97), followed by control market 1B
($$.87), as compared to averages of $.58 and $.63 for market structures 2 and 3,
respectively. As mentioned earlier, the null probability of drawing two
markets from a sample of six is .067, corroborating the ANOVA results. It
would therefore appear that O players were willing to pay more for audits in
the two control markets where A players were known to be credible.

Supplementary Analyses for End Effects and Sensitivity to Risk Preferences.

All above analyses reflect thirteen laboratory market periods. In order
to examine whether results might have been sensitive to subjects' knowing when
the game would end, an additional two periods (14 and 15) were run where time
permitted (all markets except 1A), following an announcement that there would
be only two more periods. Supplementary analysis of these two "end-effect"
periods indicated that neither prices for M player units nor A player investi-
gation fees dropped in periods 14 and 15. M players continued to elect the
high cost choice substantially all (95%) of the time, and reported low outcomes
truthfully in 16 out of 27 cases (59%) in the auditor moral hazard markets.
A players in market structures 2 and 3 paid the investigation cost for 7 of the 15 (47%) audit contracts in periods 14 and 15. It would therefore appear that any end effects on strategic incentives and price negotiations were not severe.

Although the model motivating this paper assumed risk neutrality, most of the hypothesis tests are not sensitive to this assumption. Further, the cost and redemption values were chosen to dominate any moderate risk aversion on the part of subjects. Some experimental evidence regarding insensitivity to risk preferences is provided by a questionnaire patterned after the certainty equivalent approach described by Harrison [1986], which all subjects completed. There were no correlations at any conventional significance level between estimated risk preferences and market choices.

CONCLUDING COMMENTS, LIMITATIONS, AND EXTENSIONS

This study tested the effects of managerial and auditor moral hazard in competitive laboratory markets. Previous experimental research (e.g., DeJong et al. [1985]; Dopuch, King, and Wallin [1989]; Plott and Wilde [1982]) has shown that behavior consistent with notions of multiperiod reputations can mitigate welfare losses due to moral hazard. This research extended these findings in a relatively complex market setting where one would presume it to be more difficult for reputations to emerge. That is, not only did managers have the opportunity to select suboptimal costs and misrepresent outcomes, but their representations could be corroborated only by strategic auditors, who made private cost choices in four of the six markets. Further, unlike prior studies, outcomes were not revealed until the end of the experiment.

Despite the difficulties associated with sustained nonshirking behavior in these markets, neither managerial nor auditor shirking prevailed. Managers elected to pay a higher cost an average of 78.5 percent of the time, and auditors voluntarily paid an investigation cost 73.8 percent of the time. The
frequency of truthful managerial reports given a low outcome was somewhat lower, but still exceeded 50 percent in five of the six markets. Together, these findings provide further empirical support for prior conclusions that multiperiod market mechanisms can mitigate welfare losses due to information asymmetries. There was also evidence that managers received lower prices following unfavorable audit reports, and that those managers who paid the higher cost more often received the highest contingent price premiums for high outcome reports. However, higher premiums were also associated with lower frequencies of truthful managerial reporting, indicating a relationship between the propensity to issue a deceptive report and the potential monetary reward for deception.

Managerial fees were lower and audit fees higher in two control markets where auditors had no investigation costs. The highest instances of truthful managerial reporting also occurred in these control markets. Conversely, there were no significant effects in the predicted directions of a probabilistic signal of auditor choices, which was present in two of the four auditor moral hazard markets.

The results of this research are subject to the limitations inherent in an experimental study with student subjects. Further, although this study has interpreted most of its findings as multiperiod reputation effects, it should be acknowledged that other explanations are possible. For example, Noreen [1988] points out that ethical predispositions may significantly affect behavior and contracts in moral hazard settings. It may be that some of the nonshirking behavior observed in this study could be due simply to subjects' ethical aversions to opportunism. Since both ethics and reputations play a fundamental role in accounting, future behavioral research should consider ways to isolate the conditions under which these phenomena occur.
APPENDIX

Derivation of Equilibrium Strategic Incentives

This appendix derives the equilibrium incentives for truthful reporting (inequality (2) in the text) and optimal effort (inequalities (3) and (4)). These inequalities result from differentiating the M player's profit function, expressed as follows:

\[ \Pi_M = \alpha \left( 0.5R_H + 0.5\beta R_L + 0.5(1-\beta)(R_H - \gamma\theta U) - C_H \right) + (1-\alpha)\left( \beta R_L + (1-\beta)(R_H - \gamma\theta U) - C_L \right) \]

where:

- \( \Pi_M \) = M player's profit from the sale of an experimental unit,
- \( \alpha \) = probability that the M player will choose to pay the high cost,
- \( \beta \) = probability that the M player will report truthfully, given that the true observed value is low,
- \( \gamma \) = perceived probability that an A player will elect to pay the cost of investigation,
- \( \theta \) = probability of audit, approximated by the ratio of the number of available A players to the number of M players,
- \( U \) = penalty for an audit detection of a false high report,
- \( R_H \) (\( R_L \)) = contingent price negotiated for a high (low) reported outcome, and
- \( C_H \) (\( C_L \)) = exogenously specified cost of high (low) effort.

Differentiation and optimization of this profit function with respect to \( \beta \) indicates that profit is maximized by choosing \( \beta = 1 \) if:

\[ (1-0.5\alpha)[\gamma\theta U - (R_H - R_L)] > 0 \]

which simplifies to:

\[ \gamma\theta U > R_H - R_L \]

for all \( \alpha \in [0, 1] \). This is indicated as inequality (2) in the text.

In order to derive optimal cost choices, the profit function must be differentiated with respect to \( \alpha \), indicating an optimal choice of \( \alpha = 1 \) if:
\[ .5\beta(R_H - R_L) + .5(1-\beta)\gamma\theta U - (C_H - C_L) > 0. \]

Restricting the analysis to pure strategies (which inequality (2) implies in the general case), the optimal choice of \( \alpha = 1 \) simplifies to:

\[ .5(R_H - R_L) > C_H - C_L \]

for \( \beta = 1 \), which is indicated in the text as inequality (3). For \( \beta = 0 \), a choice of \( \alpha = 1 \) is still optimal if:

\[ .5\gamma\theta U > C_H - C_L \]

as stated in inequality (4).
NOTES

1. See Smith, Schatzberg, and Waller [1987] for a review of the application of experimental economics to auditing issues.

2. A progression similar to that envisioned in this study has taken place in the analytic literature. Ng and Stoeckenius [1979] introduced auditing as a verification mechanism for outcomes privately observed by agents. Antle [1982; 1984] and Baiman, Evans, and Noel [1987] expanded this analysis to the more complex setting with private effort choices on the part of both managers and auditors, finding tractable equilibria only with restrictive assumptions regarding auditor independence [Antle, 1984] or on the message space [Baiman, Evans, and Noel, 1987]. This research does not directly test any of these models, since a market environment of repeated price negotiation differs from the contract optimization used in agency models.

3. In the general case, the high cost would maximize the total expected surplus whenever \((p_1 - p_2)(V_H - V_L) > C_H - C_L\), where \(p_1\) (\(p_2\)) is the conditional probability of the low outcome \(V_L\) given the low cost \(C_L\) (high cost \(C_H\)). In this study, \(p_1 = 1\) and \(p_2 = .5\). In addition, the cost and redemption value choices should satisfy two other inequality constraints, necessary to ensure that higher expected gains to trade are associated with the high cost irrespective of the perceived frequency of truthful report representations (details available upon request):

   i) \(C_H - C_L < (V_H - V_L)[(1 - p_2)^2 - (1 - p_1)^2]\)

   ii) \(p_1(1 - p_1) < p_2(1 - p_2)\).

4. Inequality (1) is based on expected net payoffs, and therefore implies risk neutral decision makers. Although no formal adjustment is made for risk aversion on the part of actual experimental subjects, a practical approach is to choose actual parameters such that the inequality in (1) holds by some amount large enough to offset moderate risk aversion. A supplementary analysis of the influence of actual subject risk preferences on observed behavior is presented later in the paper.

5. The assumption that payoffs are ultimately received by owners precludes the possibility of managerial embezzlement of productive revenues.

6. Although audit contracts cannot be made contingent upon the payment of \(T\), it still might be possible to induce \(A\) players to pay \(T\) by making audit fees explicitly contingent upon audit reports. This possibility is beyond the scope of the present research, which is limited to fixed fee audit contracts.

7. For another example of audit failure even in the absence of explicit collusive opportunities between auditors and managers, see example 2.1 in Antle [1982], as commented on by Antle [1984, p. 15].
8. The word "reputation" is used with some caution. Some of the recent literature (e.g., see Young [1985]) suggests that the "reputation" label should be restricted to games where "weak" players emulate the dominant strategies of "strong" players, as in Kreps and Wilson [1982]. In the present research, a "reputation-building" strategy is defined more broadly as any strategy which departs from the dominant single-period strategy in an attempt to capture higher aggregate multi-period payoffs.

9. Only fourteen subjects showed up for market 3A, so only 0 players were used in that market. This still preserved the desired structure of excess demand for M player units, but by a margin of only one player. Fortunately, average prices for M player units were similar in markets 3A and 3B, so it is unlikely that this shortage of one player resulted in any substantial bias.

10. A structure of excess demand is a particularly simple way to provide for positive returns to favorable reputations. For more complex models which generate reputation price premiums even when sellers face perfectly elastic demand schedules, see Shapiro [1983] and Klein and Leffler [1981].

11. There is a conflict between the game theoretic assumption of common knowledge and the precept of privacy usually associated with laboratory markets [Smith, 1976]. That is, strategic games require that players know the payoff functions facing their opponents. However, due to concern over potentially confounding preferences for equity, Smith [1976] and others generally advise against the full public revelation of specific payoff parameter values. In this study a compromise view was adopted, whereby the ordinal structure of different costs and redemption values was publicly revealed, but the specific amounts were kept private. Certain other parameters (e.g. the penalty transfer amount from M to 0 players) were publicly announced. In this manner, an attempt was made to preserve the strategic structure of the game without introducing the concerns underlying Smith's privacy precept.

12. The high cost leads to an expected surplus of $1.60 (=.5($4.00 + $1.00) - $.90), whereas the low cost leads to an expected surplus of only $.70 (= $1.00 - $.30).

13. There would be no reason for an M player to pay the high cost, get a high valued unit, and then report low. Given a desire to report low, an M player could simply pay the low cost, which would ensure a low valued unit.

14. The recision of audit contracts for low reported values stems from the fact that M players could only report "high" or "low." In a more realistic setting of continuous report levels, audit contracts would be enforced over a wide range of possible reports, as intuition suggests. The dichotomous reports of high or low in this study correspond to the "audit region" and "nonaudit region" in the analytic model of Baiman, Evans, and Noel [1987].

15. Even if the A player faces possible forfeiture of his/her fee, there is no single-period incentive to pay the investigation cost T as long as the expected penalty is less than T. In the public signal markets (3A and 3B),
the expected penalty (negotiated investigation fees multiplied by .25) was much less than the investigation cost of $.35, satisfying this requirement.

16. Qualitative judgment is necessary because the statistical evaluation of H1 is trivial. Since the binomial distribution has zero variance at the null frequency of 0 (or 1), any observed instance of behavior counter to the null statistically refutes the null.

17. Thus, this percentage is a measure of market efficiency, typically calculated in laboratory market studies as the percentage of total possible profits actually captured by market participants.

18. The ANOVA views the market itself as the experimental unit of analysis for identifying a treatment effect, resulting in low statistical power but separating treatment effects from extraneous sources of market variance. See Glass and Hopkins [1984, §19.22] for a general discussion of this point.

19. In addition to the separate ANOVAs for prices conditional on high and low reported values, an ANOVA was run on all prices negotiated by M players, with the contingent reported value (high or low) defined as a separate fixed factor. Aside from a highly significant main effect for this factor (indicating that "high" prices were significantly higher than "low" prices), the results paralleled the reported findings.

20. Attempts to identify possible reasons for variations between replications, such as chronological order or relative subject composition (mostly MBAs versus mostly accounting majors), failed to explain observed differences. Nontrivial differences between replications are common in this type of research, making it important to identify only those treatment effects incremental to chance variations between replications.
REFERENCES


Davis, L.H., "No Chain Store Paradox," Theory and Decision (March 1985), pp. 139-44.


____, "Markets as Economizers of Information: Experimental Examination of the 'Hayek Hypothesis'," Economic Inquiry (April 1982a), pp. 165-79.


### Possible Auditor Reports

<table>
<thead>
<tr>
<th>Auditor Pays Investigation Cost</th>
<th>Manager's Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>False</td>
</tr>
<tr>
<td>Yes</td>
<td>(1)</td>
</tr>
<tr>
<td>No</td>
<td>(2)</td>
</tr>
</tbody>
</table>

Types of Audit Reports:

1. Auditor reports that the manager's report is false.
2. Auditor reports that s/he has not found the manager's report to be false.

*Figure 1*
LEGEND: Digits represent the average price by period over two runs of each market type:
1 = No auditor moral hazard (markets 1A and 1B).
2 = Auditor moral hazard with no public signal (markets 2A and 2B).
3 = Auditor moral hazard and a public signal of auditor choices (markets 3A and 3B).

Figure 2: Prices for M Player Units Conditional on a Low Reported Value
LEGEND: Digits represent the average price by period over two runs of each market type:
1 = No auditor moral hazard (markets 1A and 1B).
2 = Auditor moral hazard with no public signal (markets 2A and 2B).
3 = Auditor moral hazard and a public signal of auditor choices (markets 3A and 3B).

Figure 3: Prices for M Player Units Conditional on a High Reported Value
LEGEND: Digits represent the average price by period over two runs of each market type:
1 = No auditor moral hazard (markets 1A and 1B).
2 = Auditor moral hazard with no public signal (markets 2A and 2B).
3 = Auditor moral hazard and a public signal of auditor choices (markets 3A and 3B).

Figure 4: A Player Investigation Fees
### TABLE 1

**Strategic Choice Frequencies**

<table>
<thead>
<tr>
<th>Markets with</th>
<th>Markets with</th>
<th>Markets with</th>
</tr>
</thead>
<tbody>
<tr>
<td>no Auditor</td>
<td>no Auditor</td>
<td>no Auditor</td>
</tr>
<tr>
<td>Moral Hazard</td>
<td>Moral Hazard</td>
<td>Moral Hazard</td>
</tr>
<tr>
<td>1A</td>
<td>1B</td>
<td>2A</td>
</tr>
<tr>
<td>2A</td>
<td>2B</td>
<td>3A</td>
</tr>
<tr>
<td>3B</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Managerial strategies:

<table>
<thead>
<tr>
<th>Category</th>
<th>1A</th>
<th>1B</th>
<th>2A</th>
<th>2B</th>
<th>3A</th>
<th>3B</th>
</tr>
</thead>
<tbody>
<tr>
<td># of high cost choices</td>
<td>49/65</td>
<td>53/65</td>
<td>53/65</td>
<td>52/65</td>
<td>51/65</td>
<td>48/65</td>
</tr>
<tr>
<td></td>
<td>.754</td>
<td>.815</td>
<td>.815</td>
<td>.800</td>
<td>.785</td>
<td>.738</td>
</tr>
<tr>
<td>Total # of cost choices</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># of low reported values</td>
<td>30/43</td>
<td>23/33</td>
<td>24/39</td>
<td>28/42</td>
<td>24/41</td>
<td>9/39</td>
</tr>
<tr>
<td></td>
<td>.698</td>
<td>.697</td>
<td>.615</td>
<td>.667</td>
<td>.585</td>
<td>.231</td>
</tr>
</tbody>
</table>

#### Auditor strategies:

<table>
<thead>
<tr>
<th>Category</th>
<th>1A</th>
<th>1B</th>
<th>2A</th>
<th>2B</th>
<th>3A</th>
<th>3B</th>
</tr>
</thead>
<tbody>
<tr>
<td># of invested cost payments</td>
<td>N/A</td>
<td>N/A</td>
<td>23/26</td>
<td>10/17</td>
<td>16/20</td>
<td>19/28</td>
</tr>
<tr>
<td># of invested cost decisions</td>
<td>N/A</td>
<td>N/A</td>
<td>.885</td>
<td>.588</td>
<td>.800</td>
<td>.679</td>
</tr>
</tbody>
</table>
### TABLE 2

Average Change in M Player Prices Conditional on a High Reported Value from Period t to Period t+1 for M Players Audited in Period t

<table>
<thead>
<tr>
<th>Period t Audit Findings</th>
<th>M Player's Report Supported</th>
<th>M Player's Report Found to Be False</th>
<th>One-Tailed p Value Under Ho: No Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>Price Change in Period t+1</td>
<td>n</td>
<td>Price Change in Period t+1</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-----------------------------</td>
<td>-----------------------------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>All markets</td>
<td>86</td>
<td>$.025</td>
<td>42</td>
</tr>
<tr>
<td>Market 1A</td>
<td>10</td>
<td>-.060</td>
<td>9</td>
</tr>
<tr>
<td>Market 1B</td>
<td>20</td>
<td>.095</td>
<td>7</td>
</tr>
<tr>
<td>Market 2A</td>
<td>14</td>
<td>.007</td>
<td>10</td>
</tr>
<tr>
<td>Market 2B</td>
<td>13</td>
<td>.046</td>
<td>2</td>
</tr>
<tr>
<td>Market 3A</td>
<td>12</td>
<td>.007</td>
<td>5</td>
</tr>
<tr>
<td>Market 3B</td>
<td>17</td>
<td>.006</td>
<td>9</td>
</tr>
</tbody>
</table>
TABLE 3

Average Change in A Player Fees from Period t to Period t+1 for A Players Who Audited an M Player in Period t

<table>
<thead>
<tr>
<th>Period t Audit Findings</th>
<th>M Player's Report Supported</th>
<th>M Player's Report Found to Be False</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Price Change in Period t+1*</td>
<td>Price Change in Period t+1*</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>All markets</td>
<td>86  .017</td>
<td>42  .020</td>
</tr>
<tr>
<td>Market 1A</td>
<td>10  .070</td>
<td>9   .044</td>
</tr>
<tr>
<td>Market 1B</td>
<td>20  .020</td>
<td>7   .057</td>
</tr>
<tr>
<td>Market 2A</td>
<td>14  .009</td>
<td>10  .013</td>
</tr>
<tr>
<td>Market 2B</td>
<td>13  .015</td>
<td>2   .000</td>
</tr>
<tr>
<td>Market 3A</td>
<td>12  -.006</td>
<td>5   .000</td>
</tr>
<tr>
<td>Market 3B</td>
<td>17  .009</td>
<td>9   -.011</td>
</tr>
</tbody>
</table>

* In those few instances where an A player did not negotiate an audit contract in period t+1, the price difference for the next period (t+2) was used instead.
### Table 4

Regression of Price Spreads on M Player Choice Parameters

Regression model:

\[
\text{SPREAD}_i = \phi_1 D_{1A} + \phi_2 D_{1B} + \phi_3 D_{2A} + \phi_4 D_{2B} + \phi_5 D_{3A} + \phi_6 D_{3B} + \phi_7 \text{COST}_i + \phi_8 \text{REPT}_i + \epsilon
\]

<table>
<thead>
<tr>
<th>Coefficient estimates</th>
<th>1.88</th>
<th>1.40</th>
<th>2.11</th>
<th>1.95</th>
<th>2.34</th>
<th>2.25</th>
<th>0.541</th>
<th>-0.285</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-statistic</td>
<td>22.32</td>
<td>17.85</td>
<td>24.64</td>
<td>23.28</td>
<td>26.15</td>
<td>17.77</td>
<td>3.20</td>
<td>-1.99</td>
</tr>
<tr>
<td>Two-tailed p-value</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.004</td>
<td>.059</td>
</tr>
</tbody>
</table>

where: SPREAD\(_i\) is the average difference between the high and low conditional prices negotiated by M player \(i\),

\(D_{1A} - D_{3B}\) are indicator variables set equal to 1 if the observation is from the market indicated and 0 otherwise,

\(\text{COST}_i\) is the percentage frequency of high cost choices by M player \(i\),

\(\text{REPT}_i\) is the conditional percentage frequency of low outcome reports by M player \(i\), given that the observed (true) outcome is low, and \(\epsilon\) is the regression residual.
### TABLE 5
Laboratory Market Prices

<table>
<thead>
<tr>
<th>Markets with no Auditor Moral Hazard</th>
<th>Markets with Auditor Moral Hazard but no Public Signal</th>
<th>Markets with Auditor Moral Hazard and a Public Signal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1A</td>
<td>1B</td>
<td>2A</td>
</tr>
<tr>
<td>R_L: Price for M player units</td>
<td></td>
<td></td>
</tr>
<tr>
<td>conditional on a low reported value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average: Periods 1-13---------------$0.67  $0.90</td>
<td>$0.69  $0.76</td>
<td></td>
</tr>
<tr>
<td>Period 13 only---------------------$0.78  $0.97</td>
<td>$0.72  $0.85</td>
<td></td>
</tr>
<tr>
<td>Low redemption value---------------1.00  1.00</td>
<td>1.00  1.00</td>
<td>1.00  1.00</td>
</tr>
<tr>
<td>R_H: Price for M player units</td>
<td></td>
<td></td>
</tr>
<tr>
<td>conditional on a high reported value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average: Periods 1-13---------------$2.51  $2.27</td>
<td>$2.81  $2.69</td>
<td></td>
</tr>
<tr>
<td>Period 13 only---------------------2.41  2.66</td>
<td>2.71  2.95</td>
<td></td>
</tr>
<tr>
<td>High redemption value--------------4.00  4.00</td>
<td>4.00  4.00</td>
<td>4.00  4.00</td>
</tr>
<tr>
<td>F: A Player investigation fees</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average: Periods 1-13---------------$0.97  $0.87</td>
<td>$0.59  $0.67</td>
<td></td>
</tr>
<tr>
<td>Period 13 only---------------------1.13  1.00</td>
<td>0.73   0.68</td>
<td></td>
</tr>
<tr>
<td>R_H + F combined*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average: Periods 1-13---------------$3.48  $3.14</td>
<td>$3.40  $3.36</td>
<td></td>
</tr>
<tr>
<td>Period 13 only---------------------3.54  3.66</td>
<td>3.44   3.63</td>
<td></td>
</tr>
</tbody>
</table>

* Sum of R_L + F = Average total outlay for a unit reported to be of high value that is covered by an investigation contract.
### Table 6

**ANOVA on Negotiated Prices**

<table>
<thead>
<tr>
<th>Panel A: M player units conditional on a low reported value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Source</strong></td>
</tr>
<tr>
<td>------------</td>
</tr>
<tr>
<td>TRMT</td>
</tr>
<tr>
<td>REP(TRMT)</td>
</tr>
<tr>
<td>PERIOD</td>
</tr>
<tr>
<td>PERIOD × TRMT</td>
</tr>
<tr>
<td>PERIOD × REP(TRMT)</td>
</tr>
<tr>
<td>SELLER(REP(TRMT))</td>
</tr>
<tr>
<td>PERIOD × SELLER(REP(TRMT))</td>
</tr>
<tr>
<td><strong>Total</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: M player units conditional on a high reported value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Source</strong></td>
</tr>
<tr>
<td>------------</td>
</tr>
<tr>
<td>TRMT</td>
</tr>
<tr>
<td>REP(TRMT)</td>
</tr>
<tr>
<td>PERIOD</td>
</tr>
<tr>
<td>PERIOD × TRMT</td>
</tr>
<tr>
<td>PERIOD × REP(TRMT)</td>
</tr>
<tr>
<td>SELLER(REP(TRMT))</td>
</tr>
<tr>
<td>PERIOD × SELLER(REP(TRMT))</td>
</tr>
<tr>
<td><strong>Total</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: A player investigation fees</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Source</strong></td>
</tr>
<tr>
<td>------------</td>
</tr>
<tr>
<td>TRMT</td>
</tr>
<tr>
<td>REP(TRMT)</td>
</tr>
<tr>
<td>PERIOD</td>
</tr>
<tr>
<td>PERIOD × TRMT</td>
</tr>
<tr>
<td>PERIOD × REP(TRMT)</td>
</tr>
<tr>
<td>SELLER(REP(TRMT))</td>
</tr>
<tr>
<td>PERIOD × SELLER(REP(TRMT))</td>
</tr>
<tr>
<td><strong>Total</strong></td>
</tr>
</tbody>
</table>

\(^a\) \(p < .01\).

\(^b\) \(p < .05\).
### TABLE 7

**Pairwise Comparisons Between Treatment Means**

<table>
<thead>
<tr>
<th>Comparison</th>
<th>df</th>
<th>SS</th>
<th>F Value</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRMT 1 vs. TRMT 2 and TRMT 3</td>
<td>1</td>
<td>27.30</td>
<td>17.40</td>
<td>.025</td>
</tr>
<tr>
<td>TRMT 2 vs. TRMT 3</td>
<td>1</td>
<td>10.33</td>
<td>6.59</td>
<td>.083</td>
</tr>
<tr>
<td>Total</td>
<td>2</td>
<td>37.63</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Comparisons for A player investigation fees:

<table>
<thead>
<tr>
<th>Comparison</th>
<th>df</th>
<th>SS</th>
<th>F Value</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRMT 1 vs. TRMT 2 and TRMT 3</td>
<td>1</td>
<td>4.86</td>
<td>9.30</td>
<td>.055</td>
</tr>
<tr>
<td>TRMT 2 vs. TRMT 3</td>
<td>1</td>
<td>0.11</td>
<td>0.21</td>
<td>.678</td>
</tr>
</tbody>
</table>

Imprecision due to unbalanced cell sizes    |       | 0.02↑ |

**Total**                                    | 2  | 4.99 |

↑ Since not all A players negotiated an investigation contract in every period, the ANOVA for investigation fees is mildly unbalanced, creating this small difference in reconciling to the total treatment sum of squares in Table 6.
NINTH SYMPOSIUM
ON AUDITING RESEARCH

October 4-6, 1990

Research Papers

An Analysis of Auditor Litigation Disclosures

by

Zoe-Vonna Palmrose

Discussant: Katherine Schipper

Section 3
An Analysis of Auditor Litigation Disclosures

"Feeling Safe and Secure"
Legislators

Catherine Schipper
"Front or the Market" Thesis
A Reliance on the Efficient Market Theory

Zoe-Vonna Palmrose*
Price Waterhouse Fellow
Associate Professor
School of Accounting
University of Southern California
Los Angeles, California 90089-1421

September 1990

*The financial assistance of the Peat Marwick Foundation Research Opportunities in Auditing Program and the USC Accounting Circle; the research assistance of Joan Cummings, Walt Lipinski, and Lenny Soffer; and the helpful comments of participants in accounting research workshops at several universities in the United States and Canada, particularly Bob Ashton, Mike Duffy, and Wanda Wallace, are gratefully acknowledged.
This study examines disclosures surrounding auditor litigation. The objective is to evaluate the observability of auditors' litigation activities as observability is fundamental to any influence of litigation on perceptions of audit quality and reputations.

The study develops a disclosure spectrum and classifies information sources along the spectrum from general (those disseminating information widely and at minimal cost to market participants), to specialized (those providing information to subsets of the market), to private sources. While commenting on the nature and representativeness of information from all sources, the study focuses on general source disclosures. General sources capture widespread public disclosures cost-effectively observable by all market participants, so that general sources are arguably the most relevant to issues of audit quality and auditor reputations. The Wall Street Journal (WSJ) serves as the representative general source.

Although this study is descriptive, it represents a necessary component for understanding the role of litigation in the audit services market. To elaborate, the legal-economic literature recognizes that legal liability has two primary objectives—compensation and deterrence (e.g., see Shavell [1987]). In the audit services market, legal liability deters auditors from associating with financial statements containing material omissions or misstatements. Liability functions as a deterrence because litigation imposes costs on auditors. Litigation costs consist of both direct and indirect elements. Indirect costs include the effects of litigation on perceptions and assessments of auditor reputations and audit quality.
In turn, these quality assessments or perceptions potentially impact a variety of decisions, such as the pricing of audit services [Palmrose, 1986], selecting and changing auditors [Simunic and Stein, 1987], and settling or trying legal disputes [Priest and Klein, 1984].

The significance of indirect litigation costs, and therefore, the effect of litigation on auditor reputations and audit quality, is an open question. Some preliminary evidence suggests losses in market share and attributes these losses to the reputation effects of litigation, at least for a subset of cases involving SEC enforcement actions against auditors (e.g., see Dopuch and Simunic [1982]; Wilson and Grimlund [1990]). Yet, Schroeder, Solomon, and Vickrey [1986] found both audit committee chairpersons and auditors rated "litigation audit firms had been involved in" as having only slight affect on perceived audit quality. 4

To help reconcile this evidence requires considering how indirect costs occur with litigation. In settings outside accounting and auditing, researchers recognize the importance of observable information in forming perceptions and influencing behavior (e.g., Combs and Slovic [1979]; Lichtenstein, et al. [1978]), including in measuring and valuing attributes of goods and services [Barzel, 1977, 1982]. In an audit setting, DeAngelo [1981] emphasizes observability as necessary for variables to function as quality surrogates. 5

For litigation to affect perceptions and assessments of auditor reputations and audit quality, market participants must be informed of auditors' litigation activities. If only a subset of litigation activities are disclosed to market participants, then it is the readily observable subset with the potential for affecting reputation and quality assessments. This study describes auditors' observable litigation activities.
Because audit quality and reputation measurement serves as a framework for examining litigation disclosures, several, perhaps subtle, issues arise in conjunction with observability. First, measurement is comparative—an audit firm relative to competitors in the market. Second, from a conceptual standpoint, some research (e.g., Palmrose [1988]; Simuni and Stein [1987]) supports relevant comparisons encompass audit firms' overall levels of litigation (litigation occurrences) and overall levels of meritorious litigation. Palmrose discusses the need for final resolution data to distinguish meritorious from non-meritorious legal actions alleging audit failures. Based on these issues, this examination of litigation disclosures will include assessing both observable litigation activity and observable litigation resolutions relative to "true" underlying levels.

The remainder of the study is organized as follows. The next section explains the disclosure spectrum. The third section describes the sample selection criteria and estimates the portion of overall audit litigation captured by the sample. Subsequent sections overview litigation disclosures and analyze WSJ disclosures. The final section contains summary and conclusion remarks.
DISCLOSURE SPECTRUM

A disclosure spectrum exists because no source or sources systematically capture and report all auditor litigation information to market participants. Private information denotes one end of the spectrum, wide-spread public information the other end, and specialized information sources fall in between.

Private information consists of unpublished or unreported auditor litigation activities. Private includes litigation information both unobservable and not cost-effectively observable by market participants. Several factors explain why litigation information remains private.

As partnerships or professional corporations, audit firms generally decline to voluntarily provide litigation information and few disclosure requirements apply to audit firms. Members of the AICPA-SECPS Division of Firms must report occurrences of "litigation against the firm or its personnel and proceedings or investigations publicly announced by regulatory agencies alleging deficiencies in the conduct of an audit of the financial statements" of a present or former SEC registrant or client of public interest [AICPA, 1986, p. 2], however, the Division treats reported information as confidential. Occasionally special reporting to regulators occurs (e.g., the Resolution Trust Corporation [JofA, 1990, p. 19] and the SEC [Business Week, 1985, p. 34]).

Much case-specific information, although available at a court, remains private because considerable barriers exist to acquiring and utilizing such information; barriers include the large number of federal and state courts, the diversity in court locations, and the volume of court information.
Courts also render relevant information unobservable by sealing, restricting, or protecting all or portions of legal records. Furthermore, some relevant information remains outside court records. Resolution information illustrates both points. Resolution information need not appear in court records as courts can dismiss actions (with prejudice) without revealing the details (including payment amounts) of private agreements among litigating parties. As a condition of settlement, litigants can request courts preclude all parties from ever revealing information by placing it under protective order. Even in class actions requiring court approval of reasonableness, courts and litigants (including plaintiffs' counsel) can be unaware of individual defendant contributions, if any, to settlement funds.

Specialized and general sources disseminate auditor litigation information. Specialized sources furnish information to subsets of market participants. These sources need not be considered unavailable to all market participants. Rather, specialized sources involve limited circulation and/or specialized skills, which create accessing barriers or costs, thereby inhibiting the observability of specialized source information by the market as a whole. Specialized sources consist of law and accounting reporters and legal and accounting periodicals that provide information to professional groups such as lawyers and accountants. Specialized sources also furnish information to subsets of market participants based on geographical location (through regional press disclosures) and organizational affiliations (through client-specific financial reporting).
General sources furnish information to all market participants at minimal cost. Being readily (cost-effectively) observable by all market participants, general sources potentially represent the most influential sources of aggregate market information (i.e., information on overall litigation activities in the audit services market) and the most applicable when comparing litigation among auditors and making relative assessments of audit quality and auditor reputations.

General sources include the national business press. This study uses the WSJ as the representative general source. In part, the WSJ's selection derives from its role in publishing significant corporate and business events [Thompson, Olsen, and Dietrich, 1987]. Also, the WSJ is the U.S. newspaper with the largest circulation [USA Today, 1989, p. 4B]. The study reviews auditor litigation disclosures from a number of additional general sources to supplement the WSJ.

In summary, the disclosure spectrum—private, specialized, and general sources—provides a framework for examining litigation observability. The next section identifies the private, specialized, and general sources used to obtain the sample observations, explains the sampling criteria, and provides some insights on the portion of overall auditor litigation captured by the sample.
SAMPLE

The study uses a sample of 761 cases related to independent audit services rendered in the U.S. market during 1960 through 1990. The cases involve the services of the 15 or so largest audit firms. Table 1 lists the firms. The 761 cases were identified primarily from general, specialized, and private sources summarized in Table 2.

Insert Table 1

Insert Table 2

Table 3 classifies the 761 cases by source(s). Note from Table 3 that 95 of the 761 cases (13 percent) are from private sources. Of course this understates the true population rate for private, unobservable, cases. I can furnish some limited information on this true rate which, in turn, produces an estimate of the overall level of auditor litigation.

Insert Table 3

First, from November 1, 1979 through June 30, 1990, 380 cases were reported to the AICPA-SECPS SIC/CQIC. The sample of 761 cases includes 214 reported cases. Of the 214 reported cases contained in the sample, 175 are from general and/or specialized sources. The remaining 39 reported cases in the sample together with the 166 (380 less 214) reported cases not captured by the sample, yield 205 (39 plus 166) private unobservable cases. This translates into an unobservable litigation rate of 54 percent (205 of 380). This rate would be a maximum estimated rate since the AICPA-SECPS has over 400 (average) member audit firms and some unknown number of reported cases not in this sample relate to member audit firms not in Table 1.
Second, data from U.S. District Courts in Denver, Los Angeles, and Seattle allow estimation of unobservable litigation rates. Comparing cases first detected by visiting the court with the total number of sample observations filed at the court, yielded unobservable litigation rates from 35 to 45 percent.

These unobservable litigation rate estimates ranging from 35 to 54 percent could understate the extent of private cases as the estimates do not incorporate data from state court filings. Therefore, translating these rates into an estimate that the sample of 761 cases encompasses about 50 percent of the overall audit-related litigation for firms in Table 1 might be optimistic rather than conservative. Nonetheless, later discussions use this 50 percent estimate to give an overall perspective to litigation disclosures. The next section overviews disclosures before discussing WSJ disclosures.

OVERVIEW OF LITIGATION DISCLOSURES

Table 3 shows that 44 percent (338 of 761) of the sample cases are found in general sources either alone or in combination with specialized sources; 43 percent (328 of 761) are from one or more specialized sources. Before focusing on wide-spread public disclosures via the WSJ, a few comments are relevant on specialized and other general source litigation information.

Although not obvious from Table 3, nearly 80 percent (262 of 328) of the cases in specialized sources are from legal sources, either alone or in combination with accounting sources. The nature of information in legal sources varies greatly and much appears not readily usable for the measurement of audit quality and reputations. Because of an orientation toward legal not factual matters, legal sources cannot be relied on to identify the auditor, client, and type of service at issue or to provide
background information on the nature of alleged audit failures. Perhaps the most consistently comprehensive legal source information occurs in the SEC ASR's and AAER's. Note, from a representativeness perspective, Loebbecke and Willingham [1988] conclude the SEC pronouncements are strongly biased in the direction of the SEC's charges against those involved in the cases.

Client company financial reports represent a unique disclosure source for 30 of the 761 cases. While not shown in Table 3, financial reports were located for 173 sample companies (this comprises 33 percent (173 of 519) of the public companies in the sample). Of the 173 company reports, 122 contained some reference to auditor litigation and 51 had no information. The 122 company reports with auditor litigation information did not always mention the auditor by name. For example, some reports stated that litigation involved former auditors. Most disclosures mentioned the existence of auditor litigation. Only a few instances of financial reports with auditor resolution information occurred in the sample of 122.

Based on the somewhat limited review of regional press disclosures described in Table 2, the regional press does not appear as a major unique source. Note in Table 3, only 4 of the 761 cases (1 percent) have regional press as the only disclosure source. Nevertheless, this does not imply that any direct and indirect costs of litigation are not significant in regional markets [Wilson and Grimald, 1990].

Table 3 reveals 18 cases with general source disclosures other than the WSJ. About half of these cases are from PR Newswire; the rest involve recent legal actions related to savings and loans. All 18 disclose the existence of auditor litigation not resolutions of actions.
ANALYSIS OF WSJ DISCLOSURES

This section examines general disclosures of auditor litigation as reflected in the WSJ as the representative general source. A WSJ disclosure case required having at least one article in the WSJ mentioning litigation involving the audit firm. If an article discussed more than one legal case, the article counted as a disclosure for each case.

For the 320 cases with WSJ disclosures, Table 4 provides a classification based on the number of disclosures. The 320 cases yield 913 WSJ disclosures. As shown in the table, 43 percent of the cases (138 of 320) have just one WSJ disclosure and 16 percent of the cases (52 of 320) have five or more.

The 52 cases with five or more WSJ disclosures represent 7 percent of the total sample of 761 cases. If estimates of unobservable cases and the overall level of litigation are reasonable, these 52 cases represent 3 to 4 percent of overall auditor litigation. Yet, these 52 cases account for 48 percent (439 of 913) of the WSJ disclosures as shown in Table 4. Because of the disproportional level of disclosure surrounding these 52 cases, these cases are referred to as "highly visible" and the remaining analyses separately present these cases. A later section contains an extended discussion of these highly visible cases.

Insert Table 4

Characteristics of WSJ Disclosure Cases

Table 5 summarizes the number of cases with and without WSJ disclosures for a variety of client, case, environment, and auditor-related factors. These potentially relevant factors were identified
(inferred) from the auditor litigation literature (e.g., see Palmrose [1987, 1988]; St. Pierre and Anderson [1984]) and the corporate disclosure literature (e.g., see Gibbins, Richardson, and Waterhouse [1990]; Thompson, Olsen, and Dietrich [1987]).

Insert Table 5

Client ownership, public or non-public, is not independent of WSJ disclosure status (yes or no). As shown in Table 5, 253 out of 519 (49%) of the public company cases have WSJ disclosures, while only 67 out of 242 (28%) of the non-public client cases do. The association between ownership and disclosure is statistically significant at less than the .01 level (based on a 2 x 2 contingency table and a computed chi-square statistic of 30.45).

Because of the unavailability of client size data for many cases, public company exchange is used to proxy for size. Even so, the exchange could not be determined for all client companies and these cases are included in Table 5 as unknown. While NYSE clients' WSJ disclosure rates are greater than rates for AMSE and OTC clients (54%, 51%, and 44%, respectively), the association between exchange and disclosure is not statistically significant as noted in Table 5.

Research suggests the relevance of severe financial distress (including bankruptcy) and fraud to auditor litigation [Palmrose, 1987; St. Pierre and Anderson, 1984]. Table 5 presents the number of cases with and without WSJ disclosures in one of four categories: financial distress only, fraud only, both financial distress and fraud, and neither financial distress nor fraud. The percentage of cases with WSJ disclosures for each of
these categories is 39 percent (90 out of 233), 40 percent (50 out of 125), 69 percent (131 out of 190), and 23 percent (49 out of 213), respectively. The presence of both financial distress and fraud is associated with WSJ disclosure frequencies greater than expected and the absence of both is associated with lesser disclosure frequencies (statistically significant at less than the .01 level).

In terms of case-related factors, all cases in the sample with criminal actions against auditors have WSJ disclosures; in fact, all are highly visible cases. The vast majority of cases with SEC enforcement actions against auditors also have WSJ disclosures (51 out of 55 or 93%). A later section discusses case resolutions and disclosure.

Table 5 provides case and disclosure data classified by time-period and auditor. The 1960-1990 time-period is subdivided into five groupings of years and cases are assigned to groupings based on the reporting period of the questioned financial information. If the financial information spanned more than one period, cases are classified according to the latest period. Neither characteristic, time-period nor auditor, is significantly associated with WSJ disclosure. Note, these data do not examine time-period and auditor characteristics from the standpoint of volume of disclosures. The discussion of highly visible cases provides some insights regarding these characteristics and disclosure volumes.

To summarize, this univariate analysis suggests that client ownership (public companies), client condition (both fraud and financial distress), and SEC enforcement actions against auditors are associated with WSJ disclosure. The next section explores the nature of auditor litigation information disclosed in the WSJ.
Classification of WSJ Disclosures by Litigation Stage(s)

To help understand the nature of litigation information widely disclosed to market participants, WSJ disclosures are classified into four stages intended to capture the primary stages in the litigation process, from pre-litigation to resolution, as follows:

Stage 1: Pre-Litigation

Allegations and inferences of audit failure in the context of auditor changes, auditor and client disagreements, financial restatements, regulatory investigations of financial reporting, etc.

Stage 2: Legal Actions

Initiation of legal actions against auditors where disclosure occurs simultaneously with or subsequent to the legal action. If the latter, existence of legal actions against auditors can be disclosed in the context of auditor changes; client operational financial reporting, or regulatory problems; audit firm problems; professional problems or issues, etc.

Stage 3: Resolution Process

Information on lower, appeal, and supreme courts' activities (including court decisions that involve preliminary findings but not final resolutions). Also, information on plaintiff and defendant negotiations and negotiations with regulatory agencies.

Stage 4: Final Resolution

Information on final resolution of actions (both court imposed verdicts and out-of-court settlements); settlements to avoid litigation; and final determinations by regulatory agencies.

Classifying the 913 WSJ disclosures by stage revealed that 11 percent related to stage 1, 46 percent to stage 2, 21 percent to stage 3, and 22 percent to stage 4.

Table 6 classifies the 320 cases with WSJ disclosures according to the litigation stage(s) information in the WSJ. The table separates highly visible and other cases.
Insert Table 6

Data in Table 6 using number of cases present a message consistent with that from classifying number of WSJ disclosures by stage. A relatively small number of cases have any pre-litigation (stage one) disclosures, only about 57 out of 320 (18%). Of these, 24 involve highly visible cases and 33 involve other cases, which represent 46 percent (24 of 52) and 12 percent (33 of 268) of their respective groups.

Focusing on other cases, 43 percent (115 of 268) have stage two disclosure(s) on the existence of litigation but no disclosures for other stages. Although not shown in Table 6, most of the stage two disclosures for other cases occur about the time of initiating legal actions against the auditor; for the majority, this represents the only WSJ disclosure.

Not surprisingly, given the definition of highly visible, the highly visible cases tend to have disclosures during most stages. Thirty-five of 52 (67%) have information in the WSJ on three or more stages, while only 18 of the 268 (7%) other cases have such disclosures.

Final resolution (stage four) data are widely disclosed to market participants for the majority of highly visible cases (34 of 52); but resolution disclosures are fewer for the other cases. At least 12 of the 18 highly visible cases lacking final resolution disclosures are unresolved.

To put this data in some perspective, the overall sample of cases breakdown is as follows:

<table>
<thead>
<tr>
<th>Description</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>No WSJ Disclosure</td>
<td>441</td>
<td>(58%)</td>
</tr>
<tr>
<td>WSJ Disclosure of Litigation Existence and/or Preliminary Resolution Information</td>
<td>182</td>
<td>(24%)</td>
</tr>
<tr>
<td>WSJ Disclosures that include Final Resolution Information</td>
<td>138</td>
<td>(18%)</td>
</tr>
<tr>
<td>Total</td>
<td>761</td>
<td>(100%)</td>
</tr>
</tbody>
</table>
Adjusting these rates for unobservable litigation not captured by this sample (i.e., assuming sample captures about 50 percent of overall auditor litigation for firms in Table 1), would result in rates of 79 percent, 12 percent, and 9 percent in each respective category (none, existence, resolution). The next section elaborates on the nature of resolution disclosures.

WSJ Resolution Disclosures

Table 7 provides additional data on the nature of resolution information widely available to market participants. Table 7 summarizes the type of resolution disclosed in the WSJ for the 138 cases with such disclosures. The table presents separately data for highly visible and other cases. Disclosed final resolutions are categorized as follows (1) auditor dismissed and/or no auditor payments, (2) auditor payments of less than $1 million, (3) auditor payments of $1 million or more, (4) auditor payments but amount not disclosed in the WSJ, and (5) regulatory actions only—usually by the SEC. While category (5) consists of cases where the only final resolution disclosed in the WSJ involves regulatory actions against auditors, some cases in categories (2), (3), and (4) have regulatory enforcement actions disclosed in the WSJ. All but one of the regulatory resolutions included in category (5) involve sanctions (consent decrees) against auditors. In addition to number of cases, Table 7 also presents number of WSJ disclosures in each of the five categories.

Insert Table 7

Palmrose [1988] provides support for using litigation outcomes to establish the merits of allegations of audit failure, so that resolution
data supplement litigation occurrence data for assessing auditor quality and reputations. One proposed framework uses auditor payments and/or regulatory sanctions to identify meritorious allegations of audit failure and auditor dismissals and/or no payments to identify non-meritorious allegations. Applying this framework to observable litigation resolutions in the WSJ summarized in Table 7, shows that for all but 14 of the 138 cases (13 dismissals plus the one unsuccessful SEC action) resolution disclosures confirm allegations of audit failure. This means that 90 percent (124 of 138) of the resolutions readily observable by market participants support audit failure occurrences.

Moving from number of cases to number of disclosures, Table 7 also shows that some reinforcement of resolutions supporting allegations of audit failure occurs through additional resolution disclosures regarding the same case. This reinforcement does not occur for auditor dismissals/no auditor payments as the number of WSJ disclosures (13) equals the number of cases (13).

Further insights on the representativeness of WSJ resolution information are available from comparing WSJ information from Table 7 to resolution information for the overall sample from all sources. I have resolution information for 348 of the 761 sample observations (46%) as follows: 19

<table>
<thead>
<tr>
<th>Description</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auditor Dismissed or No Payments</td>
<td>133</td>
<td>38%</td>
</tr>
<tr>
<td>Auditor Payments of Less than $1 Million</td>
<td>105</td>
<td>30%</td>
</tr>
<tr>
<td>Auditor Payments of $1 Million or More</td>
<td>85</td>
<td>24%</td>
</tr>
<tr>
<td>Regulatory Actions Only Available Final Resolution: Sanctions Imposed</td>
<td>23</td>
<td>7%</td>
</tr>
<tr>
<td>Regulatory Actions Only Available Final Resolution: No Sanctions</td>
<td>2</td>
<td>1%</td>
</tr>
<tr>
<td>Total</td>
<td>348</td>
<td>100%</td>
</tr>
</tbody>
</table>
Classifying resolutions into two categories, (1) payments or sanctions identifying audit failures and (2) dismissals, no payments, or no sanctions, and comparing WSJ disclosure (yes or no) reveals greater than expected frequencies for resolution disclosures identifying audit failures (computed chi-square of 67.38, significant at less than the .005 level).

Before concluding this analysis of WSJ resolution disclosures, several additional issues arise related to the measurement of auditor quality and reputations. Table 6 provides some data on stage 3 disclosures, i.e., information on intermediate events in the resolution process. Some of these stage 3 disclosures are problematic because at the time of disclosure it is not obvious the information is preliminary. For example, disclosed trial court decisions are later overturned on appeal. Long delays between preliminary and final resolution results and the possibility final resolutions lack disclosure compound the difficulties. A series of events for one observation, Auto-Train an AG/GT client (having two WSJ auditor litigation disclosures--stage 3* and stage 4**), illustrates these issues:

1978 Company announces SEC investigation
1978 and 1979 Audits with qualified opinions (not in WSJ)
1980 Company announces SEC concluded private investigation without taking any action
1980 Company declares bankruptcy
1983 Bankruptcy trustee files suit against auditor
1985* Jury decision, judgment against auditor for $11 million, WSJ states auditor will request judge reverse decision or will appeal
1990** Federal appeals court overturns trial court decision and throws out $11 million judgment against auditor

Note, the periods of time among the providing of services, filing of a legal action, and the rendering of initial and final decisions on the action; also, note the total elapsed time exceeds ten years.
To summarize the evidence in this section, observable final resolution information primarily confirms allegations of audit failure via disclosures of auditor payments. This observable information overstates the proportion of meritorious allegations in the overall sample. Other problems with resolution data arise because of the length of time between filing and resolution and the uncertainties as to whether disclosed information represents final resolutions.

**Highly Visible WSJ Disclosure Cases**

Table 8 lists the highly visible cases—the cases with five or more WSJ disclosures. Table 8 gives the identities of auditors and clients, and, the number and time-frame of WSJ disclosures.

Insert Table 8

For the cases listed in Table 8, disclosure periods vary from 1-20 years, with the mean being 6.2 years and the median being 5 years. The distribution is as follows:

<table>
<thead>
<tr>
<th>Duration</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 years or less</td>
<td>14</td>
<td>(27%)</td>
</tr>
<tr>
<td>4-6 years</td>
<td>25</td>
<td>(48%)</td>
</tr>
<tr>
<td>7-9 years</td>
<td>5</td>
<td>(10%)</td>
</tr>
<tr>
<td>10 or more years</td>
<td>8</td>
<td>(15%)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>52</strong></td>
<td></td>
</tr>
</tbody>
</table>

For many of the cases, the time spans reflect time-frames for case resolutions. For other cases, WSJ discussions of auditor litigation or SEC enforcement activities include references to prior events (e.g., see five PMM highly visible cases from the early 1970's, incorporated in one SEC ASR, that have references in the late 1980's).

I examined the highly visible cases for characteristics commonly associated with auditor litigation (e.g., see Table 5). Almost all cases
involve management fraud and/or severe financial distress, 87 percent (45 of 52) and 89 percent (46 of 52), respectively.

Regulatory characteristics figure prominently in highly visible cases. The majority involve clients in regulated industries (including government contractors and financial services) (28 of 52 or 54%). Forty-five of the highly visible cases involve public companies. Eighteen of the cases (35%) include attempted SEC enforcement actions against audit firms or firm personnel, with all but two being successful (one unsuccessful and one in-progress) (Table 5 notes only successful actions). Furthermore, six of the cases have criminal charges against audit firm personnel, four of which were successful. These six cases constitute the only criminal audit-related cases that I am aware of for the audit firms in Table 1.

One final descriptive comment on highly visible cases relates to the litigation stages data in Table 6. Cases achieve highly visible status through three disclosure routes: (1) a clustering of disclosures with the occurrence of litigation (stages 1 and 2), (2) a continuous series of disclosures as litigation occurs and progresses (throughout all stages), and (3) a clustering of disclosures during the resolution process (stages 3 and 4). Route (1) characterizes some recent highly visible cases with highly visible status occurring quickly.

SUMMARY AND CONCLUDING REMARKS

In summary, this study described the nature of auditor litigation disclosures for a sample of 761 cases from 1960-1990 for the largest audit firms. The study's objective included providing insights on the observability of auditor litigation from the perspective of market
participants, since observability constitutes an important condition in the measurement of auditor quality and reputations. Observability was examined using a disclosure spectrum, whereby information sources were classified along that spectrum as general, specialized, or private sources. Being widely and cost-effectively observable to market participants, general sources became the focus for much of the study. The WSJ served as the representative general source. The analysis revealed insights including the following:

- Using data on private litigation, it was estimated the sample encompasses about half of the overall audit-related litigation for the firms in the study. Given less than half of the sample cases had one or more WSJ disclosures, this means only about 20 percent of auditor litigation cases have any widely disclosed information.

- For the portion of the sample cases with WSJ disclosures mentioning some aspect of auditor litigation, about half had just one article, usually on the existence of legal actions against auditors around the time of filing.

- Public ownership and fraud and financial distress were client characteristics associated with greater frequencies of cases with WSJ disclosures.

- A small number of cases (52) accounted for nearly half of the total WSJ disclosures. These cases were termed highly visible cases. Both management fraud and severe financial distress characterized highly visible cases. In addition, SEC actions against auditors and regulated industry affiliations were prominent features of these clients. Highly visible case disclosures occurred over relatively long time periods, the median being 5 years.

- The availability of final resolution information in general sources was limited for most cases except the highly visible cases. Most (90%) of the final resolutions disclosed in the WSJ confirmed allegations of audit failure.

- In addition to representativeness, the length of time between rendering services and resolving litigation and ambiguities related to whether information represented final resolutions were identified as potential issues regarding litigation resolutions and quality and reputation measurement.
This study documenting observable auditor litigation activities raises a number of research questions—both normative and descriptive—related to litigation and audit quality and reputations. From a normative standpoint, given the issues regarding the representativeness of observable litigation, should litigation be used in quality and reputation measurement? If, indeed, measurement requires resolution data, then the results undermine a strong form role for litigation in measurement, i.e., litigation as a surrogate or proxy for quality [Palmrose, 1988].

On the other hand, a number of questions exist as to whether market participants use litigation information. If so, how is information from pre-litigation through resolution used? For example, researchers found biases in newspaper coverage of lethal events "correlated closely" to biases found in research asking people to judge the frequency of the same causes of death [Combs and Slovic, 1979, p. 843]. Thus, in spite of any unrepresentativeness, highly visible cases and resolutions confirming audit failures may affect perceptions and assessments of audit quality and reputations. In fact, market participants may infer litigation occurrences represent meritorious allegations because observed resolutions primarily confirm such allegations.

Perhaps litigation can be viewed as consisting of two broad categories—routine and non-routine, as largely reflected in the other and highly visible case categories in the study. The Schroeder, Solomon, and Vickrey [1986] study results may reflect a lack (or minimal) influence for most routine litigation, while the Dopuch and Simunic [1982] and Wilson and Grimald [1990] studies' results reflect the influence of the non-routine.
The issues raised in the study have implications for more than just audit quality and reputation measurement. For example, market participants' ability to assess the risks of audit failure, and biases in these assessments, affect the role of liability rules in deterrence (see Shavell [1987]).

Priest and Klein [1984] discuss the role of expected reputation effects in the resolution of individual legal disputes. Past disclosure practices, along with current disclosure status and prospects for the case, can influence auditors' assessments of reputation effects during the resolution of specific legal actions. In turn, reputation effects, or lack thereof, influence the nature of disputes tried to verdict. The study results suggest that indirect litigation costs (reputation effects) may be insignificant for non-public, unregulated clients without both fraud and financial distress. (An example of this situation is auditor litigation regarding failure to detect employee defalcations for a non-public, unregulated client.) If so, using Priest and Klein's framework, under ceteris paribus conditions this leads to a prediction of different auditor trial success rates for these cases as compared to other cases.

These represent some of the issues raised by the study. While the study seems to raise more questions than it answers, this is a strength if documenting auditor litigation disclosures helps to address the right questions.
FOOTNOTES

1 Market participants are defined as users of audited financial information engaged in voluntary contracting. Market participants include current and potential investors and creditors, advisers thereof (such as analysts and underwriters), boards of directors, and management. This study concentrates on users in the U.S. audit services market.

2 Direct costs include penalties, judgments, and settlements not covered by insurance; defense expenses (e.g., fees and salaries for lawyers, experts, and firm personnel; and travel, supplies, and court costs); and insurance cost increases.

3 Audit quality can be defined in terms of the level of assurances -- the probability financial statements contain no material omissions or misstatements. Higher levels of assurance (with lower variances) are consistent with higher quality services. This definition is compatible with DeAngelo [1981], the professional literature, and the literature on audit attributes. The professional literature describes audit quality in terms of audit risk, with higher quality services reflecting lower audit risk. Using an attribute approach [Dopuch and Simunic, 1982; Palmrose, 1982; Simunic and Stein, 1987], level of assurances (credibility) represents one attribute and auditor reputations (brand name) another. Under the attribute approach, audit services with greater amounts of all attributes are unambiguously defined as higher quality. Finally, both the professional and research literature [Palmrose, 1988] recognize client acceptance/continuation decisions are endogenous to decisions on audit quality.

4 Anecdotal evidence suggests some clients monitor their auditors' involvement in litigation and regulatory difficulties, some clients monitor these activities for all major audit firms [Macchiavenna, 1981].

5 DeAngelo [1981] discusses that quality surrogates need to be both observable and correlated with underlying audit quality. Palmrose [1988] examines the correlation between audit quality and auditor litigation, however, correlation and observability issues overlap, so that portions of this study also relate to correlation issues.

6 Legal cases can be classified into three types, (1) cases not catalogued in law reporters and legal data bases such as LEXIS, (2) cases catalogued without any explanation or information on court rulings, and (3) cases catalogued with written court rulings. Law reporters and legal data bases such as LEXIS typically contain (2) and (3) but refer to (2) as unreported, unpublished, or unwritten cases. In this study, these terms apply to (1). As an aside, many written rulings respond to motions attempting to get judges to clarify trial issues or procedures, or resolve specific legal questions before trial or settlement (e.g., class action status) [Shanley and Peterson, 1987, p. 5]. Criteria for publishing court rulings differ among jurisdictions, but usually include rulings expected to establish new law or to be appealed.
Member firms must include in their annual report to the SECPS a "disclosure regarding pending litigation as required under GAAP and indicating whether such pending litigation is expected to have a material effect on the firm's financial condition or its ability to serve clients" [POB, 1984, p. 31].

Member firms report to the Quality Control Inquiry Committee (QIC)—formerly the Special Investigations Committee (SIC). The Committee evaluates the existence of and compliance with appropriate quality control standards and the adequacy of applicable auditing standards. The Committee does not form an opinion about the performance of any audit, adjudicate, or punish audit firms or auditors [Hall and Renner, 1988].

This represents an expanded sample from Palmrose [1988]; however, similar criteria were followed for specifying cases. A client and audit firm combination counts as one case. Therefore, multiple actions (criminal and civil; actions by investors, creditors, regulators, and bankruptcy trustees) regarding similar services and issues count as one case. Nearly all cases involve lawsuits filed in federal or state courts. The few remaining cases involve regulatory actions or auditor payments to avoid litigation over allegations of audit failure. Some actions include more than one audit firm as defendants, usually because of related companies with different auditors or auditor changes. Since actions occurred against all auditors, these situations count as a case for each firm.

Legal actions in the sample involving the merged firms of Ernst & Young and Deloitte & Touche also involve the client's pre-merger auditor and are classified with the latter. Post-merger litigation disclosures in 1989 and 1990 frequently identify the pre-merger audit firm.

Work with the New York Court data is too preliminary to use for rate estimation.

Information was not consistent across the types of client reports, so that no primary financial report source emerges. For example, financial statement footnotes might be devoid of references to auditor litigation, while 10-K's for the same period contained references; or, both 10-K's and financial statement footnotes might omit auditor litigation information, while proxy statements contained such information (usually in discussing auditor reappointment).

Although no consideration is given to length or location of articles, most feature article disclosures involve the cases classified as highly visible. Exceptions include recent coverage of some audit firms' litigation related to savings and loans and LH's litigation focused primarily on PTL.
14. The definition of a WSJ disclosure does not consider articles that discuss clients' operational, reporting, or regulatory difficulties without mentioning the auditor. This is not necessarily a trivial exclusion. It can be argued that auditors need not be mentioned by name to impact perceptions and assessments of audit quality and reputations, as long as market participants connect appropriate audit firms with alleged financial reporting problems. Because I reviewed the client company WSJ disclosures, I can provide some data on this issue. About 30 cases might fall into the category of meaningful company disclosures but no auditor litigation disclosures. This would increase slightly the percentage of cases with WSJ disclosures. While this considers a lack of WSJ auditor litigation disclosures for companies with auditor litigation, it fails to consider the number of clients with meaningful financial reporting problems but no auditor litigation. Reviewing the WSJ over a four-year period (1985-1989), I identified over 200 companies that appeared to have financial reporting difficulties for which further investigation revealed no auditor litigation. In contrast, auditor litigation disclosures are not an unambiguous signal of audit failure; however, the lack of any disclosures has additional difficulties.

15. Many auditor litigation cases also involve client companies, officers and directors, and related parties (e.g., underwriters) as defendants.

16. Non-public includes broker-dealers and limited partnerships that may have some SEC reporting requirements.

17. Fraud includes management fraud, irregularities and illegal acts by employees, and defalcations (diversions of assets). For the sample cases the presence of fraud was defined by: (1) SEC or other regulatory agency enforcement actions against the company, management, and/or independent auditors regarding financial reporting issues; (2) criminal indictments or convictions of management, employees, and/or auditors for defalcation or fraudulent financial reporting; or (3) internal investigations by the company reporting management fraud, other irregularities and illegal acts or defalcations.

18. The neither category consists of cases with no indication of financial distress or fraud from the information available through sources listed in Table 2. As great differences in the amount of useful information occur among sample cases, some misclassifications are likely.

19. As a cautionary note, the underrepresentation of private cases in the sample likely means an underrepresentation of cases with auditor dismissals or no auditor payments, and, perhaps of cases with smaller payments (less than $1 million).
Specialized sources have similar problems. Even legal sources typically require additional work to determine if information is preliminary or final. For example, court decisions dismissing auditors or granting summary judgment for auditors can be appealed, can be refiled via amended complaints, can represent some but not all claims (e.g., RICO but not negligence claims), and can involve one but not all actions related to the audit client (e.g., dismissal of creditor action but another action maintained with investors as plaintiffs).
REFERENCES

American Institute of Certified Public Accountants Division for CPA Firms, Report on the Activities of the Special Investigations Committee of the SEC Practice Section of the AICPA Division for CPA Firms (AICPA, 1986).


Business Week, "Auditors Own Up to the Cost of Their Errors," (April 1, 1985) p. 34.


Causey, D.Y., Jr., Duties and Liabilities of Public Accountants (Dow Jones-Irwin, 1982).


Gormley, R.J., "Developments in Accountants' Liability to Nonclients for Negligence," Journal of Accounting Auditing & Finance (Summer 1988), pp. 185-212.


USA Today, "Newspaper Circulations on the Rise," (November 9, 1989), p. 4B.

<table>
<thead>
<tr>
<th>(AA)</th>
<th>Arthur Andersen &amp; Co.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(AY)</td>
<td>Arthur Young &amp; Co. ⁹</td>
</tr>
<tr>
<td>(CL)</td>
<td>Coopers &amp; Lybrand (Lybrand, Ross Bros. &amp; Montgomery)</td>
</tr>
<tr>
<td>(DHS)</td>
<td>Deloitte Haskins &amp; Sells (Haskins &amp; Sells) ⁸</td>
</tr>
<tr>
<td>(EW)</td>
<td>Ernst &amp; Whinney (Ernst and Ernst) ⁹</td>
</tr>
<tr>
<td>(PMM)</td>
<td>Peat Marwick Main &amp; Co. (Peat, Marwick, Mitchell &amp; Co.)</td>
</tr>
<tr>
<td>(PW)</td>
<td>Price Waterhouse</td>
</tr>
<tr>
<td>(TR)</td>
<td>Touche Ross &amp; Co. (Touche, Ross, Bailey &amp; Smart) ⁸</td>
</tr>
<tr>
<td>(GT)</td>
<td>Grant Thornton</td>
</tr>
<tr>
<td>(AG)</td>
<td>Alexander Grant &amp; Company</td>
</tr>
<tr>
<td>(Fox)</td>
<td>Fox &amp; Company</td>
</tr>
<tr>
<td></td>
<td>Including: Elmer Fox &amp; Co.</td>
</tr>
<tr>
<td></td>
<td>Westheimer, Fine, Berger &amp; Co.</td>
</tr>
<tr>
<td></td>
<td>Elmer Fox, Westheimer &amp; Co.</td>
</tr>
<tr>
<td>(MH)</td>
<td>KMG Main Hurdman ⁶</td>
</tr>
<tr>
<td></td>
<td>Including: Main Lafrentz &amp; Co.</td>
</tr>
<tr>
<td></td>
<td>Hurdman &amp; Cranstoun</td>
</tr>
<tr>
<td></td>
<td>Hurdman &amp; Cranstoun, Penney</td>
</tr>
<tr>
<td></td>
<td>Main, Hurdman &amp; Cranstoun</td>
</tr>
<tr>
<td></td>
<td>L.H. Penney &amp; Co.</td>
</tr>
<tr>
<td></td>
<td>John F. Forbes &amp; Co.</td>
</tr>
<tr>
<td>(LH)</td>
<td>Laventhal &amp; Horwath (Laventhal, Krekstein, Horwath, and Horwath)</td>
</tr>
<tr>
<td>(JKL)</td>
<td>J.K. Lasser &amp; Co. ⁶</td>
</tr>
<tr>
<td>(SDL)</td>
<td>S.D. Leidesdorf &amp; Co. ⁸</td>
</tr>
<tr>
<td>(McP)</td>
<td>McGladrey &amp; Pullen</td>
</tr>
<tr>
<td></td>
<td>Including: McGladrey Hendrickson &amp; Co.</td>
</tr>
<tr>
<td></td>
<td>Broeker, Hendrickson &amp; Co.</td>
</tr>
<tr>
<td></td>
<td>McGladrey, Hansen, Dunn &amp; Co.</td>
</tr>
<tr>
<td></td>
<td>A.M. Pullen &amp; Co.</td>
</tr>
<tr>
<td></td>
<td>McGladrey, Hendrickson, Pullen &amp; Co.</td>
</tr>
<tr>
<td>(PKF)</td>
<td>Pannell Kerr Forster (Harris Kerr Forster)</td>
</tr>
<tr>
<td>(SS)</td>
<td>BDO Seidman (Seidman &amp; Seidman)</td>
</tr>
<tr>
<td></td>
<td>Including: Wolfson, Weiner, Ratoff &amp; Lapin</td>
</tr>
</tbody>
</table>

⁹AY and EW merged (1989) to form Ernst & Young.
⁸DHS and TR merged (1989) to form Deloitte & Touche.
⁶⁴Merged with TR (1977).
⁶⁵Merged with EW (1978).
Table 2

Sources Classified Along Disclosure Spectrum

General Sources


Auditor litigation news from reviewing various national business periodicals and national periodicals with business sections (e.g., Business Week, Forbes, Fortune, the New York Times, and USA Today) and press releases in PR Newswire.

Specialized Sources

Accounting


Legal

Cases reported in Mead Data's LEXIS system in the general federal file, state files, and federal bankruptcy file (January 1960–December 1989) where audit firms in Table 1 were named as either plaintiff/defendant or appellee/appellant.

SEC Accounting Series Releases (ASR) and Accounting and Auditing Enforcement Releases (AAER) (January 1960–July 1990).

Auditor litigation cases reported in BNA Securities Regulation Law Reports.

Combination Accounting, Legal, and Other

Cases disclosed in Mead Data's NAARS system in files for general annual reports and proxy statements (December 1985, May 1989, May 1990). Supplemented by examining annual reports, 10-K's, and proxy statements available in the UC-Berkeley and USC libraries for sample client companies.

Review of regional press disclosures in several geographic areas (Los Angeles, San Francisco, Seattle, and Tucson) and the AICPA clipping service.

Lawsuits mentioned in studies of accountants' liability or auditor litigation [e.g., Gormley, 1988; Kellogg, 1984; St. Pierre and Anderson, 1984].

Miscellaneous sources including auditing texts [e.g., Arens and Loebbecke 1984; Carmichael and Willingham, 1979] and books on accountants' liability [e.g., Causey, 1982; Davies, 1983; Gormley 1981; Jaenicke, 1977].

Private

Table 3
Cases Classified by Disclosure Source(s)

<table>
<thead>
<tr>
<th>General Sources (Only or in Combination With Specialized Sources)</th>
<th>Number of Cases</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>WSJ (only)</td>
<td>78</td>
<td>10%</td>
</tr>
<tr>
<td>WSJ and Legal Sources</td>
<td>69</td>
<td>9%</td>
</tr>
<tr>
<td>WSJ and Accounting Sources</td>
<td>45</td>
<td>6%</td>
</tr>
<tr>
<td>WSJ, Legal, Accounting and Other Specialized Sources</td>
<td>128</td>
<td>17%</td>
</tr>
<tr>
<td>Subtotal WSJ</td>
<td>320</td>
<td>42%</td>
</tr>
<tr>
<td>Other General Sources</td>
<td>18</td>
<td>2%</td>
</tr>
<tr>
<td>Subtotal</td>
<td>338</td>
<td>44%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Specialized Sources</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Legal Sources (only)</td>
<td>184</td>
<td>24%</td>
</tr>
<tr>
<td>Accounting Sources (only)</td>
<td>24</td>
<td>3%</td>
</tr>
<tr>
<td>Legal and Accounting Sources</td>
<td>55</td>
<td>7%</td>
</tr>
<tr>
<td>Other Sources (only)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regional Press</td>
<td>4</td>
<td>1%</td>
</tr>
<tr>
<td>Financial Reports</td>
<td>30</td>
<td>4%</td>
</tr>
<tr>
<td>Legal, Accounting, and Other Specialized Sources</td>
<td>31</td>
<td>4%</td>
</tr>
<tr>
<td>Subtotal</td>
<td>328</td>
<td>43%</td>
</tr>
</tbody>
</table>

| Private Sources                                                  | 95             | 13%        |
| Total                                                           | 761            | 100%       |
Table 4
Number of WSJ Disclosures

<table>
<thead>
<tr>
<th>Number of Cases</th>
<th>Number of WSJ Disclosures</th>
</tr>
</thead>
<tbody>
<tr>
<td>138 (43%)</td>
<td>138 (15%)</td>
</tr>
<tr>
<td>79 (25%)</td>
<td>158 (17%)</td>
</tr>
<tr>
<td>26 (8%)</td>
<td>78 (9%)</td>
</tr>
<tr>
<td>25 (8%)</td>
<td>100 (11%)</td>
</tr>
<tr>
<td>52 (16%)</td>
<td>439 (48%)</td>
</tr>
<tr>
<td>320 (100%)</td>
<td>913 (100%)</td>
</tr>
<tr>
<td>Client</td>
<td>Total Sample</td>
</tr>
<tr>
<td>----------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Public Company</td>
<td>519</td>
</tr>
<tr>
<td>Non-Public Company</td>
<td>242</td>
</tr>
<tr>
<td></td>
<td>761</td>
</tr>
<tr>
<td>Exchange: NYSE</td>
<td>177</td>
</tr>
<tr>
<td>AMSE</td>
<td>65</td>
</tr>
<tr>
<td>OTC</td>
<td>208</td>
</tr>
<tr>
<td>Unknown</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>519</td>
</tr>
<tr>
<td>Financial Distress (Only)</td>
<td>233</td>
</tr>
<tr>
<td>Fraud (Only)</td>
<td>125</td>
</tr>
<tr>
<td>Both Financial Distress/Fraud</td>
<td>190</td>
</tr>
<tr>
<td>Neither</td>
<td>213</td>
</tr>
<tr>
<td></td>
<td>761</td>
</tr>
<tr>
<td>Case</td>
<td></td>
</tr>
<tr>
<td>Criminal Action Against Auditor</td>
<td>6</td>
</tr>
<tr>
<td>SEC Action Against Auditor</td>
<td>55</td>
</tr>
<tr>
<td>Resolution (see Table 7)</td>
<td></td>
</tr>
<tr>
<td>Environment (Time-Period)</td>
<td></td>
</tr>
<tr>
<td>Year(s) of Financial Statements</td>
<td></td>
</tr>
<tr>
<td>Prior 1970</td>
<td>85</td>
</tr>
<tr>
<td>1970-1974</td>
<td>179</td>
</tr>
<tr>
<td>1975-1979</td>
<td>106</td>
</tr>
<tr>
<td>1980-1984</td>
<td>225</td>
</tr>
<tr>
<td>1985-1990</td>
<td>166</td>
</tr>
<tr>
<td></td>
<td>761</td>
</tr>
<tr>
<td>Auditor</td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>97</td>
</tr>
<tr>
<td>AY</td>
<td>52</td>
</tr>
<tr>
<td>CL</td>
<td>74</td>
</tr>
<tr>
<td>DHS</td>
<td>55</td>
</tr>
<tr>
<td>EW</td>
<td>69</td>
</tr>
<tr>
<td>PMM</td>
<td>132</td>
</tr>
<tr>
<td>PW</td>
<td>69</td>
</tr>
<tr>
<td>TR</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>632</td>
</tr>
<tr>
<td>Other</td>
<td>129</td>
</tr>
<tr>
<td></td>
<td>761</td>
</tr>
</tbody>
</table>

1Chi-square computed by comparing cases with and without WSJ disclosure for characteristic.

2Significant at less than .10 level.
Table 6
Cases With WSJ Disclosures
Summarized by Litigation Stage(s) Disclosed

<table>
<thead>
<tr>
<th>Existence of Litigation</th>
<th>Total</th>
<th>Other</th>
<th>Highly Visible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disclosed (Stages 1 and/or 2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stages 1 and 2</td>
<td>22</td>
<td>17</td>
<td>5</td>
</tr>
<tr>
<td>Stage 2 (only)</td>
<td>116</td>
<td>115</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>138 (43%)</td>
<td>132 (49%)</td>
<td>6 (12)</td>
</tr>
<tr>
<td>Existence of Litigation and/or</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information on the Resolution Process Disclosed (Stages 1, 2, and/or 3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stages 1, 2, and 3</td>
<td>8</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Stages 1 and 3</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Stages 2 and 3</td>
<td>21</td>
<td>16</td>
<td>5</td>
</tr>
<tr>
<td>Stage 3 (only)</td>
<td>12</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>44 (14%)</td>
<td>32 (12%)</td>
<td>12 (23)</td>
</tr>
<tr>
<td>Existence of Litigation, Information on the Resolution Process, and/or Final Resolution Information Disclosed (Stages 1, 2, 3, and/or 4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stages 1, 2, 3, and 4</td>
<td>8</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Stages 1, 2, and 4</td>
<td>10</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Stages 1, 3, and 4</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Stages 1 and 4</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Stages 2, 3, and 4</td>
<td>25</td>
<td>7</td>
<td>18</td>
</tr>
<tr>
<td>Stages 2 and 4</td>
<td>30</td>
<td>27</td>
<td>3</td>
</tr>
<tr>
<td>Stages 3 and 4</td>
<td>11</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>Stage 4 (only)</td>
<td>48</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>138 (43%)</td>
<td>104 (39%)</td>
<td>34 (65)</td>
</tr>
</tbody>
</table>

Stage 1: Pre-Litigation
Stage 2: Initiation or Existence of Legal Action(s)
Stage 3: Resolution Process
Stage 4: Final Resolution
Table 7
Resolution Information Disclosed in the WSJ

<table>
<thead>
<tr>
<th>Number of WSJ Disclosures</th>
<th>Final Resolution</th>
<th>Number of Cases</th>
<th>Highly Visible</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>Other</td>
</tr>
<tr>
<td>13 (6%)</td>
<td>Auditor Dismissed or No Auditor Payments</td>
<td>13 (9%)</td>
<td>10 (10%)</td>
</tr>
<tr>
<td>30 (15%)</td>
<td>Auditor Payments of Less Than $1 Million</td>
<td>26 (19%)</td>
<td>25 (24%)</td>
</tr>
<tr>
<td>69 (34%)</td>
<td>Auditor Payments of $1 Million or More</td>
<td>49 (36%)</td>
<td>27 (26%)</td>
</tr>
<tr>
<td>20 (10%)</td>
<td>WSJ Discloses Existence of Auditor Payments But Not Amount Paid</td>
<td>15 (11%)</td>
<td>12 (11%)</td>
</tr>
<tr>
<td></td>
<td>Subtotal</td>
<td>103 (75%)</td>
<td>74 (71%)</td>
</tr>
<tr>
<td></td>
<td>Regulatory Actions Only</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Disclosed Final Resolutions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>66 (33%)</td>
<td>SEC</td>
<td>30 (22%)</td>
<td>25 (24%)</td>
</tr>
<tr>
<td>5 (2%)</td>
<td>Other</td>
<td>5 (3%)</td>
<td>5 (5%)</td>
</tr>
<tr>
<td>71 (35%)</td>
<td>Subtotal</td>
<td>35 (25%)</td>
<td>30 (29%)</td>
</tr>
<tr>
<td>203 (100%)</td>
<td>Total</td>
<td>138 (100%)</td>
<td>104 (100%)</td>
</tr>
</tbody>
</table>
### Table 8
Cases With Five or More WSJ Disclosures

<table>
<thead>
<tr>
<th>AA: (5)</th>
<th>DSC Communications (1985-90)b</th>
</tr>
</thead>
<tbody>
<tr>
<td>(8)</td>
<td>DeLorean Motor Co. (1984-88)</td>
</tr>
<tr>
<td>(10)</td>
<td>Drysdale Gvt. Sec. (1982-87)</td>
</tr>
<tr>
<td>(9)</td>
<td>Financial Corp. of Am. (1984-89)</td>
</tr>
<tr>
<td>(7)</td>
<td>Four Seasons (1972-77)</td>
</tr>
<tr>
<td>(6)</td>
<td>Frigitemp (1979-84)</td>
</tr>
<tr>
<td>(5)</td>
<td>Fund of Funds (1977-85)</td>
</tr>
<tr>
<td>(6)</td>
<td>Home State S&amp;L (1985-88)</td>
</tr>
<tr>
<td>(8)</td>
<td>Marsh and McLennan (1984-88)</td>
</tr>
<tr>
<td>(7)</td>
<td>NuCorp (1982-88)</td>
</tr>
<tr>
<td>(5)</td>
<td>SeaFirst (1984-86)</td>
</tr>
<tr>
<td>(10)</td>
<td>Southern Cos. (1988-90)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TR: (6)</th>
<th>Ampex (1972-79)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(10)</td>
<td>Beverly Hills S&amp;L (1985-89)</td>
</tr>
<tr>
<td>(7)</td>
<td>Data Access (1981-85)</td>
</tr>
<tr>
<td>(9)</td>
<td>Giant Stores (1973-84)</td>
</tr>
<tr>
<td>(5)</td>
<td>Reliable Investors (1981-84)</td>
</tr>
<tr>
<td>(6)</td>
<td>U.S. Financial (1973-78)</td>
</tr>
<tr>
<td>(7)</td>
<td>Ziff Davis/CBS (1985-90)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(47)</td>
<td>ESM (1985-90)</td>
</tr>
<tr>
<td>(6)</td>
<td>Southmark (1989-90)</td>
</tr>
<tr>
<td>(5)</td>
<td>Tiffany Ind. (1979-82)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fox: (6)</th>
<th>Flight Transport. (1982-85)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(10)</td>
<td>Saxon (1983-86)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MH: (10)</th>
<th>Technical Equities (1986-88)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(6)</td>
<td>Wedtech (1987-89)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SS: (7)</th>
<th>Cenco (1976-82)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(15)</td>
<td>Equity Funding (1973-87)</td>
</tr>
<tr>
<td>(5)</td>
<td>SaCom (1976-79)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DHS: (7)</th>
<th>Equity Funding (1973-77)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(5)</td>
<td>First Peoples Bank NJ (1983-84)</td>
</tr>
<tr>
<td>(7)</td>
<td>United American Bank (1983-87)</td>
</tr>
<tr>
<td>(9)</td>
<td>ZZZZ Best (1987-88)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PMM: (7)</th>
<th>Coated Sales (1988-89)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(5)</td>
<td>Itel (1980-85)</td>
</tr>
<tr>
<td>(14)</td>
<td>National Student Mkt. (1972-89)</td>
</tr>
<tr>
<td>(9)</td>
<td>Penn Central (1970-89)</td>
</tr>
<tr>
<td>(9)</td>
<td>Penn Square Bank (1982-86)</td>
</tr>
<tr>
<td>(9)</td>
<td>Regina Co. (1988-89)</td>
</tr>
<tr>
<td>(9)</td>
<td>Republic Mt. Life/Rel.Eq. (1974-89)</td>
</tr>
<tr>
<td>(7)</td>
<td>Stirling Homex (1975-89)</td>
</tr>
<tr>
<td>(7)</td>
<td>Talley Industries (1973-89)</td>
</tr>
</tbody>
</table>

| PW: (5)  | AM International (1981-90) |

---

aNumber of WSJ disclosures.

bTime-period of WSJ disclosures.
NINTH SYMPOSIUM
ON AUDITING RESEARCH

October 4-6, 1990

Research Papers

Evaluating Representations by Management in Analytical Review: The Importance of Domain-Specific Evidence

by

Jean C. Bedard
Stanley F. Biggs

Discussant:  David M. Frederick

Section 4
THE EFFECT OF DOMAIN-SPECIFIC EXPERIENCE ON EVALUATION OF
MANAGEMENT REPRESENTATIONS IN ANALYTICAL PROCEDURES

Jean C. Bedard

Stanley F. Biggs

Both at:
Department of Accounting, Box U-41A
School of Business Administration, 368 Fairfield Road
University of Connecticut
Storrs, CT 06268

September, 1990

The authors would like to thank Jean Bédard, Cindy Moeckel, Andy Rosman, Arnie Wright, and participants of the accounting workshops at Northeastern University and the University of Connecticut, and two anonymous reviewers for helpful comments. The research assistance and comments of Gail Sergenian are gratefully acknowledged. Financial support for this research was provided by the program of Corporate Associates Research Grants at the University of Connecticut School of Business Administration.
THE EFFECT OF DOMAIN-SPECIFIC EXPERIENCE ON EVALUATION OF MANAGEMENT REPRESENTATIONS IN ANALYTICAL PROCEDURES

ABSTRACT

While conducting an audit, auditors often seek explanations from clients regarding material findings. Such representations by management are to be considered as evidence, but not necessarily as competent evidence because lack of independence may reduce validity. This study evaluates the effect of management representations that vary in completeness on auditor performance of analytical procedures. The following features of the study are important. First, its experimental task enables determination of outcome quality, in order to distinguish performance. Second, the task provides a complex context for the study of auditor expertise. Third, several surrogate measures for expertise are employed, from general (rank and years of experience) to specific (recent time spent auditing manufacturing companies).

Our findings indicate three important conclusions. First, the type of representation given by management as an explanation for analytical results significantly influenced the quality of the auditors' responses. Specifically, while the presence of an incomplete representation by management did not adversely affect judgment in the task, performance was significantly improved when the explanation by management was consistent with the cues. Second, domain-specific experience was more highly associated with improved performance than were general measures of experience. Third, improvement associated with high levels of domain experience was apparent only when auditors were given a representation by management that completely explained all discrepancies in the case situation.
THE EFFECT OF DOMAIN-SPECIFIC EXPERIENCE ON EVALUATION OF MANAGEMENT REPRESENTATIONS IN ANALYTICAL PROCEDURES

While conducting an audit, auditors often seek explanations from clients regarding material findings. Such inquiries are a common audit procedure and do yield evidence useful in signalling errors [Hylas and Ashton, 1982; Kreuzfeldt and Wallace, 1986; Wright and Ashton, 1989]. Auditing standards suggest that an explanation given by client management (termed a "representation" by AU 333) is to be considered as evidence, but not necessarily as competent evidence because of lack of independence (AU 326.18). The ability to evaluate and disconfirm invalid explanations is an important determinant of audit quality, as is illustrated by management attempts to "explain away" fraudulent transactions in the Equity Funding case [Seidler et al., 1977].

Management representations are essentially "inherited" hypotheses regarding the nature of events underlying accounting numbers, until the auditor subsequently accepts or rejects their validity through accumulation of evidence. There is evidence that the presence of inherited or self-generated hypotheses may inhibit achieving an adequate explanation of observed events [Patel and Groen, 1986; Johnson et al., forthcoming]. If management representations about the cause of discrepancies affect auditor judgment, the potential exists for reduction of audit efficiency and effectiveness.

In this research, we examine the effect of management representations on the quality of hypotheses generated by auditors as they perform analytical procedures, and how that relationship varies with amount and type of audit experience.
This study builds on a previous investigation [Bedard and Biggs, 1989], which studied auditors' recognition and explanation of patterns of financial statement discrepancies. The experimental task of that study contained an incomplete explanation of discrepancies by the client's management. However, that study was concerned with individual decision processes, and did not examine the effect of the management representation experimentally. Here, we gave the same basic case information to a large number of auditors, and manipulated the completeness of representations given by management.

Bedard and Biggs [1990] found that managers were better than seniors at explaining a recognized pattern of cues. This result suggests that general experience measures such as rank can be associated with auditors' ability to fit combinations of cues obtained from audit evidence with patterns in memory. According to Libby and Frederick [1989], more experienced auditors have a larger number of plausible errors available, and so are less likely to miss consideration of a particular error because that error is not present in memory. However, in the Bedard and Biggs study, experience alone was not sufficient to produce correct performance, since one-half of the managers did not generate the correct hypotheses. In this task, it may be that specialized experience (i.e., experience in the particular domain of the financial statement error) is more important than general experience. Davis and Solomon [1989] for example, propose that specialized experience is likely to result in more rapid, efficient accumulation of representative error patterns in
memory. Thus, expected differences in performance across types of experience motivate examination here of how general and domain-specific experience are associated with auditor performance in the task.

This design responds to several recent calls for improvement in auditor judgment research (e.g., Biggs [1985], Bédard [1989], and Davis and Solomon [1989]). First, this study uses an experimental task with a criterion outcome to distinguish performance. Second, the task is designed to be a complex context for the study of auditor judgment. The experimental case was based on a situation from practice in which significant overbudget activity was caused by "truncated" analytical procedures [Kinney, 1987].

Also, prior results using the same basic case information indicate that the task is difficult, yet solvable for experienced auditors. Third, we incorporate several surrogate measures of expertise, from general (rank and years of experience) to specific (relative amount of recent time spent auditing manufacturing companies, the domain of the case).

This paper is divided into four remaining sections. Section I presents relevant background literature that motivates this research, and specifies the hypotheses tested. Section II describes the experimental task and outlines the empirical methods used. In Section III, results of the hypothesis tests are presented. Section IV provides some conclusions based on the

---

1. Kinney [1987] refers to "truncated" analytical procedures as "acceptance of a seemingly plausible explanation without adequate analysis of all its implications" (p. 69).
experimental results and their implications for research and practice.

I. MOTIVATION AND RELEVANT LITERATURE

Management Representations.

As a result of audit procedures (e.g., inquiry) or through less formal means, oral or written representations from client management can be received at various points in the audit. AU 333 guides auditors to consider these representations as evidence, but not to rely on them to the exclusion of other evidence, due to the lack of independence of their source. Thus in the context of analytical procedures, auditors should investigate an explanation represented by client management, but remain able to disconfirm the explanation in the face of contradictory evidence.

However, both theory and empirical evidence support the notion that explanations received early in a decision process may influence outcomes. Einhorn and Hogarth [1983] found that alternative explanations significantly affect judgment. Empirical studies of medical experts [e.g., Patel and Groen, 1986] have also suggested that having a particular diagnostic hypothesis in mind may inhibit access to others. Judgment errors that result from the inability to disconfirm hypotheses in mind have been termed "Garden Path" errors by Johnson et al. [forthcoming], since later disconfirming evidence tends to be ignored or explained away. These studies suggest that auditors may have difficulty treating management representations as auditing standard's direct.
Some evidence has recently been obtained on the effect of suggested explanations on auditor judgment. Libby [1985] found some evidence that hypotheses generated in analytical procedures were influenced by hypotheses received from other members of the audit team. Heiman [1990] also studied analytical hypotheses and found that making alternative explanations more available reduced assessments of the likelihood of a particular hypothesized cause. Anderson and Kida [1989] studied the effect of credibility of the source of a hypothesis (i.e., client personnel are less credible than other members of the audit team due to lack of independence) on information search in an internal control decision. They found that source credibility did not affect information search, but did affect control risk assessments.

According to the decision process model given in Bedard and Biggs [1990], auditors who have recognized the pattern of discrepancies to be explained will attempt to match that pattern with patterns characteristic of various financial statement errors. If a management representation is present, it is likely that the fit of that proposed explanation will be evaluated along with other errors available in memory. The task is essentially to recognize whether the management representation provides an adequate explanation of the discrepancies. If the auditor determines that the management representation is inconsistent with the observed pattern, further processing must take place to recall from memory an explanation that is consistent with the pattern. This subsequent step is analogous to cued recall, because the management representation has already been attended
to and may influence their access to other explanations. A task with no management representation is analogous to a free recall situation. Gillund and Schiffrin's [1984] theory postulates a general familiarity process for recognition, while recall is assumed to involve cue-dependent probabilistic sampling and recovery from an associative network. Thus, recall is a more complex and difficult process than recognition. The dependency of sampling and recovery on available cues suggests that the presence of a suggested explanation can inhibit processing of alternative explanations, and this theory is in fact consistent with results of studies in medicine such as Patel and Groen [1986].

These theories and empirical results suggest two interesting comparisons for this study. First, we compare results from two cases that differ only by the presence or absence of the incomplete management representation used in the Bedard and Biggs study. As stated above, we expect that the incomplete explanation will interfere with the process of recalling another plausible explanation, leading to the following hypothesis:

H1A: A greater percentage of auditors receiving no management representation will propose correct hypotheses than will auditors given an incomplete explanation as a management representation.

Second, results from two cases differing only in the presence or absence of the correct answer as the management

---

2. These tasks are considered analogous to the elementary tasks of recognition since the experimental situation used here is more complex than tasks such as word recognition and recall used in psychological experiments.
representation will be compared. Given the theories of ease of recognition vs. recall tasks, we expect that performance should be improved in the "complete" representation case, leading to the following hypothesis:

H1B: A greater percentage of auditors given the correct explanation as a management representation will propose correct hypotheses than auditors given the incomplete representation or no representation.

Expertise and Performance

Both the value and the difficulty of developing objective criteria for evaluating performance in auditing tasks are well known (Biggs [1985], Davis and Solomon [1989]). Because performance criteria are often lacking, researchers turn to assessing the performance of auditors who might be accomplished at the task because of long experience or attainment of rank. Bédard [1989] and Davis and Solomon [1989] both emphasize that common surrogates for expertise, such as years of experience or rank in an audit firm, do not translate directly into expertise in any particular auditing task. Empirical studies such as Abdolmohammadi and Wright [1987], Bonner [1990], and Rosman [1990] reinforce the point that experience specific to the task domain is often important, especially in relatively complex, unstructured tasks.

In the present study, the nature of the task concerns application of analytical procedures in a manufacturing audit; specifically, detecting that the cause of a pattern of financial statement discrepancies is an error in allocating inventory costs. Thus, the "domains" of the task could be considered as
(1) analytical procedures in general (developing expectations for unaudited balances, and analyzing reasons for discrepancies found between expected and actual balances); and (2) specifically, analytical procedures in manufacturing accounts. General skill in analytical procedures should develop over time as the number of audits performed increases. However, to the extent that auditors specialize as they advance within the firm, expertise in auditing manufacturing accounts may not be improved with additional analytical experience. The marginal contribution of these general or specific aspects of experience to performance in this context is unknown. However, only half of the managers studied in Bedard and Biggs [1990] hypothesized an error that explained all discrepancies present in the cues. Although their performance was better than that of the seniors, the number of incorrect managers was somewhat surprising. While subjects for that study had been selected by firm personnel on the basis of their ability to perform analytical procedures, a specific measure of experience in manufacturing was not obtained.

H2 predicts that greater experience will positively affect performance. H2 is operationalized as four separate hypotheses, each using a different experience measure, with the more general in H2A and H2B, and the more specific in H2C and H2D. It is likely that more specialized measures will result in improvement in the ability to distinguish performance. Therefore, support is weaker for H2A and H2B, and stronger for H2C and H2D.

H2A. A greater percentage of managers will propose correct hypotheses than will seniors.
H2B. A greater percentage of auditors with more years of auditing experience will propose correct hypotheses than will those with fewer years of auditing experience.

H2C. A greater percentage of auditors with more recent inventory experience will propose correct hypotheses than will those with less recent inventory experience.

H2D. A greater percentage of auditors with more recent manufacturing experience will propose correct hypotheses than will those with less recent manufacturing experience.

Interactions.

The above theories may also lead to an expectation that the relationships between performance and versions of the management representation received by auditors will be conditional on experience. For instance, due to their better developed memory structures, we might expect that auditors with greater experience (especially domain-specific experience) would be more able to disconfirm an incorrect hypothesis. Results of Bedard and Biggs [1990] are consistent with this explanation, as are Libby and Frederick’s [1989] findings on experience and availability of plausible errors in memory. However, both Johnson et al. [forthcoming] and Patel and Groen [1986] found that hypothesis generation behavior (and thus diagnostic performance) of even expert physicians can be affected by the presence of early hypotheses. We therefore state the hypothesis on interactions as exploratory and non-directional:

H3: There will be a relationship between performance, the version of the management representation received, and measures of expertise.
As in H2, experience is operationalized by the four variables: rank, years of audit experience, percentage of recent audit experience in clients with inventory, and percentage of recent audit experience with manufacturing clients specifically.

II. RESEARCH METHODS

Task Development and Validation

This study uses the same basic case information as Bedard and Biggs [1990]. To develop the task from practice, the authors worked with a senior manager at a Big Six accounting firm to identify a situation based on his experience, in which better use of analytical procedures could have improved audit efficiency. A situation involving an error in allocation of overhead to inventory was selected as the focus of the task. Considerable over-budget audit activity had been incurred in identifying the nature of that error (the inventory had been recounted, and tests of controls over inventory had been increased). Significant audit resources could have been saved had the audit focus been narrowed through analytical procedures.

The case was developed by first establishing correct balances and ratios, and then seeding the misallocated overhead. The senior manager who participated in identifying the audit problem read the case to insure that it was consistent with the firm's experience. The task was then revised several times based on comments by a partner at a second Big Six firm, to insure that: (1) the case was not firm-specific; (2) it was reasonably realistic; and (3) subjects would need no other materials in order to solve it within a reasonable time period. Pilot testing
was then performed on the final version by two auditors in one of the Big Six firms, two doctoral students, and one auditing professor (all with public accounting experience).

Task Content and Instructions

Financial information in the task was presented as "projected" and "unaudited," which differed only by the presence of the seeded error. Projections were described as summaries of expectations based on results from previous quarters of the current year, past audited results, and industry trends. Unaudited values were the client's representations of end-of-year balances. Subjects were told that the projections were prepared by the auditing firm, and could be considered reliable. In that way, their attention was focused on the client's numbers as the source of differences, not on an inaccurate projection. As in Libby [1985], subjects were told to assume that all discrepancies in the case were caused by a single error or misrepresentation. When performing the case, subjects assumed a supervisory role on a continuing audit, advising the senior regarding a series of discrepancies that the senior has discovered.

The seeded error resulted in the following discrepancies between projected and actual: inventory overstated, $150,000; cost of sales overstated, $50,000; selling, general and administrative expenses understated, $200,000; income before taxes overstated, $150,000; and net income overstated, $90,000. Some important discrepancies (e.g., cost of sales) had to be calculated from the information given in the case. If these cues are considered to have been caused by a single error, then the
resulting conclusion is that some part of SG&A expense was capitalized to inventory (and under a FIFO system was partially expensed in cost of sales). This pattern is consistent with an error or misrepresentation in allocating overhead (or other costs) to inventory. However, in assessing subjects' performance (see discussion of procedures below), an error in allocation of overhead costs was not presumed to be the only correct response. Rather, each proposed error was carefully analyzed to determine whether or not it was consistent with the financial statement discrepancies in the case.

To accomplish the manipulation of the management representation (MGTREP), three versions of the case were randomly assigned. Case I contained the incomplete representation of the original laboratory study, Case N contained no representation, and Case C contained the correct outcome as the management representation.

Following presentation of the introductory text and financial data described above, subjects were asked to record an error that they considered to be the "most likely" hypothesis (MLH) about the cause of the financial statement discrepancies found during analytical procedures. The "most likely" form of response is consistent with the Einhorn and Hogarth view ([1986], p. 4) that reasoning about causes of events usually involves uncertainty. They were then asked to give a retrospective protocol describing the thought processes leading to their MLH, and then to list other (less likely) hypotheses. A debriefing questionnaire gathered several types of information, including
the following measures of the subject's experience: (1) years of audit experience; (2) current rank; (3) percentage of recent (past three to five years) experience with clients in various industries, including manufacturing and retailing.

Distribution Procedures and Subjects

Materials were distributed and collected by a contact person in the offices of four Big Six firms that had previously agreed to participate in the experiment. The contact person and a cover letter from firm management were used to improve subject motivation. Subjects were instructed to perform the task alone, and to refrain from referring to the task in conversation with others. Of the 237 packets distributed to the firms, 181 were returned (76 percent). Five were omitted due to illegibility or missing information. Of the usable responses, 79 were managers and 97 were seniors. The mean number of years of auditing experience for the subjects was 5.13 (7.2 for managers and 3.5 for seniors). The subjects reported a mean percentage of audit time spent on manufacturing clients (over the past three to five years) as 30.7 percent (managers: 33.3; seniors: 28.6). The mean percentage of audit time spent on clients with inventory (manufacturing plus wholesale and retail clients) was 38.2 percent (managers: 44.5; seniors: 32.8).

III. RESULTS

Coding of Subjects' Hypotheses

In order to establish subjects' performance, each hypothesis was coded according to whether it explained the discrepancies in
the case ("Correct") or not ("Incorrect"). The following classification method was used by two independent coders, one of the authors and a doctoral student with extensive public accounting experience. Each hypothesis (the MLH, the alternative hypotheses, and any contained in the retrospective protocol) was seeded into the projected values to determine whether it would have caused discrepancies similar to those found in the case. A hypothesis was designated as "Correct" if it explained all discrepancies between projected and unaudited balances and ratios. Otherwise, the hypothesis was termed "Incorrect." Agreement between the two coders was 95%; the few differences were subsequently reconciled by the other author.

The results of hypothesis testing are presented in three parts, reflecting the statistical procedures used to assess the relationship between performance and the various measures of experience noted above. First, we present results of testing the basic relationships posited in H1 (MGTREP will affect performance) and H2 (the subject's audit experience affects performance). Tests of the interaction hypothesis (H3) take two forms. In the third section below, we report results of chi-squared tests on the data partitioned by both MGTREP and the experience measures. In addition, loglinear models were also fit

3. Allowance of only MLH responses in the determination of a correct outcome would be a restrictive criterion. Although it is probable that audit effort would be directed toward errors considered most likely, auditors in practice often design tests to fulfill multiple objectives. Therefore, if a correct hypothesis appeared anywhere in any of the three areas listed, the response was deemed correct. Statistical conclusions do not differ for the more restrictive definition.
to the data to discover simultaneous relationships. Those results are presented in the fourth section below.

H1: Effect on Performance of a Management Representation.

The first set of hypotheses concerns the effect of MGTREP on performance. Table 1 shows numbers of subjects achieving the correct response in each version of the case. H1A proposed that an incomplete explanation would reduce subjects' ability to generate and evaluate the correct hypothesis. Table 1 results (using the likelihood ratio chi-squared test $G^2$) show that H1A is not supported. This result indicates that auditors inheriting the incomplete hypothesis from management performed as well as those inheriting no hypothesis. Further evidence of the influence of the management representation is presented later.

Insert Table 1 about here

Results of testing H1B provide support for the proposition that when MGTREP is correct (the recognition task), performance is improved relative to no MGTREP. When compared to performance of subjects given either Case I (incomplete representation) or Case N (no representation), significantly more subjects given Case C were correct. This result suggests that availability of the correct hypothesis early in the process is important in assisting these auditors in the task, and is consistent with expectations derived from Gillund and Schiffrin [1984]. Implications of this result are discussed in the final section.

Process Evidence from Retrospective Protocols. Since H1A was not rejected, we sought further evidence from the
retrospective protocols regarding how subjects evaluated the incomplete management representation. Specifically, examination of the hypotheses proposed as outcomes does not differentiate between two alternative theories: (1) that subjects given Case I evaluated and then disconfirmed that explanation in sufficient numbers to result in no significant difference with Case N; or (2) that subjects given Case I failed to attend to the MGTRP cue. Collection of a concurrent trace of the decision process (as in Bedard and Biggs [1990]) is the most accurate means of discriminating between these alternative theories. Since that method is precluded in a study of this size, we instead used immediate retrospective protocols to gain some understanding of subjects' hypothesis evaluation behavior. According to Ericsson and Simon [1984], the quality of a retrospective protocol as an indicator of decision process may be reduced by intermediate processing during search of long-term memory. Our analysis of these reports is directed toward positive evidence of evaluation of the MGTRP, in keeping with Ericsson and Simon's statement (p. 167) that ".. information that is reported is information that is heeded." However, this analysis is limited by the condition that no reference to the MGTRP in the protocol does not necessarily indicate that the subject did not attend to it.

Figure 1 shows results of analysis of the retrospective protocols directed toward the following specific points. (1) Was there evidence that the management representation was evaluated? (2) If evaluated, was there evidence that the management representation was specifically disconfirmed (i.e., an
inconsistency with at least one cue was mentioned)? (3) The final column notes whether or not the subject referred to the management representation in the MLH.4

Of the 47 incorrect subjects who could have adopted the MGTREP as a hypothesis, only 13 (28 percent) did so. There are varying degrees of evidence available as to how the others evaluated the MGTREP. Evidence is clearest for the 14 subjects who mentioned and specifically disconfirmed the MGTREP. The next clearest evidence is available from the three subjects who mentioned MGTREP in their protocol, did not specifically disconfirm it, and did not propose it as a hypothesis. The fact that they did not propose it is some evidence that it was disconfirmed. Finally, evidence is least clear for the 17 subjects who did not mention the MGTREP in the protocol. These subjects either did not attend to it, or did not consider it important enough to mention.

Some differences between managers and seniors are also evident from Figure 1. First, more managers showed specific evidence of disconfirmation ($G^2 = 3.608, p = .058$). Second, of the managers who did not disconfirm, all retained management's representation as being the most likely cause of discrepancies, whereas seniors were split among those who did and did not retain.

4. This serves as a form of validity check on the protocol analysis, since no subject whose protocol did not mention the MGTREP, or who disconfirmed it, referred to it in the MLH.
The large percentage of subjects who either disconfirmed or did not adopt the MGTREP provides evidence that professional skepticism was widely applied to the MGTREP. On the other hand, the fact that 13 subjects adopted the MGTREP is cause for concern.

**H2: Relationship of Performance to Experience.**

The four subhypotheses of H2 reflect expectations that the four measures of experience captured in this study would be related to performance in the task, although support varies with specificity of that experience. Table 2 gives results of hypothesis testing. H2A and H2B used general measures of experience that have been common in past audit judgment studies. For H2A (rank), 53 percent of managers were correct, compared to 41 percent of seniors, although the difference is not significant ($G^2 = 2.493, p = .114$). For H2B (years of experience) differences are again in the expected direction, although differences among the categories are not significant.5

| Insert Table 2 About Here |

The two domain-specific measures captured were degree of specialization in auditing clients with inventory (manufacturing, wholesale, or retail industries; H2C), and in auditing manufacturing clients (H2D). In both cases, auditors with relatively high recent experience with clients in those

---

5. Years of experience was divided into high, medium, and low categories based on evidence by Ashton (1988) that the effects of experience may be U-shaped instead of linear.
industries (measured as greater than fifty percent) were significantly more likely to achieve the correct response \( G^2 = 5.214 \) (\( p = .022 \)) and \( G^2 = 3.655 \) (\( p = .056 \)), respectively. Thus, H2C and H2D are supported, while H2A and H2B are not. These results provide evidence in favor of the arguments presented by Davis and Solomon [1989] and Bédard [1989] for developing more specific measures of experience as surrogates for expertise in audit judgment research.

Tests of H3: Interaction Between MGTREP and Experience

H3 postulated that the effect on performance of management representations might be contingent on levels of experience (domain or general) of the subjects. As an initial analysis, the data were partitioned by MGTREP and the four experience variables, giving the contingency tables shown in Table 3.

---

Insert Table 3 about here
---

Several important conclusions can be drawn from Table 3. First, the far right column shows results of testing for Performance differences by Case within each experience category (the rows). The Case variable is significantly associated with Performance in all rows. Second, below each contingency table is reported the test for column differences; e.g., did managers outperform seniors in Case I? Differences between managers and

---

6. To check for sensitivity of results to the cutoff level for high vs. low domain experience, tests were also run for cutoff levels of 40 and 60 percent, without a difference in statistical conclusion. However, when the cutoff was specified as 30 percent, differences were not significant.
seniors are in the expected direction, although not significant. Relative performance within Case varies for years of experience categories, and none of the comparisons is significant.

Since there was not a significant association between rank (or years of experience) and performance in Table 2, failure to find a significant association by Case (Table 3) is not surprising. However, a partition of the data by Case provides supplementary information on the significant relationship between domain-specific experience and performance overall. Table 3 shows that while performance of high domain experience subjects was better in five of the six column comparisons, the significant overall association is primarily driven by Case C results ($G^2 = 9.874 \ (p = .002)$ and $G^2 = 5.040 \ (p = .025)$) for inventory and manufacturing experience, respectively). Domain-specific experience was most associated with recognizing the fit of an "inherited" hypothesis to a pattern of discrepancies, rather than in recalling from memory a hypothesis that fits the pattern.

Tests of H3: Loglinear Modelling

In contrast to the separate column/row testing approach above, loglinear modeling gives an overall picture of simultaneous relationships among the three variables: Performance, the various measures of experience, and MGTREP. To select the best-fitting model, the BMDP4F procedure was used with backward elimination of terms, as outlined in Berenson et al. (1983, pp. 488-498). Table 4 shows results of fitting threeway loglinear models using the two domain experience variables found to be significantly associated with Performance overall. Common
to each are two variables: Performance (P; whether the subject's MLH was correct or incorrect, as described above), and the version of the management representation received (C). Each of the two experience variables is adopted in turn as the third variable in the model.

Insert Table 4 about here

Models selected as the best fit when both domain-specific experience variables are used show significant interactions. In both models, the effect of the Case version (MGTREP) on Performance is different for auditors with high vs. low experience in that domain. In addition, the relationship between domain experience and Performance also differs according to Case. Model I (experience with inventory), the selected model explains about 26 percent of the variance from the overall mean. Model II is a better fit, explaining about 57% of the overall variance. Computation of standardized residuals from both models produced none in excess of 0.9, which also indicates a good fit. Thus we find further evidence to support the expected interaction (H3) between performance, domain-specific experience and type of management representation.

IV. CONCLUSIONS AND IMPLICATIONS

The objectives of this study were to assess how hypothesis generation performance is affected by the presence of a

7. The notation used here is exemplified by (PXC), which denotes a significant interaction between Performance and Case (version of MGTREP). If the interaction between two variables is significant, those main effects are also included in the model.
management representation, and whether that association differs in auditors with various types and levels of experience. Findings indicate a significant overall relationship of performance with some types of experience, and one type of management representation studied.

The first important result was that good performance in the task was more associated with recent domain-specific experience than with general audit experience. This result provides evidence supporting arguments presented by Davis and Solomon [1989] and Bédard [1988] for developing more specific, task-related measures of auditor experience as surrogates for expertise. Similarly, domain experience is a key element of the cognitive computational model developed by Selfridge, Biggs and Krupka [1990] in the context of going concern judgments. A common feature of analytical procedures and going concern judgments is the analysis of patterns of financial statement data, which requires knowledge of specialized industry accounting and trends. As acknowledged by Davis and Solomon [1990] and Bonner [1990], this result may not hold for all audit tasks. Perhaps for more simple tasks, expertise acquired through general audit experience is sufficient for good performance.

The second important result of this study is that auditor performance in correctly hypothesizing the accounting error causing the financial statement discrepancies was significantly influenced by whether the correct solution to the case was given as a management representation. However, unlike other recent studies of the effect of inherited hypotheses (e.g., Moser
performance in this task was not significantly affected by the presence of an incomplete representation by management, relative to the case in which no inherited hypothesis was given. This latter result is consistent with the explanation that auditors are behaving as auditing standards suggest: i.e., evaluating management's representations as evidence but not being unduly influenced by them.

A deeper look at decision processes through evidence in retrospective protocols provides further evidence that many subjects disconfirmed the incomplete MGTREP. However, twelve of 62 subjects did not disconfirm and adopted the MGTREP as their hypothesis. Their failure to disconfirm is troubling if such behavior were to be found in a situation involving fraud. Thus, further research on processes of hypothesis confirmation and disconfirmation is needed to better understand how auditors evaluate management representations.

We also found that performance in the different tasks was contingent on experience. Specifically, domain-specific experience was associated with improvement in ability to recognize the correct answer when it was provided, which is essentially a recognition task. Surprisingly, audit experience as measured in this study did not significantly improve performance when the correct answer was not provided (i.e., subjects had to generate an MLH from memory). Theories of memory indicate that mechanisms needed for recall are more complex than those needed for recognition (e.g., Gillund and Schiffrin
The incremental benefit in performance from industry specialization appears to increase with task complexity.

This study's findings have important implications for practice. Although the lack of a significant interference effect caused by the incomplete hypothesis implies adherence to auditing standards, the percentage of correct hypotheses in groups not receiving the complete hypothesis was low. Audit situations that are most troublesome are those in which management does not supply a correct answer, either intentionally or due to lack of knowledge. The results of this and the prior study (Bedard and Biggs [1990]) suggest that audit efficiency and/or effectiveness could be compromised in these situations.
TABLE 1. CONTINGENCY TABLES AND RESULTS OF TESTING HYPOTHESIS 1

**H1A:**

<table>
<thead>
<tr>
<th>Case I (Incomplete Representation)</th>
<th>Correct</th>
<th>Incorrect</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20 (32%)</td>
<td>42</td>
<td>62</td>
</tr>
<tr>
<td>Case N (No Representation)</td>
<td>19 (35%)</td>
<td>36</td>
<td>55</td>
</tr>
<tr>
<td>G² (p)</td>
<td>.069 (.793)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**H1B:**

<table>
<thead>
<tr>
<th>Case I (Incomplete Representation)</th>
<th>Correct</th>
<th>Incorrect</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20 (32%)</td>
<td>42</td>
<td>62</td>
</tr>
<tr>
<td>Case C (Complete Representation)</td>
<td>43 (73%)</td>
<td>16</td>
<td>59</td>
</tr>
<tr>
<td>G² (p)</td>
<td>20.600 (.000**)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Case N (No Representation)         | 19 (35%)| 36        | 55 |
| Case C (Complete Representation)   | 43 (73%)| 16        | 59 |
| G² (p)                             | 17.291 (.000**) |          |    |
TABLE 2. CONTINGENCY TABLES AND RESULTS OF TESTING HYPOTHESIS 2

<table>
<thead>
<tr>
<th>H2A: Rank</th>
<th>Correct</th>
<th>Incorrect</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seniors</td>
<td>40 (41%)</td>
<td>57</td>
<td>97</td>
</tr>
<tr>
<td>Managers</td>
<td>42 (53%)</td>
<td>37</td>
<td>79</td>
</tr>
<tr>
<td>$G^2$ (p)</td>
<td>2.493 (.114)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>H2B: Years of Experience</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Low (≤3)</td>
<td>24  (42%)</td>
<td>33</td>
<td>57</td>
</tr>
<tr>
<td>Medium (4-5)</td>
<td>29  (48%)</td>
<td>31</td>
<td>60</td>
</tr>
<tr>
<td>High (≥6)</td>
<td>29  (49%)</td>
<td>30</td>
<td>59</td>
</tr>
<tr>
<td>$G^2$ (p)</td>
<td>.692 (.708)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>H2C: Recent Audit Experience with Clients with Inventory</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Low (≤50%)</td>
<td>42  (40%)</td>
<td>66</td>
<td>106</td>
</tr>
<tr>
<td>High (&gt;50%)</td>
<td>40  (57%)</td>
<td>30</td>
<td>70</td>
</tr>
<tr>
<td>$G^2$ (p)</td>
<td>5.214 (.022**)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>H2D: Recent Audit Experience with Manufacturing Clients</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Low (≤50%)</td>
<td>22  (42%)</td>
<td>104</td>
<td>126</td>
</tr>
<tr>
<td>High (&gt;50%)</td>
<td>29  (58%)</td>
<td>21</td>
<td>50</td>
</tr>
<tr>
<td>$G^2$ (p)</td>
<td>3.655 (.056**)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TABLE 3. PERCENT OF SUBJECTS ACHIEVING CORRECT RESPONSE, AND RESULTS OF TESTING HYPOTHESIS 3

<table>
<thead>
<tr>
<th>RANK</th>
<th>CASE I (INCOMPLETE)</th>
<th>CASE N (NONE)</th>
<th>CASE C (COMPLETE)</th>
<th>G² (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Senior</td>
<td>26%</td>
<td>28%</td>
<td>70%</td>
<td>16.93 (.000**)</td>
</tr>
<tr>
<td>Manager</td>
<td>41</td>
<td>42</td>
<td>76</td>
<td>9.19 (.010**)</td>
</tr>
</tbody>
</table>

G² (p) | 1.569 (.210) | 1.317 (.251) | 0.388 (.534) |

EXPERIENCE

<table>
<thead>
<tr>
<th>EXPERIENCE</th>
<th>CASE I (INCOMPLETE)</th>
<th>CASE N (NONE)</th>
<th>CASE C (COMPLETE)</th>
<th>G² (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low (&lt;4)</td>
<td>29%</td>
<td>18%</td>
<td>64%</td>
<td>9.36 (.009**)</td>
</tr>
<tr>
<td>Medium (4-5)</td>
<td>33</td>
<td>38</td>
<td>87</td>
<td>12.84 (.009**)</td>
</tr>
<tr>
<td>High (&gt;5)</td>
<td>35</td>
<td>40</td>
<td>74</td>
<td>7.06 (.029**)</td>
</tr>
</tbody>
</table>

G² (p) | 0.212 (.899) | 1.798 (.407) | 2.612 (.271) |

DOMAIN EXPERIENCE - CLIENTS WITH INVENTORY

<table>
<thead>
<tr>
<th>EXPERIENCE</th>
<th>CASE I (INCOMPLETE)</th>
<th>CASE N (NONE)</th>
<th>CASE C (COMPLETE)</th>
<th>G² (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low (&lt;50%)</td>
<td>30%</td>
<td>33%</td>
<td>57%</td>
<td>6.48 (.039**)</td>
</tr>
<tr>
<td>High (≥50%)</td>
<td>36</td>
<td>36</td>
<td>92</td>
<td>23.82 (.000**)</td>
</tr>
</tbody>
</table>

G² (p) | 0.261 (.610) | 0.053 (.817) | 9.874 (.002**) |

DOMAIN EXPERIENCE - MANUFACTURING CLIENTS

<table>
<thead>
<tr>
<th>EXPERIENCE</th>
<th>CASE I (INCOMPLETE)</th>
<th>CASE N (NONE)</th>
<th>CASE C (COMPLETE)</th>
<th>G² (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low (&lt;50%)</td>
<td>29%</td>
<td>35%</td>
<td>64%</td>
<td>11.52 (.003**)</td>
</tr>
<tr>
<td>High (≥50%)</td>
<td>40</td>
<td>33</td>
<td>90</td>
<td>15.74 (.000**)</td>
</tr>
</tbody>
</table>

G² (p) | 0.530 (.467) | 0.013 (.908) | 5.040 (.025**) |

*significant, ≤.05; **significant, ≤.10
<table>
<thead>
<tr>
<th>Model Selected:</th>
<th>Probability of Fit</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(PxI,PxC)</td>
<td>5.32</td>
<td>Performance (P) is related to Case (C), conditional on level of domain (inventory) experience (I); also, the relationship of P to I is conditional on C.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model Selected:</th>
<th>Probability of Fit</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(PxC,PxM)</td>
<td>2.99</td>
<td>Performance (P) is related to Case (C), conditional on level of domain (manufacturing) experience (M); also, the relationship of P to M is conditional on C.</td>
</tr>
</tbody>
</table>
FIGURE 1. CLASSIFICATION OF HYPOTHESIS EVALUATION BEHAVIOR USING RETROSPECTIVE PROTOCOLS (CASE I)

EVIDENCE THAT MGTREP EVALUATED?

--- YES: 7 M 4 S ---

--- YES: 11 M 15 S ---

--- NO: 4 M 11 S ---

INCORRECT: 17 M 25 S

--- NO: 5 M 10 S ---

EVIDENCE THAT MGTREP DISCONFIRMED?

--- YES: ---

--- NO: ---

MLH REFERS TO MGTREP?

--- YES: 4 M 8 S ---

--- NO: 0 M 3 S ---

--- NO: [ALL] ---
REFERENCES


Experimental Evidence on the Effects of Accountability on Auditor Judgments

by

Van E. Johnson
Steven E. Kaplan

Discussant: Mark E. Peecher

Section 5
Experimental Evidence on the Effects of Accountability on Auditor Judgments

Van E. Johnson
Assistant Professor
Accounting Department
Temple University
Philadelphia, Pennsylvania

Steven E. Kaplan
Associate Professor
School of Accountancy
Arizona State University
Tempe, Arizona

September, 1990

Van Johnson gratefully acknowledges the financial support of Ernst and Young. This research has benefitted from the comments of Cindy Moeckel and Phil Reckers.
purpose of this paper is to present additional evidence about the effects of accountability on auditor judgments.

An experiment was conducted in which auditors were given an inventory task and were required to identify potentially obsolete inventory items. Practicing auditors were assigned to a non-accountable (control) condition or to an accountable condition. The obsolescence judgments of the two groups were compared for consensus, consistency, stability, and self-insight.

Ashton [1990b] previously examined the effects of justification on auditor judgments. This study extends this prior work in three ways. First, while Ashton [1990a; 1990b] used experienced auditors as subjects, he employed a bond rating task. The inventory obsolescence task used in this study is more representative of one auditors' normally perform. Second, while Ashton examined the effects of justification, this study examines the effects of accountability. The two constructs, both present in many audit settings, are theoretically distinct. Finally, initial evidence is presented on the effects of accountability on self-insight and judgment stability among auditors.

The next section discusses the role of accountability in auditing. Section three details the hypotheses tested in this study tests and sections four and five describe the study and present the results. The final section discusses the results.
ACCOUNTABILITY IN AUDITING

Accountability is defined as the existence of social pressure to justify one's judgments to significant others [Tetlock, 1983a; 1985b]. The effects of accountability have been investigated across a wide variety of tasks such as social inference [Ford and Weldon, 1981; Rozelle and Baxter, 1981; Tetlock, 1983a, 1985a; Tetlock and Kim, 1987], attitude formation and change [Chaiken, 1980; Tetlock, 1983b], risky choice [Cvetkovitch, 1978], multiple cue probability learning [Hagafors and Brehmer, 1983], business simulation [McCallister, Mitchell, and Beach, 1979], negotiations [Ben-Yoav and Pruitt, 1984], and social choice [Adelberg and Batson, 1978]. The results from these studies have generally found that subjects who expect to be held accountable attend to information more thoroughly and process information more vigilantly [see Tetlock, 1985b for a thorough review of this literature].

While accountability is present in most natural environments, it is particularly important in auditing. An audit team is generally composed of at least one member from each of the following levels: partner, manager, senior, and staff. Audit teams are responsible for conducting routine, though not necessarily cognitively simple tasks [Solomon, 1987]. The review process is an inherent element of audit teams. The work performed, the methods used, and the conclusions drawn by a subordinate auditor are all subject to review by a superior auditor [Emby and Gibbins, 1988]. As a result of the review
process, a subordinate auditor (whether staff, senior, or manager) expects to be (and is) accountable to a superior. Additionally, outcome feedback is frequently delayed or incomplete in an audit setting. Accordingly, measures of judgment quality often depend on the apparent wisdom of the judgments rather than on the consequences [Gibbons, 1984].

Evidence provided by Gibbins and Emby [1984] and Emby and Gibbins [1988] suggests that auditors clearly perceive the importance of accountability in their environment. Auditors across all levels responded in a questionnaire that the ability to justify a decision is one of the most important qualities of professional judgment [Gibbins and Emby, 1984]. More recently Emby and Gibbins [1988], also employing a questionnaire methodology, found that justification influences auditors' evaluations of good judgments.

Ashton [1990a; 1990b] extends this line of research by providing experimental evidence on the effects of justification on auditor judgment. Practicing auditors were given a task in which they provided bond ratings for 16 companies based on the values of three financial ratios. In Ashton [1990b], the responses of a control group were compared to those of a treatment group in which the auditors were required to provide written justifications for each judgment. Ashton describes his justification manipulation as "private, written justifications to an experimenter of unknown beliefs and preferences who has no organizational authority with respect to the subjects" [1990b, p. 13]. Thus, the justification
variable did not necessarily bolster the extent to which subjects felt accountable for their judgments to a significant other.

Three results from Ashton's research are relevant for the current study. First, mean decision accuracy (measured as the correlation between the auditor ratings and the actual Moody ratings) was significantly higher for the justification group (0.378 versus 0.275). Second, mean consistency (measured as the multiple correlation coefficient obtained from regressing auditor ratings on the three ratios) was significantly higher for the justification group (0.906 versus 0.867) based on raw correlations. Consistency between the two groups did not significantly differ when the analysis was based on Fisher's Z-transformed correlations. Third, mean consensus (measured as the correlation between pairs of auditor ratings) was significantly higher for the justification group (0.732 versus 0.631).

Ashton [1990a] examined potential interactive effects between the existence of decision aids and the requirement for justification. The results show a significant interaction between justification and decision aids. Mean accuracy (measured as the number of ratings correctly predicted) of unaided judgments was improved by justification while the mean accuracy of aided judgments was lower as a result of justification.

HYPOTHESES

Tetlock's theory of accountability characterizes the decision maker as a politician. Within this characterization, people are assumed to be motivated to maintain the approval and respect of
those to whom they are accountable. The objective of the theory is "to identify the behavioral strategies that people have developed for coping with fundamental or invariant features of natural decision environments (features likely to be present in at least some degree in all social and organizational settings)" [1985, p. 306].

The theory has been tested by examining how strategies and styles of information processing change in response to changes in the demands for accountability in the environment. Results of these tests have generally been consistent with the view of people as cognitive misers [Taylor, 1980]. Information processing requires costs, in the form of cognitive effort which, other things being equal, individuals attempt to minimize. However, when an individual is accountable to a significant other, the benefits or stakes (rewards and/or penalties) associated with a judgment increase. Thus, the presence of accountability is expected to result in more vigilant (and more costly) information processing due to the increased benefits associated with the judgment.

With the exception of Ashton [1990a; 1990b], tests of Tetlock's theory have not involved professional subjects making professional judgments. While Ashton's results suggest that the theory may generalize to professionals settings, two observations can be noted. First, auditors rarely determine a bond rating for a company as part of the conduct of a normal audit, therefore the task given to auditors may be beyond the scope of normal audit judgment. Consequently, it is possible that the justification
effects found by Ashton may not extend to professional subjects performing familiar audit tasks.

Second, Ashton investigates the effect of justification rather than accountability. Although overlap exists between the two, they are not identical. Justification is the actual physical and/or mental process of explaining a judgment. Alternatively, accountability is a pre-existing expectation that an individual may be called on to justify his/her judgments to a significant other. Ashton's justification manipulation required auditors to provide written justifications to an experimenter (who may not have been seen as a significant other). Accordingly, it is not clear that Ashton's justification manipulation necessarily made auditors feel accountable for their judgments.

This study examines auditor judgment in the context of a routine audit task. Further, the focus is on accountability as opposed to justification. Specifically, four effects of accountability are predicted. The first two hypotheses represent replications of Ashton [1990b] while using an accountability manipulation and a more representative task. Auditors who are accountable are expected to process information more vigilantly. Consequently, this group of auditors is expected to make judgments that contain less error (e.g., are less random) than a group of auditors who are not accountable. Ashton [1974] defines consensus as agreement across auditors at the same point in time. Holding other factors constant, the proportion of error in auditor judgment will be negatively associated with auditor consensus. That is, increases in the random component of auditor judgment
will reduce the extent to which auditors' judgments will be in agreement. Consensus is an important criterion for many auditing tasks, particularly those tasks where accuracy measures are unavailable [Ashton, 1983; Ashton et al. 1989]. In addition to the significant results regarding consensus reported by Ashton [1990b], Rozelle and Baxter [1981] also found that accountability led to greater consensus in a non-auditing task.

In addition to displaying higher consensus, judgments by accountable subjects are also expected to be more consistent than non-accountable subjects. Consistency is defined as the proportion of variance captured by a model of the individual's judgments in relation to the cues. Because accountability reduces the extent of error in an individual's judgment, the variance (in individual judgments) explained by a model of the individual's judgments should increase. This leads to the first two hypotheses:

H1: Mean consensus for accountable subjects will be significantly greater than mean consensus for non-accountable subjects.

H2: Mean consistency for accountable subjects will be significantly greater than mean consistency for non-accountable subjects.

Accountability is also expected to increase the stability of auditor judgment. Stability represents the extent to which an auditor makes similar judgments for similar circumstances. Stability is high when an auditor, facing similar circumstances at
two points in time, makes the similar judgments. Because accountability reduces the extent of error in an auditor's judgment, stability of judgments over time should increase. This leads to hypothesis three.

H3: Mean stability for accountable subjects will be significantly greater than mean stability for non-accountable subjects.

Finally, accountability is expected to increase auditors' self-insight. Self-insight represents the extent to which an auditor is aware of his/her own judgment process. Auditors who are accountable are expected to attend to information more thoroughly when making judgments. Further, accountable subjects will anticipate the need to be able to explain and defend their judgments to others. Tetlock, Skitka, and Boettger [1989] refer to this process as pre-emptive self-criticism. This type of cognitive preparation is expected to improve self-insight. That is, if an auditor is to be better able to explain and defend judgments to others, the auditor should implicitly be better able to explain it to oneself. In a non-audit task, Cvetkovich [1978] found accountability increased self-insight. This leads to hypothesis four.

H4: Mean self-insight for accountable subjects will be significantly greater than mean self-insight for non-accountable subjects.
METHOD

An experiment was conducted to test the above four hypotheses. The task, independent variable, dependent variables, subjects, and data analysis are discussed below.

Task

The experimental task required auditors to review a sample of inventory items as part of a preliminary review to identify potentially obsolete inventory. The experimental materials consisted of instructions (including a brief client description), 20 inventory items to be reviewed, and a debriefing questionnaire.

The instructions described the client as a catalogue merchandiser of men's and women's clothing and indicated that preliminary analytical procedures were being performed to detect potentially obsolete inventory items as part of the normal year-end inventory procedures. It was emphasized that the review to be performed was preliminary in nature and that further audit work would be performed to determine the necessity and materiality of any inventory write-downs. Accordingly, auditors were told to assume that each inventory item was material for purposes of the analytical procedure. Further, auditors were told that prior year workpapers included write-downs for obsolete inventory items and that the inherent risk of inventory obsolescence for the client was considered fairly high.

An inventory task was selected because experienced auditors should be familiar with and should have worked extensively on the account. Inventory commonly represents a large portion of current
assets and frequently contains accounting errors [Hylas and Ashton, 1982]. The importance of this account may be further illustrated by considering that audit standards require auditors to perform substantive tests over inventory.

The 20 inventory cases presented to each auditor contained information related to five cues. The cues were selected after consulting several sources. Blocher and Cooper [1988] identify factors used by auditors in identifying potential inventory errors. In a protocol analysis experiment, the trend in gross profit percentage and in inventory turnover were among the most frequently used items. SAS 56 also emphasizes the importance of industry and non-financial client information in analytical procedures. Accordingly, industry sales information, competitor product information, and information on the product's life cycle were also included. As shown in Table 1, a high risk level and a low risk level was constructed for each of the five information cues. Table 1 also shows the two orders the information cues were presented to auditors.  

[INSERT TABLE 1 ABOUT HERE]

Based on the five cues, 16 inventory items were generated according to a one-half fractional replication of a $2^5$ factorial design. An additional four items were generated by randomly selecting four of the original 16 items to repeat. The order of the original 16 items was randomized and the four repeat cases were added to the end of the sequence in a random order. All
auditors received the same 20 cases in the identical order. Each inventory item (together with the response scale) was presented on a separate page. A sample page of the instrument is included as Figure 1.

[Insert Figure 1 About Here]

**Independent Variable**

Auditors were randomly assigned to one of two groups: accountable or non-accountable. The accountability manipulation was included in the instructions as shown in Table 2. In the accountable group, the first page of the instructions included a statement that their judgments would be reviewed by the researchers in conjunction with the firm's personnel and that they would be asked to explain their judgments during a breakout session later in the seminar. Subjects were told that firm personnel would review their judgments in order to enhance the perception of a significant reviewer. The accountability condition was restated at the end of the instructions. Auditors in the accountable condition were asked to sign their booklets to indicate that they understood the instructions and so that the booklets could be returned at the subsequent session. In the non-accountable group the instructions included a statement that the responses on the task would be completely anonymous.

[Insert Table 2 About Here]
Dependent Measure

The auditor's assessment of the risk of inventory obsolescence served as the dependent measure. Specifically, auditors responded to the statement, "Please mark an X on the scale below corresponding to your assessment of the risk of obsolescence for this item". Auditors responded using a nine-point continuous scale anchored at the endpoints by the statements "very low (high) risk of inventory obsolescence".

Debriefing Questionnaire

After reviewing the 20 inventory items, auditors were asked to provide demographic information and a self-insight measure. The self-insight measure required auditors to allocate 100 points between the five cues based on the relative importance of each cue to the auditor's judgments. The debriefing questionnaire also contained a manipulation check for accountability status.

Subjects

A total of 101 practicing auditors from one Big 6 accounting firm participated as subjects in the study. The auditors were attending a national training seminar. The mean age and audit experience of the auditors was 25 years and 38 months, respectively.

Data Analysis

Consensus. Consensus was measured by computing the correlation between the original 16 risk assessment judgments of each pair of auditors within a group.
Consistency. Consistency was measured as the square-root of the proportion of variance explained by each auditor's judgment model. This is equivalent to the square root of the proportion (SSmodel/SStotal) from the analysis of variance of each auditor's judgments. The individual ANOVA models were based on the original 16 cases.

Stability. Stability was measured by computing the correlation between each auditor's judgments for the first and second presentation of the four repeated items.

Self-Insight. Omega-squared statistics were calculated for each significant cue in an auditor's analysis-of-variance model. Insignificant cues were assigned a value of zero. The individual ANOVA models were based on the original 16 cases. Self-insight was measured by correlating the omega-squared statistics from the objective model with the subjective weights provided by each auditor.

RESULTS

The results are presented in five sections. In the first section the results of the manipulation check are given. The following four sections detail the results for consensus, consistency, stability, and self-insight. In these four sections results presented are based on analyses of both raw correlations and transformed Fisher's Z-values. Both sets of results are presented because disagreement exists over the conditions favoring the use of transformed Fisher's Z-values [e.g., see Silver and Dunlap, 1987].
Manipulation Check

Each auditor indicated whether they were told that they would be expected to justify their judgments. Of the 101 participating auditors, five failed the manipulation check and were dropped from the analysis.

Consensus

A t-test compared the consensus levels of the accountable and non-accountable groups. Descriptive statistics for the raw correlations and Fisher's Z-transformed correlations are shown in Table 3. A statistically significant difference in the predicted direction was found, as shown in Table 3. Accountable auditors achieved a mean consensus of 0.542 (based on raw correlations) whereas mean consensus was only 0.431 for non-accountable auditors. Thus, the results support hypothesis one. The improvement in consensus for accountable auditors is approximately 26%, which would appear to represent a practical as well as statistical difference. This improvement is somewhat greater than that found by Ashton [1990b], who reports an improvement from justification of approximately 16%.

[INSERT TABLE 3 ABOUT HERE]

Consistency

A t-test compared the consistency of accountable and non-accountable auditors. As shown in Table 3 the consistency between the two groups is not statistically different. Thus, hypothesis
two was not supported. The consistency of both groups was quite high. Further, the mean of the non-accountable group (based on the raw correlations) exceeded the mean of accountable group.

The consistency results are slightly different from those reported by Ashton [1990b]. Ashton found significant consistency differences based on raw correlations but not based on transformed correlations. In comparing the current result to Ashton's, two observations may be offered. The mean consistency for the control group was lower in Ashton's study than in this study. The accountable auditors in both studies performed in a relatively similar fashion. Secondly, there was a much higher dispersion in consistency among the accountable auditors in this study relative to the justification group in Ashton's study. The dispersion among control auditors was relatively similar in both studies.

**Stability**

Stability measures the correlation between the four original and repeated cases each auditor evaluated. The results of a two-way ANOVA with accountability and cue order as independent variables are presented in Table 4. As shown, a significant interaction was found using both raw and transformed correlations. The results are in the predicted direction for order one, with significantly greater stability for accountable than non-accountable auditors. However, the difference between accountable and non-accountable auditors for order two is statistically significant and in the opposite direction. Thus, the results do not support hypothesis three. However, the results should be
interpreted cautiously because the correlations are likely to be unstable since each correlation was based on only four observations per auditor.

[Insert Table 4 About Here]

**Self-Insight**

Self-insight was computed by correlating omega-squared statistics from the objective model with the subjective weights provided by each auditor. Because the omega-squared statistics for each individual cue were not normally distributed, Spearman rank correlation coefficients were employed. The t-test results, shown in Table 3, indicate a significant accountability effect based on raw correlations and a marginally significant effect based on transformed values. The mean self-insight for accountable and non-accountable auditors was .577 and .351 (based on raw correlations), respectively. These means are in the predicted direction and are consistent with hypothesis four. Again, the results should be interpreted cautiously because each correlation was based on only five observations per auditor.

**DISCUSSION**

The effect of many aspects of the auditor's naturally occurring environment on auditor judgment has been left largely unexplored in prior research. This study has been motivated by the need for additional research on these effects. In particular, the focus of the current study has been to provide evidence on the
role of accountability in individual auditor judgment. Accountability is an important part of the auditor's naturally occurring environment. Further, the existence of accountability might be expected to mitigate some of the documented shortcomings in auditor judgment [see Ashton et al., 1989 for a review]. The evidence presented here extends Ashton [1990a; 1990b] by considering a routine audit task and a broader set of judgmental characteristics.

Before discussing the results two limitations should be noted. First, while the sample of auditors appeared to be representative of third year auditors with the participating firm, the auditors were participating in in-house training scheduled by the accounting firm. Accordingly, because auditors were not randomly selected from the population of auditors inferences cannot be made to the larger population of auditors. However, auditors were selected so that their relative experience level would match task demands [Abdolmohammadi and Wright, 1987]. Second, the study employs a surrogate for the kind of accountability present in the auditors naturally occurring environment. Differences exist between accountability as manipulated in this study and as found by practicing auditors. In particular, the ability to account for one's work successfully may impact an auditor's job assignments, compensation package, and promotion decision. Such outcomes were not directly at risk in this study.

Turning to a discussion of the results, four hypotheses were tested. The results related to the first two hypotheses indicate
that accountable subjects: a) achieved significantly higher levels of consensus than non-accountable subjects, and b) achieved consistency levels not significantly different from non-accountable subjects. The results from the fourth hypothesis found that the level of self-insight was higher for accountable auditors than non-accountable auditors.

These results appear to suggest that although both accountable and non-accountable auditors applied a decision approach in a consistent fashion, accountability resulted in two important changes in decision outcomes. First, accountability appears to reduce idiosyncratic differences in decision approaches as indicated by greater consensus in judgments. That is, higher levels of judgment consensus would be expected to the extent that high levels of agreement existed on decision approaches.

Second, accountability was associated with improved self-insight. This finding suggests that when an explanation is expected, auditors anticipate and prepare by being more aware of their judgment processes. This suggests that self-monitoring activities are effortful and the extent to which they are applied by auditors reflects a cost-benefit analysis. When auditors were held accountable, the benefits from engaging in additional self-monitoring activities increased, and consequently self-insight improved.

The results of this study, along with Ashton's [1990a; 1990b] results demonstrate the importance of environmental factors to auditor judgments. Specifically, this line of research suggests that at least part of the shortcomings noted in auditor judgments
are mitigated by elements of the auditors existing environment. Environmental elements such as accountability, justification, and financial rewards, are all likely to affect the individual's motivation level. As a result, the individual may apply more effortful strategies in the natural environment than in the laboratory. In this regard, Beach and Mitchell's [1978] contingency theory predicts exactly such motivational effects. Thus, this research suggests that studies which ignore motivational factors or assume motivated subjects (based perhaps on self-report) should be interpreted with caution.

However, the incorporation of motivational factors should not be interpreted as a criticism of the cognitive research paradigm that has emerged in accounting. Instead, the two approaches should be complementary, each with their own region of validity [Tetlock, 1985b]. Different components of the judgment process may be subject to varying degrees of personal control [Abelson and Levi, 1985]. Tetlock, for example, emphasizes the difference between structural and control processes. The former are difficult to change and unlikely to be subject to conscious control (e.g., the capacity of short term memory) while the latter are much more flexible and apt to be subject to control (e.g., information search strategies). While the cognitive approach is most appropriate for investigating structural processes, research on control processes is likely to be enhanced by the consideration of economic-based or organizational theory-based perspectives [Ashton, 1990a].
ENDNOTES

1 The subjects analyzed in Ashton [1990b] are a subset of the subjects analyzed in Ashton [1990a].

2 The first two hypotheses are not exact replications since the present study applies the concept of accountability whereas the Ashton studies applied the concept of justification. However, both concepts lead to similar predictions.

3 The cues were presented in two orders. With the exception of one interaction, order did not significantly affect auditors' risk ratings. In order to simplify presentation of the results, this variable was collapsed when it was not significant.

4 Of the five subjects dropped, four were in the non-accountable group. One additional subject was dropped from the accountable group. This subject apparently misunderstood or did not take the task seriously. The subject's responses were negatively correlated with every other accountable subject's responses.
REFERENCES


**Table 1**

**Cues for Inventory Task**

<table>
<thead>
<tr>
<th>Cue</th>
<th>Low Risk Cue</th>
<th>High Risk Cue</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Industry sales for this product class have remained strong both in terms of volume and dollar amounts.</td>
<td>Industry sales for this product class have declined in both volume and dollar amounts in the third and fourth quarter of the current year.</td>
</tr>
<tr>
<td>2</td>
<td>A review of competitor catalogues revealed no other products with features similar to this one at a comparable price.</td>
<td>Two of the company's main competitors introduced a product early in the year which is very similar to this item but which is priced substantially lower.</td>
</tr>
<tr>
<td>3</td>
<td>There have been no changes in the retail price for the current year. The gross profit percentage for the item has remained approximately the same.</td>
<td>The retail price of the product was marked down by 40% of the original selling price in June of the current year. As a result, the gross profit percentage has fallen substantially.</td>
</tr>
<tr>
<td>4</td>
<td>The inventory turnover ratio for this item has remained approximately 4 throughout this year and the previous year.</td>
<td>The inventory turnover ratio for this item has declined substantially from the prior year. While the ratio was approximately 4 throughout last year, it has declined to below 2 in the third and fourth quarter of this year.</td>
</tr>
<tr>
<td>5</td>
<td>This item remains essentially the same as last year. Substantial space is dedicated to it in the company's catalogues.</td>
<td>During the year the company introduced an &quot;improved&quot; version of this item. Since then, this item (the old version) has received little space in the catalogues</td>
</tr>
</tbody>
</table>

The cues were presented in two randomly selected orders. The first order is shown in the table. The second order two was 4 3 2 5 1.
Table 2
Independent Variable Manipulation

<table>
<thead>
<tr>
<th>Accountable</th>
<th>Non-Accountable</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General Instructions</strong></td>
<td><strong>PLEASE NOTE</strong></td>
</tr>
<tr>
<td>Your responses will be reviewed by the researchers in conjunction with staff at the national office during the next week. These materials will be returned to you in small group breakout sessions next week at which time you will be asked to explain the reasoning behind your judgments. Each of you will be required to explain one or more of the judgments (chosen by the reviewers) made on this task. So that we may return these materials to you, please print your name on the following line.</td>
<td>Your responses to this task are completely anonymous. No one at your firm will see these materials and the researchers have no way of tying responses back to individuals.</td>
</tr>
<tr>
<td>Performance Measure</td>
<td>Non-Accountable Group</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Consensus</td>
<td>.542 (.194)</td>
</tr>
<tr>
<td>Consistency</td>
<td>.892 (.099)</td>
</tr>
<tr>
<td>Self Insight</td>
<td>.577 (.395)</td>
</tr>
</tbody>
</table>

Panel B: Means of Z-transformed correlations

<table>
<thead>
<tr>
<th>Performance Measure</th>
<th>Non-Accountable Group</th>
<th>Accountable Group</th>
<th>T-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consensus</td>
<td>.571</td>
<td>.462</td>
<td>11.65 ***</td>
</tr>
<tr>
<td>Consistency</td>
<td>.921</td>
<td>.914</td>
<td>.60</td>
</tr>
<tr>
<td>Self Insight</td>
<td>.722</td>
<td>.550</td>
<td>1.48 *</td>
</tr>
</tbody>
</table>

* p ≤ .1  ** p ≤ .05  *** p ≤ .01

The reported means of the z-transformed correlations are the back-transformed mean z scores.
### Table 4

#### Effects of Accountability on Stability of Judgments

**PANEL A. Analysis of Variance: Transformed Correlations**

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>F-statistic</th>
<th>p&lt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group (G)</td>
<td>1</td>
<td>.328</td>
<td>.57</td>
<td>.452</td>
</tr>
<tr>
<td>Order (O)</td>
<td>1</td>
<td>.015</td>
<td>.03</td>
<td>.874</td>
</tr>
<tr>
<td>G x O</td>
<td>1</td>
<td>3.347</td>
<td>5.83</td>
<td>.018</td>
</tr>
<tr>
<td>Error</td>
<td>89</td>
<td>51.129</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**PANEL B. Analysis of Variance: Raw Correlations**

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>F-statistic</th>
<th>p&lt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group (G)</td>
<td>1</td>
<td>.010</td>
<td>.11</td>
<td>.742</td>
</tr>
<tr>
<td>Order (O)</td>
<td>1</td>
<td>.003</td>
<td>.04</td>
<td>.351</td>
</tr>
<tr>
<td>G x O</td>
<td>1</td>
<td>.396</td>
<td>4.29</td>
<td>.041</td>
</tr>
<tr>
<td>Error</td>
<td>89</td>
<td>8.294</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**PANEL C. Stability Cell Means for Raw (Transformed) Correlations**

<table>
<thead>
<tr>
<th></th>
<th>Accountable Group</th>
<th>Non-Accountable Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Order 1</td>
<td>.787 (0.897)</td>
<td>.640 (.749)</td>
</tr>
<tr>
<td>Order 2</td>
<td>.672 (.805)</td>
<td>.784 (.883)</td>
</tr>
</tbody>
</table>

28
Figure 1
Risk Assessment Example

Inventory Item 139

1. Industry sales for this product class have remained strong both in terms of volume and dollar amounts.

2. Two of the company's main competitors introduced a product early in the year which is very similar to this item but which is priced substantially lower.

3. There have been no changes in the retail price for the current year. The gross profit percentage for the item has remained approximately the same.

4. The inventory turnover ratio for this item has remained approximately 4 throughout this year and the previous year.

5. This item remains essentially the same as last year. Substantial space is dedicated to it in the company's catalogues.

Please mark an X on the scale below corresponding to your assessment of the risk of obsolescence for this item.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1   2   3   4   5   6   7   8   9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Very Low
- Risk of
- Inventory
- Obsolescence
- Average
- Risk of
- Inventory
- Obsolescence
- Very High
- Risk of
- Inventory
- Obsolescence
NINTH SYMPOSIUM ON AUDITING RESEARCH

October 4-6, 1990

New Scholars Forum

Strategic Auditing and the Value of Internal Controls

by

Richard C. Sansing

Section A
Strategic Auditing
and the Value of Internal Controls

Richard Sansing*

August 1990

*Assistant Professor
Yale School of Organization and Management
Box 1A
New Haven, CT 06520-7368

This paper is based on Chapter 4 of my dissertation. I am grateful for the comments of committee members Paul Newman (chairman), Robert Freeman, Reed Smith, Paul Rothstein, and Richard Macminn. Financial support was provided by the Deloitte & Touche Foundation.
Strategic Auditing
and the Value of Internal Controls

ABSTRACT

This paper examines a hidden action model with fixed payoffs featuring an auditee and an auditor. The auditee chooses whether to comply with or violate a policy. The auditor chooses whether to investigate the auditee's behavior. The auditee prefers to comply when the auditor investigates and violate otherwise; the auditor prefers to investigate when the auditee violates and not investigate otherwise. Before the auditor makes his choice, he observes a signal that is imperfectly correlated with the auditee's action. This signal represents the output of the firm's internal control system, and an increase in the correlation between the signal and the auditee's action represents an improvement in internal controls.

In equilibrium, both the auditee and auditor use mixed strategies. Analysis of the model yields predictions regarding the relationship between the players' payoffs, the degree of correlation between the signal and the auditee's action, and the frequency of Type I and Type II errors. This analysis indicates that for certain parameter values, an improvement in the internal control system induces less compliance and more Type II errors.
1. Introduction

The interaction between an auditor and an auditee can be modeled as a game between players with conflicting interests. Recent papers have developed game-theoretic models that have focused on the effect of the players' payoffs on equilibrium strategies and error rates. The principal contribution of this paper is to examine the effect of the quality of the audit technology on strategies and error rates. In addition, this paper examines the value of the audit technology to the owner of the firm. These issues are considered in an internal auditing context, but the analysis could easily be applied to an external audit setting.

Shibano [1990] distinguishes between hidden action models and hidden information models. In a hidden action model, the auditee is supposed to take a particular action for reasons that are exogenous to the model. The action actually chosen by the auditee determines the probability distribution of a signal privately observed by the auditor. The set of distributions that could be induced by the auditee's action is called the audit technology. After observing the signal, the auditor accepts or rejects the null hypothesis that the auditee has taken the proper action. In a hidden information model, the auditee privately observes an outcome and issues a report. The auditor then observes a signal, the distribution of which is determined by the outcome, and tests the hypothesis that the auditee's report conforms to his observation. This paper analyzes a hidden action setting, not a hidden information setting.

Fellingham and Newman (FN) [1985] consider the difference between modeling the audit as two-player game as opposed to a single-person decision problem in a hidden action setting. They conclude that the game theory approach is conceptually superior, since the single-person decision theory model fails to consider the effect of the auditor on the auditee's actions. In that paper, the external auditor chooses whether to extend audit procedures (which would reveal the auditee's action with certainty) before issuing
an audit opinion. In this paper, the internal auditor observes a signal from the firm's internal control system (the audit technology) before deciding whether to investigate further.

In Newman and Noel (NN) [1989], the auditor observes a noisy signal before he accepts or rejects the auditee's reported balance. In that paper, the audit signal is the realization from a normal distribution with mean determined by the auditee's hidden action and an exogenously specified variance. In this paper, the audit signal is binary, revealing the auditee's action correctly with probability \( \pi \) (where \( 0.5 \leq \pi < 1 \)) and incorrectly with probability \( 1 - \pi \). Modeling the audit technology as a binary signal simplifies the analysis of the effects of changes in the audit technology's precision and thereby permits clearer insights into its effects.\(^1\)

Melumad and Thoman [1990] also examine the role of a utility-maximizing auditor with an imperfect audit technology. In their model, the firm's decision to hire an auditor signals its type to investors. In my model, the auditee's strategic choice is his compliance with or violation of a company policy. In their model, audit effort is costly, and the imperfect audit technology never makes Type I errors; in my model, auditing is costless, but both Type I and Type II errors can occur. Modeling the audit technology in this manner enables me to focus on the way the auditor's economic incentives affect the way information is used, rather than the extent to which information is gathered.

I analyze the strategic interaction between a manager who chooses whether to comply with a company requirement and an auditor with access to a noisy audit technology. The auditor has an incentive to discover whether the manager has complied, but also has an incentive to minimize the number of investigations. The focus of the analysis in this paper is the effect of the auditor's technology on the strategic choices

\(^1\) The audit technology in this model most closely resembles the arms control detection model in Brams and Davis [1987]; however, Brams and Davis did not use the Nash equilibrium solution approach in their analysis.
made by the auditor and auditee, which in turn affects the frequency of Type I and Type II errors that occur in equilibrium.

The remainder of this paper is organized as follows. Section 2 describes the setting. Section 3 describes the model. Equilibria are derived in section 4. Comparative statics are derived and the value of internal controls to the owners are calculated in section 5. Section 6 contains conclusions and discusses possible extensions.

2. Setting

The setting I consider features an auditee who either complies with or violates a company policy. Examples of violations are theft of company property and failure to follow proper procedures. An internal auditor chooses whether to make a detailed investigation into the auditee's behavior. Before making this choice, the auditor observes a signal generated by the company's internal control system. The signal reveals the auditee's action correctly with probability \( \pi, 0.5 \leq \pi < 1 \). An investigation is costly, but will reveal with certainty the auditee's action. If an investigation does not take place, the auditee's action will not be discovered.

The auditee prefers to comply with the policy if the auditor investigates and violate it if the auditor does not investigate. Since investigation is costly, the auditor prefers to investigate if the auditee has violated the requirement and not investigate if the auditee is in compliance. The auditee's gain from violating the requirement and not being investigated does not necessarily equal to his loss from violating the requirement and being investigated. Likewise, the auditor's loss from investigating when the auditee has complied is not necessarily equal to his loss from not investigating when the auditee has not complied.
3. Model

There are two players, an auditee (S) and an auditor (T). S chooses an action $\theta$, $\theta \in \{C,V\}$, where C represents compliance and V represents a violation. T then privately observes a signal $z$, $z \in \{C,V\}$, where $\text{Prob}[z = \theta] = \pi$, $1/2 \leq \pi < 1$. Note that if $\pi = 1/2$, the signal is useless, since it gives no indication whether or not S has complied; as $\pi$ approaches 1, the signal increasingly tends to reveal S's action. After observing $z$, T chooses an action $\omega$, $\omega \in \{N,Y\}$, where N indicates that T does not investigate and Y indicates that T does investigate.

The payoffs for each player are functions of the S's and T's actions, as shown below.

\begin{table}[h]
\centering
\begin{tabular}{ccc}
\hline
 & $\omega = N$ & $\omega = Y$ \\
\hline
$\theta = V$ & (K, L) & (-1, 0) \\
$\theta = C$ & (0,0) & (0, -1) \\
\hline
\end{tabular}
\caption{Payoffs for S and T}
\end{table}

I assume that $K > 0$, which ensures that S prefers to choose $\theta = V$ when $\omega = N$ and $\theta = C$ when $\omega = Y$. The parameter K represents the relative gain to S from violating the requirement when he is not investigated; the payoffs in the other cells are normalized to -1 and 0, without loss of generality. This normalization implies that an increase in K can represent either an increase in the gain from an undetected violation or a decrease in the loss from a detected violation. I also assume that $L > 0$, which ensures that T prefers to choose $\omega = Y$ when $\theta = V$ and $\omega = N$ when $\theta = C$. The parameter L represents the relative loss to T of failing to investigate when S is not in compliance, where once again the other payoffs are normalized to -1 and 0 without loss of generality. An increase in L represents an increase in the cost of an undetected violation or the
decrease in the loss of an investigation when S has complied. These assumptions regarding payoffs ensure that the interests of S and T conflict.

The players' strategy spaces are:

(i) an action rule for S, \( \theta \), and;

(ii) an investigation rule for T, \( \omega = \omega(z) \).

A solution to this game is a Nash equilibrium, in which each player's strategy is a best response to the other player's strategy. An Nash equilibrium in this game is defined as a pair of strategies \((\theta, \omega(z))\) such that

(a) \( \theta^* \) solves

\[
\max_{\theta} \sum_z \left[ U_S(\theta, \omega^*(z)) \right] \left( \text{Prob}[z|\theta] \right)
\]

(b) for each \( z, \omega^*(z) \) solves

\[
\max_{\omega} \sum_{\theta^*} \left[ U_T(\theta^*, \omega) \right] \left( \text{Prob}[\theta^*|z] \right)
\]

Since T observes the signal \( z \) before choosing whether to investigate, there are four pure strategies he can pursue. He can choose to not investigate regardless of \( z \), which is denoted strategy NN. He can choose to investigate regardless of \( z \), which is denoted strategy YY. He can choose to investigate when \( z = V \) and not investigate when \( z = C \), which is denoted strategy YN. Finally, he can choose to not investigate when \( z = V \) and investigate when \( z = C \), which is denoted strategy NY. The payoff matrix generated by combining the two players' pure strategies is shown in Table 2 below.
4. Equilibrium

An examination of the payoff matrix in Table 2 indicates that there is no pure strategy Nash equilibrium for this game. In general, S pursues a mixed strategy, choosing $\theta = V$ with probability $\lambda$, and $\theta = C$ with probability $(1 - \lambda)$. T also pursues a mixed strategy. With probability $\alpha$, T chooses strategy NN; with probability $\beta$, T chooses strategy YY; and with probability $\delta$, T chooses strategy YN. T never chooses strategy NY, as it is dominated by YN. T randomizes between NN and YN when $\pi \geq \left[ K/(K + 1) \right]$ and between YY and YN when $\pi \leq \left[ K/(K + 1) \right]$. (Recall that $K$ is the payoff to S when he is not in compliance and T does not investigate.) The equilibrium strategies are stated formally in Propositions 1 and 2; the proofs are in the appendix.

**Proposition 1:** Suppose $\pi \geq \left[ K/(K + 1) \right]$. Then the following strategies constitute a Nash equilibrium.

(a) With probability $\lambda$, S chooses $\theta = V$; with probability $(1 - \lambda)$, S chooses $\theta = C$, where $\lambda = \left[ (1 - \pi)/(1 - \pi + L\pi) \right]$;

(b) With probability $\delta$, T chooses YN; with probability $\alpha = (1 - \delta)$, T chooses NN, where $\delta = \left[ K/(\pi(K + 1)) \right]$. 

\[ \begin{array}{cccc}
\text{T} & \text{NN} & \text{YY} & \text{YN} \\
\theta = V & [K, L] & [-1,0] & [K(1 - \pi) - \pi, -L(1 - \pi)] \\
\theta = C & [0, 0] & [0, -1] & [0, \pi - 1] \\
\end{array} \]

(payoffs are in the order \([S,T]\))
Proposition 2: Suppose \( \pi \leq \frac{K}{(K + 1)} \). Then the following strategies constitute a Nash equilibrium.

(a) With probability \( \lambda \), \( S \) chooses \( \theta = V \); with probability \( 1 - \lambda \), \( S \) chooses \( \theta = C \), where \( \lambda = \frac{\pi/(\pi + L - L\pi)}{1/(1 - \pi)(K + 1)} \);
(b) With probability \( \delta \), \( T \) chooses \( YN \); with probability \( \beta = 1 - \delta \), \( T \) chooses \( YY \), where \( \delta = \frac{1}{1/(1 - \pi)(K + 1)} \).

5. Comparative Statics and Welfare Analysis

Differentiating the strategy functions of \( S \) and \( T \) with respect to the exogenous parameters \( K \), \( L \), and \( \pi \) yields insights into the effects of these parameters on the players' strategies and the game's outcome. First consider \( S \), who violates with probability \( \lambda \). When Proposition 1 applies:

\[
\begin{align*}
1) & \quad \frac{\partial \lambda}{\partial K} = 0 \\
2) & \quad \frac{\partial \lambda}{\partial L} < 0 \\
3) & \quad \frac{\partial \lambda}{\partial \pi} < 0.
\end{align*}
\]

Equation (1) indicates that increasing the gain from an undetected violation (or equivalently, reducing the penalty from a detected violation) has no effect on compliance.\(^2\) Equation (2) indicates that increasing \( T \)'s loss from not investigating when a violation occurs reduces the likelihood of a violation. The intuition behind (1) and (2) is that since \( S \) and \( T \) each play a mixed strategy, a change in \( T \)'s payoffs induces a change in \( S \)'s strategy, and vice versa. Finally, equation (3) indicates that an improvement in the audit technology induces more compliance by \( S \).

When Proposition 2 applies:

\[
\begin{align*}
4) & \quad \frac{\partial \lambda}{\partial K} = 0 \\
5) & \quad \frac{\partial \lambda}{\partial L} < 0 \\
6) & \quad \frac{\partial \lambda}{\partial \pi} > 0.
\end{align*}
\]

\(^2\) Note that this only holds locally; a sufficiently large increase in \( K \) would cause Proposition 2 to apply, which would result in an increase in the probability of a violation.
The intuition behind (4) and (5) is the same as it was in (1) and (2). In contrast, equation (6) indicates that an increase in π induces less compliance on the part of S.\(^3\) This result stems from the nature of the strategic interaction between S and T. When Proposition 2 holds, S chooses his compliance strategy so as to make T indifferent between YY and YN. As π increases, YN becomes a more attractive strategy to T if S’s strategy were fixed. To keep T indifferent, S increases his rate of noncompliance.

Next, consider the effect of variations in K, L, and π on T’s strategy. When Proposition 1 applies, T chooses strategy YN with probability δ, and strategy NN with probability \(\alpha = 1 - \delta\). Differentiating δ yields:

\[
\begin{align*}
(7) & \quad \frac{\partial \delta}{\partial K} > 0 \\
(8) & \quad \frac{\partial \delta}{\partial L} = 0 \\
(9) & \quad \frac{\partial \delta}{\partial \pi} < 0.
\end{align*}
\]

Equations (7) and (8) indicate the T responds to an increase in S’s gain from an undetected violation by relying more on the audit signal, but does not respond to a change in his own payoffs. As before, this is because the payoffs induce each player to make his opponent indifferent between strategies. Equation (9) indicates that an increase in the audit technology’s precision induces T to rely less on the audit signal in choosing whether to investigate. This is a logical consequence of (3); since S responds to an improvement in audit technology by increasing compliance, T should be more willing to not investigate even though \(z = V\). When Proposition 2 applies, T chooses strategy YN with probability \(\delta\), and strategy YY with probability \(\beta = 1 - \delta\). Differentiating δ yields:

\[
\begin{align*}
(10) & \quad \frac{\partial \delta}{\partial K} < 0 \\
(11) & \quad \frac{\partial \delta}{\partial L} = 0 \\
(12) & \quad \frac{\partial \delta}{\partial \pi} > 0.
\end{align*}
\]

Equation (10) indicates that as the rewards for undetected violations increase, T investigates more, i.e., T shifts from YN to YY. Equation (12) indicates that T responds to an increase in π by shifting from YY to YN.

---

\(^3\) As was the case in (1), this is only true locally; a sufficiently large increase in π will cause Proposition 1 to apply, which would increase compliance.
The average frequency of outcomes can be calculated directly from the best response strategies in Propositions 1 and 2. When Proposition 1 applies:

\[
\begin{align*}
(13) \quad \text{Prob}[\omega = Y, \theta = C] & \equiv \tau_1 = (1 - \lambda)[\delta(1 - \pi)] = \\
&= [(LK(1 - \pi))/((1 - \pi + L\pi)(K + 1))] \\
(14) \quad \text{Prob}[\omega = N, \theta = V] & \equiv \tau_2 = \lambda[1 - \delta\pi] = \\
&= [(1 - \pi)/((1 - \pi + L\pi)(K + 1))] \\
(15) \quad \text{Prob}[\omega = N, \theta = C] & \equiv \tau_3 = (1 - \lambda)[1 - \delta(1 - \pi)] = \\
&= [(L(2K\pi + \pi - K))/((1 - \pi + L\pi)(K + 1))] \\
(16) \quad \text{Prob}[\omega = Y, \theta = V] & \equiv \tau_4 = \lambda[\delta\pi] = \\
&= [((1 - \pi)K)/((1 - \pi + L\pi)(K + 1))]
\end{align*}
\]

Note that \(\tau_1\) corresponds to a Type I audit error (unnecessary investigation), \(\tau_2\) corresponds to a Type II audit error (undetected violation), \(\tau_3\) corresponds to compliance with no investigation, and \(\tau_4\) corresponds to a detected violation. Table 3 summarizes these definitions.

**Table 3**

<table>
<thead>
<tr>
<th>T</th>
<th>(\omega = N)</th>
<th>(\omega = Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\theta = V)</td>
<td>(\tau_2)</td>
<td>(\tau_4)</td>
</tr>
<tr>
<td>(\theta = C)</td>
<td>(\tau_3)</td>
<td>(\tau_1)</td>
</tr>
</tbody>
</table>

Differentiating \(\tau_1, \tau_2, \tau_3,\) and \(\tau_4,\) with respect to \(K, L,\) and \(\pi\) yields:

\[
\begin{align*}
(17) \quad \partial\tau_1/\partial K & > 0 \\
(18) \quad \partial\tau_2/\partial K & < 0 \\
(19) \quad \partial\tau_3/\partial K & < 0 \\
(20) \quad \partial\tau_4/\partial K & > 0 \\
(21) \quad \partial\tau_1/\partial L & > 0 \\
(22) \quad \partial\tau_2/\partial L & < 0 \\
(23) \quad \partial\tau_3/\partial L & > 0 \\
(24) \quad \partial\tau_4/\partial L & < 0 \\
(25) \quad \partial\tau_1/\partial \pi & < 0 \\
(26) \quad \partial\tau_2/\partial \pi & < 0 \\
(27) \quad \partial\tau_3/\partial \pi & > 0 \\
(28) \quad \partial\tau_4/\partial \pi & < 0.
\end{align*}
\]
The intuition behind (17)-(20) is that an increase in \( K \) induces \( T \) to substitute \( YN \) for \( NN \), which in turn causes an increase in \( \tau_1 \) and \( \tau_4 \) and a decrease in \( \tau_2 \) and \( \tau_3 \). Similarly, (21)-(24) indicate that an increase in \( L \) induces an increase in compliance on the part of \( S \), which in turn causes an increase in \( \tau_1 \) and \( \tau_3 \) and a decrease in \( \tau_2 \) and \( \tau_4 \).

The effects of \( \pi \) are more difficult to analyze, since a change in \( \pi \) affects the frequency of the various outcomes directly as well as indirectly via its influence on \( \lambda \) and \( \delta \). The component and net effects of an increase in \( \pi \) are shown in Table 4.

<table>
<thead>
<tr>
<th></th>
<th>( \lambda )</th>
<th>( \delta )</th>
<th>( \pi )</th>
<th>Net</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau_1 )</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( \tau_2 )</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( \tau_3 )</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>( \tau_4 )</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

Thus the effect of an increase in \( \pi \) is to decrease Type I and Type II errors, increase unaudited compliance, and decrease detected violations.

When Proposition 2 applies:

\[
\begin{align*}
(29) \quad \text{Prob}[\omega = Y, \theta = C] \equiv \tau_1 &= (1 - \lambda)[1 - \delta \pi] = \\
&= [L(1 + K - 2\pi - K\pi)/((\pi + L - L\pi)(K + 1))] \\
(30) \quad \text{Prob}[\omega = N, \theta = V] \equiv \tau_2 &= \lambda[\delta(1 - \pi)] = \\
&= [\pi/((\pi + L - L\pi)(K + 1))] \\
(31) \quad \text{Prob}[\omega = N, \theta = C] \equiv \tau_3 &= (1 - \lambda)[\delta \pi] = \\
&= [(L(1 - \pi))/((\pi + L - L\pi)(K + 1))] \\
(32) \quad \text{Prob}[\omega = Y, \theta = V] \equiv \tau_4 &= \lambda[1 - \delta + \delta \pi] = \\
&= [(\pi K)/((\pi + L - L\pi)(K + 1))].
\end{align*}
\]
Differentiating \( \tau_1, \tau_2, \tau_3, \) and \( \tau_4, \) with respect to \( K, L, \) and \( \pi \) yields:

\[
\begin{align*}
(33) \quad & \frac{\partial \tau_1}{\partial K} > 0 \\
(34) \quad & \frac{\partial \tau_2}{\partial K} < 0 \\
(35) \quad & \frac{\partial \tau_3}{\partial K} < 0 \\
(36) \quad & \frac{\partial \tau_4}{\partial K} > 0 \\
(37) \quad & \frac{\partial \tau_1}{\partial L} > 0 \\
(38) \quad & \frac{\partial \tau_2}{\partial L} < 0 \\
(39) \quad & \frac{\partial \tau_3}{\partial L} > 0 \\
(40) \quad & \frac{\partial \tau_4}{\partial L} < 0 \\
(41) \quad & \frac{\partial \tau_1}{\partial \pi} < 0 \\
(42) \quad & \frac{\partial \tau_2}{\partial \pi} > 0 \\
(43) \quad & \frac{\partial \tau_3}{\partial \pi} < 0 \\
(44) \quad & \frac{\partial \tau_4}{\partial \pi} > 0.
\end{align*}
\]

The comparative statics are identical to the ones derived in Proposition 1, except for the effect of \( \pi. \) Table 5 shows the component and net effects of an increase in \( \pi. \)

<table>
<thead>
<tr>
<th>( \lambda )</th>
<th>( \delta )</th>
<th>( \pi )</th>
<th>Net</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau_1 )</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( \tau_2 )</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>( \tau_3 )</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>( \tau_4 )</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

Thus an improvement in the audit technology can actually increase Type II audit errors due to the effect of the strategic interaction between the auditor and the auditee. In addition, an increase in \( \pi \) will decrease Type I audit errors, decrease unaudited compliance, and increase detected violations.

The final step in the analysis is to examine the effects of changes in \( \pi, K, \) and \( L \) on the value of the audit technology to the owners of the firm. I assume that both noncompliance by the auditee and investigations by the auditor are costly to the owners. I posit a utility function for the owner of \( U_O = -(\tau_2 + \tau_4) \cdot c(\tau_1 + \tau_4), c \geq 0. \) Note the loss from a violation is normalized to -1, and so the parameter 'c' determines the
relative costs of noncompliance and investigations. Using (1)-(6) and (17)-(28), when Proposition 1 holds:

\[(45) \, \frac{\partial U_O}{\partial K} < 0 \quad (46) \, \frac{\partial U_O}{\partial L} > 0 \quad (47) \, \frac{\partial U_O}{\partial \pi} > 0.\]

The intuition at work here is that since an increase in K induces more investigations but the same number of violations, the owner prefers for K to be low, i.e., he prefers relatively low rewards and high penalties for violations. Since an increase in L induces greater compliance but does not affect T's investigation strategy, the owner prefers that L be relatively high, i.e., Type II errors should be punished relatively more severely than Type I errors. Finally, since an increase in \(\pi\) decreases the frequency of both violations and investigations, the owner prefers that \(\pi\) be high.

When Proposition 2 holds:

\[(48) \, \frac{\partial U_O}{\partial K} < 0 \quad (49) \, \frac{\partial U_O}{\partial L} > 0\]

as was the case for Proposition 1. However, the sign of \(\frac{\partial U_O}{\partial \pi}\) depends on the whether

\[(50) \quad c \geq \frac{(K + 1)/(L + 1)}{1/2}\] or
\[(51) \quad c < \frac{(K + 1)/(L + 1)}{1/2}\).

If (50) holds, \(\frac{\partial U_O}{\partial \pi} \geq 0\); otherwise \(\frac{\partial U_O}{\partial \pi} < 0\). Thus as long as \(\pi\) is low (i.e., as long as \(\pi \leq [K/(K + 1)]\)) and \(c\) is low (i.e., (51) applies), then an improvement in the internal control system will make the owner worse off, even if improvements are costless.

6. Conclusion

The key insights in this paper are that when the precision of the audit technology is sufficiently low, an increase in the precision induces less compliance with legal requirements (6), less reliance on the technology (10), and more Type II audit errors.
Furthermore, when the cost of investigation to the owner is sufficiently low relative to the cost of violations, costless improvements in the audit technology make the owner of the firm worse off. These predictions, which are driven by the strategic interaction between the auditor and auditee, are different from those that would be predicted using a more traditional single-person decision theory paradigm.

One possible extension to this paper is to give S a wider set of action choices instead of the binary choice of compliance or violation. For example, the magnitude of the violation could be a choice variable, where a greater violation could increase S's reward but also increase the chance of detection by the accounting system.

A second extension could be to derive the payoffs in Table 1 endogenously by making the conflict modeled here part of a larger game. It is open question whether the results derived in this partial equilibrium setting hold in a general equilibrium setting.
APPENDIX

Proof of Proposition 1: To prove Proposition 1, I must show that each player's strategy satisfies the best response property. Since both strategies are mixed, I must show $S$ is indifferent between $\theta = V$ and $\theta = C$ given $T$'s strategy, $T$ is indifferent between $YN$ and $NN$ given $S$'s strategy, and $YY$ is dominated given $S$'s strategy. Given $S$'s and $T$'s strategy and the payoffs in Table 1,

\begin{align*}
(52) \quad & E[U_S|\theta = V] = [\delta \pi(-1) + (1 - \delta \pi)(K)] \\
(53) \quad & E[U_S|\theta = C] = 0 \\
(54) \quad & E[U_T|YN] = [\lambda(1 - \pi)(-L)] + [(1 - \lambda)(1 - \pi)(-1)] \\
(55) \quad & E[U_T|NN] = -\lambda L. \\
(56) \quad & E[U_T|YY] = -(1 - \lambda). \\
\end{align*}

Equating (52) and (53) and solving for $\delta$ yields

\begin{align*}
(57) \quad & \delta = [K/(\pi(K + 1))]. \\
\end{align*}

Since $\pi \geq [K/(K + 1)]$, $0 \leq \delta \leq 1$. Thus $S$ is indifferent between $\theta = V$ and $\theta = C$ given $T$'s strategy. Equating (54) and (55) yields:

\begin{align*}
(58) \quad & \lambda = [(1 - \pi)/(1 - \pi + \pi L)]. \\
\end{align*}

Thus $T$ is indifferent between $YN$ and $NN$ given $S$'s strategy. Substituting this value of $\lambda$ into (55) and (56) and simplifying shows

\begin{align*}
(59) \quad & E[U_T|NN] \geq E[U_T|YY]. \\
\end{align*}

Thus I have shown that $YY$ is dominated given $S$'s strategy. Q.E.D.
Proof of Proposition 2: To prove Proposition 2, I must show that each player's strategy satisfies the best response property. Since both strategies are mixed, I must show $S$ is indifferent between $\theta = V$ and $\theta = C$ given $T$'s strategy, $T$ is indifferent between $YN$ and $YY$ given $S$'s strategy, and $NN$ is dominated given $S$'s strategy. Given $T$'s strategy and the payoffs in Table 1,

\begin{align*}
(60) \quad & E[U_S|\theta = V] = [\delta(1 - \pi)(K) + (1 - \delta + \delta \pi)(-1)] \\
(61) \quad & E[U_S|\theta = C] = 0 \\
(62) \quad & E[U_T|YN] = [\lambda(1 - \pi)(-L)] + [(1 - \lambda)(1 - \pi)(-1)] \\
(63) \quad & E[U_T|YY] = -(1 - \lambda) \\
(64) \quad & E[U_T|NN] = -\lambda L
\end{align*}

Equating (60) and (61) and solving for $\delta$ yields

\[ (65) \quad \delta = \frac{1}{(1 - \pi)(1 + K)}. \]

Since $\pi \leq [K/(K + 1)], 0 \leq \delta \leq 1$. Thus $S$ is indifferent between $\theta = V$ and $\theta = C$ given $T$'s strategy. Equating (62) and (63) yields:

\[ (66) \quad \lambda = \frac{\pi/(\pi + L - \pi L)}. \]

Thus $T$ is indifferent between $YN$ and $YY$ given $S$'s strategy. Substituting this value of $\lambda$ into (63) and (64) and simplifying shows

\[ (67) \quad E[U_T|YY] \geq E[U_T|NN]. \]

Thus I have shown that $NN$ is dominated given $S$'s strategy. Q.E.D.
REFERENCES


NINTH SYMPOSIUM ON AUDITING RESEARCH
October 4-6, 1990

New Scholars Forum

Belief Perseverance in Audit Analytical Review
by
Lisa L. Koonce

Section B
BELIEF PERSEVERANCE IN
AUDIT ANALYTICAL REVIEW

Lisa Koonce

Department of Accounting
College of Business Administration
The University of Texas at Austin

June, 1990

I am grateful to Urton Anderson, Clif Brown, Bill Kinney, Garry Marchant, Jim McKeown, and Ira Solomon for their helpful discussions on this paper. Also, the workshop participants at Duke University, Ohio State University, University of Illinois, University of Texas, and University of Arizona provided many useful comments. Funding was provided by the University of Illinois and by the Ernst & Whinney and Touche Ross & Co. Foundations.
BELIEF PERSEVERANCE IN
AUDIT ANALYTICAL REVIEW

Auditors often perform the diagnostic process of analytical review to identify potential financial-statement errors, irregularities and other unusual events. Although research suggests that analytical review is beneficial for identifying unexpected fluctuations in account balances and financial ratios, little research exists on its effectiveness and efficiency for determining the causes of unexpected fluctuations. The primary purpose of this study, therefore, was to provide experimental evidence on the conditions under which auditors performing analytical review may be prone to compromising audit effectiveness by accepting a plausible, but incorrect, non-error cause for an unexpected fluctuation.

Within the framework of belief-perseverance theory, the effects of explanation, counterexplanation, and their order of performance were studied to determine how such factors affect the propensity for acceptance of a plausible, but incorrect, cause. Investigating explanation is important since auditors typically document explanations for analytical review findings. Even though current auditing standards do not require auditors to counterexplain, understanding how counterexplanation and the order of counterexplanation and explanation affect auditors’ judgments is nevertheless important since it enhances our understanding of the potential for acceptance of an incorrect cause.

Consistent with belief-perseverance theory, auditors revised their beliefs about an hypothesized non-error cause of an unexpected fluctuation in the auditee’s gross margin in an upward (downward) fashion when asked to document an explanation (counterexplanation) for that cause. Inconsistent with theory, however, auditors who were asked to initially document an explanation and then a counterexplanation judged the non-error cause to be less likely to have occurred than did auditors asked to initially document a counterexplanation and then an explanation. That is, the experimental data were suggestive of a recency, rather than the predicted primacy, order effect. Based on additional experimental evidence that initially engaging in counterexplanation was significantly more difficult, it is possible that explanation followed by counterexplanation is the natural order of comprehension and that any other task-performance order will yield judgments less impacted by the available counterexplanation evidence.

Overall, these results suggest once auditors develop written explanations (counterexplanations) for non-error causes, they are more likely (less likely) to compromise audit effectiveness by accepting an incorrect cause than are auditors who do not write such explanations (counterexplanations). In addition, the results suggest that auditors who initially document counterexplanations and then document explanations are more likely to compromise audit effectiveness by accepting an incorrect cause than are auditors who perform the same two tasks in the opposite order (i.e., explain and then counterexplain).
BELIEF PERSEVERANCE IN AUDIT ANALYTICAL REVIEW

Analytical review, the diagnostic process of identifying and determining the cause of unexpected fluctuations in account balances and financial ratios, has become an increasingly important part of auditing. Recent research suggests that analytical review is beneficial for detecting such unexpected fluctuations (Kreutzfeldt and Wallace 1986, Wright and Ashton 1989). It is, however, difficult for an auditor utilizing analytical review to know when he/she has determined the "correct" cause for an observed fluctuation, especially since no official guidance exists to assist the auditor. Thus, accepting a plausible, but incorrect, cause is a dangerous, and largely unresearched, possibility (Kinney 1987).

This paper describes two experiments that investigate conditions under which auditors may be prone to accepting a plausible, but potentially incorrect, cause when performing attention-directing analytical review. Identifying such conditions is important, because either audit effectiveness or efficiency would be compromised if such incorrect acceptance were to occur. Although there are no normatively correct causes for the analytical review problem situations employed in this study, relative evidence on such compromises will be provided. Such evidence will be obtained by investigating conditions which lead auditors to make different judgments about the cause of an unexpected fluctuation. Differences in auditors' judgments are important since they can lead to differences in decisions as to the cause of an unexpected fluctuation and subsequent auditor actions (cf. Asare 1990).
By dichotomizing causes into error and non-error classes,¹ there are several ways in which the auditor may accept a plausible, but incorrect, cause. One has especially important potential audit effectiveness implications: concluding that a non-error cause adequately accounts for an unexpected fluctuations when the "correct" cause involves a financial-statement error. Since it is unlikely that the auditor would increase planned tests of details if indications of a material financial-statement error were not present (and it is possible that the auditor may decrease such testing), the likelihood of subsequent detection of the financial-statement error would be reduced in this situation.² Conditions that may lead to this type of decision error will be investigated in this paper.

Since analytical review is characterized by sequential belief revision and causal reasoning (Gibbins 1984, Libby 1985), an appropriate theoretical framework for studying this incorrect-acceptance issue also must incorporate such factors. Based on such criteria, belief-perseverance theory was chosen to guide the investigation of the potential for acceptance of a plausible, but incorrect, non-error cause. Belief perseverance is a social-cognition theory which predicts that when information is sequentially acquired and causally evaluated, information considered first will have a greater effect on an individual's final (revised) belief than information considered later (i.e., the initial belief perseveres).

The remainder of this paper is organized as follows. Section I provides an overview of belief-perseverance theory and its relation to the environment of analytical review. Specifically, three factors suggested to affect the potential for belief perseverance -- explanation, counterexplanation, and the order in which explanation and counterexplanation occur -- are linked to the analytical review process. The experimental hypotheses also are developed in this first section. Section II of the paper provides a description of the
method and results of the first experiment which investigates the impact of explanation on auditors’ analytical review judgments. Section III presents the method and results of the second experiment. This experiment investigates the impact of counterexplanation and task-performance order (i.e., explanation-then-counterexplanation versus counterexplanation-then-explanation) on auditors’ analytical review judgments. The final section of the paper provides a discussion of the results and concluding comments.

I. THEORY AND HYPOTHESES

The Explanation Effect

When performing analytical review at the planning stage of the audit, the auditor initially determines an expectation for the account balance or financial ratio under investigation. This expectation assists the auditor in identifying any unexpected fluctuation. Once such a fluctuation has been identified, the auditor typically will secure a potential cause from client management (AICPA 1988 para. 21, Wright and Ashton 1989) and may eventually document an explanation supporting that cause in the audit workpapers. Because antecedent conditions (i.e., conditions that could have led to the unexpected fluctuation) must be considered when written explanations are required, such explanations should increase the extent of causal reasoning employed by the auditor relative to situations in which written explanations are not required.

Due to this increased causal reasoning, an event which is part of a written explanation is typically judged more likely to have occurred than an event which is not part of a written explanation. This phenomenon has been termed an explanation effect (Anderson and Sehler 1986) and has been experimentally linked to belief perseverance
(Anderson, et al. 1980). That is, individuals who develop written explanations are more likely to maintain their existing beliefs (i.e., their beliefs are more likely to persevere) even when subsequent data challenge those beliefs than are individuals who have not developed written explanations.

The foregoing cognitive mechanism is congruent with the auditor's sequential belief-revision process for analytical review. That is, since auditors frequently document explanations when performing analytical review, an explanation effect is expected when they develop written explanations for an hypothesized cause of an unexpected fluctuation. If an explanation effect is observed, it would suggest that auditors who write explanations would be more prone to accepting an incorrect cause than auditors who do not write such explanations.

This explanation-effect prediction is made even though the effect has been documented only for predictive-inference tasks. Since the underlying cognitive mechanism promoting the explanation effect does not appear to be domain specific, there is little reason to suspect that it also will not occur using a diagnostic-inference task such as analytical review. This explanation-effect prediction also is made even in light of prior research which has shown that the phenomenon does not occur using subjects experienced with the task. Specifically, Anderson and Wright (1988) found that experienced auditors' judgments about the likelihood of various payroll system events (payroll-entry error; foreman fraud) are not affected by whether auditors explain how those events could happen. These results suggested to Anderson and Wright that auditors have stored in long-term memory previously generated causal scenarios for payroll system events. Importantly, these causal scenarios are recalled from memory for both pre- and post-explanation likelihood
judgments. Consequently, written explanations are suggested to have little or no impact on likelihood judgments.

It is possible, then, that the auditor may have stored in long-term memory previously generated causal scenarios for potential causes of various unexpected fluctuations. However, Pei, et al. (1990) suggest that for complex tasks such as analytical review, it is unlikely that the auditor would have such causal scenarios (and they report experimental results consistent with this suggestion). Thus, in light of psychological research on the explanation effect and the Pei, et al. findings, the following alternative-form hypothesis is proposed:

**Hypothesis 1:** Auditors who are asked to generate a written explanation for an hypothesized non-error cause will assign a higher probability to that cause than auditors who are not asked to generate a written explanation.

**The Counterexplanation Effect**

Even though the causal reasoning that occurs when written explanations are developed should enable the auditor to acquire a more-reasoned judgment about the validity of the hypothesized cause, it also may cause him/her to prematurely cease information search and accept a seemingly plausible, but incorrect, cause. Indeed, prior research has shown that the extent to which an individual spontaneously incorporates disconfirmatory information into his/her judgment is reduced once a written explanation has been developed (e.g., Anderson 1982, Anderson and Sechler 1986). In contrast, when an individual is explicitly asked to consider disconfirmatory evidence, or to counterexplain, information not favoring a target event becomes more developed and salient in memory, thus enhancing the perceived multiplicity of causation and attenuating the individual’s belief in the validity of
the target event (Einhorn and Hogarth 1986). In other words, the potential for belief perseverance is reduced once explicit counterexplanation occurs.

In the audit domain, Heiman (1990) demonstrates that once auditors consider alternative error causes or counterexplanation evidence during analytical review, they reduce their probability assessments for an hypothesized financial-statement error cause. That is, Heiman demonstrates a counterexplanation effect. There are, however, several important differences between Heiman’s experimental procedure and the procedure to be employed in this research. First, the hypothesized cause in Heiman’s study involved a financial-statement error, while the hypothesized cause in this study will involve a non-error cause. Since clients probably are more inclined to suggest non-error causes for unexpected fluctuations during auditor inquiry, this study will not only provide a more-realistic portrayal of analytical review but it will test the generalizability of the counterexplanation effect. Second, while subjects in this study will be required to document explanations for the hypothesized cause, subjects in Heiman’s study were not asked to perform this explanation procedure. Requiring subjects to develop a written explanation will yield a more-stringent test of the counterexplanation effect, since subjects should be more committed to the (explained) hypothesized cause and less likely to revise their beliefs once explicit counterexplanation occurs.

In summary, extant audit research suggests that auditors may not fully incorporate the potential impact of disconfirmatory evidence into their judgments when performing analytical review unless they are explicitly requested to counterexplain. Even by providing a more-stringent test of the counterexplanation effect in the present study, the development and salience and, therefore, importance attributed to counterexplanation information should
be greater for an auditor who is requested to document a written counterexplanation. Thus, the following alternative-form hypothesis is proposed:

**Hypothesis 2:** Auditors who are asked to develop a written explanation for an hypothesized non-error cause will assign higher probabilities to that cause than auditors who are asked to develop a written explanation and a written counterexplanation.

**The Primacy Order Effect**

Although development of a written counterexplanation is expected to reduce the auditor's potential for belief perseverance (and, thus, the likelihood that he/she will accept a plausible, but incorrect, cause), various approaches to counterexplanation have been suggested to differentially reduce the potential for belief perseverance. In particular, belief-perseverance research indicates that the order in which explanation and counterexplanation procedures are employed can differentially affect target-event probability assessments (Hoch 1984, Moser 1989, and cf. Tetlock 1983) which, in turn, can differentially affect the potential for belief perseverance. When task-performance order is explanation-then-counterexplanation, belief-perseverance theory predicts that final probability assessments regarding the occurrence of the target event will be higher than if task-performance order were counterexplanation-then-explanation. This pattern of results is consistent with a primacy order effect (cf. Anderson 1981) and suggests that auditors who explain and then counterexplain will be more prone to accepting an incorrect cause than will auditors who follow the opposite task-performance order.  

The primacy order effect occurs because initially developing a target-event explanation frequently causes an individual to overweight that explanation relative to the opposite task-performance order (Nisbett and Ross 1980, Tetlock 1983). In some cases,
such initial generation of an explanation also can cause the individual to be less able and/or willing to subsequently generate counterexplanation arguments (Anderson, et al. 1985, Hoch 1984, Moser 1989). In contrast, when information not favoring a target event is initially considered, uncertainty surrounding the validity of the target event is greater since the perceived multiplicity of causation has been enhanced before a written explanation is constructed (Anderson 1982). Consequently, individuals who initially counterexplain and then explain tend to overweight (underweight) their counterexplanations (explanations) relative to individuals employing the opposite task-performance order.

Based on belief-perseverance theory, auditors who initially engage in explanation followed by counterexplanation are expected to assess higher final probabilities for a non-error cause (and, thus, be more prone to accepting an incorrect cause) than auditors who initially engage in counterexplanation followed by explanation. This primacy order effect is predicted even though other auditing research has consistently reported recency order effects for auditors' sequential processing of information (e.g., Ashton and Ashton 1988, Tubbs, et al. 1990). Recency occurs when information has a greater impact on beliefs when it is evaluated last, rather than first, in a two-task sequence.

The theoretical framework used for these audit studies, none of which were conducted within the analytical review context, was Hogarth and Einhorn’s (1990) belief-adjustment model. Two aspects of the Hogarth and Einhorn model make it incongruent with audit analytical review and, thus, preclude its use in this study. First, the model is based on a context in which all information items are presented to subjects who are typically requested to revise their judgments after receipt of each independent information item. However, information is typically not presented in concise, summary form to auditors
performing analytical review. Rather, auditors must integrate information items obtained from multiple sources by making assumptions and inferences among those (typically non-independent) information items. Second, Hogarth and Einhorn’s model defines disconfirmatory evidence only as evidence about conditions that might cancel or prevent the target event from causing the effect; the model does not allow disconfirmatory evidence to involve alternative potential causes. Since consideration of evidence about alternative possible causes is an important element of analytical review (Heiman 1990), the belief-adjustment model provides an incomplete representation of the analytical review environment.

Since belief-perseverance theory is not constrained by the aforementioned two factors, it appears to be more suitable than the belief-adjustment model for studying order effects within audit analytical review. Consequently, the following alternative-form hypothesis is proposed:

**Hypothesis 3:** Auditors who initially are asked to develop a written explanation and then develop a written counterexplanation will assess higher probabilities for the hypothesized non-error cause than auditors who initially are asked to develop a written counterexplanation and then develop a written explanation.

Two experiments test the three experimental hypotheses. The purpose of experiment 1 is to determine how developing a written explanation for a non-error cause affects auditors’ probability assessments for that cause (hypothesis 1). The purpose of experiment 2 is to determine how developing a written counterexplanation affects the probabilities assigned to the non-error cause (hypothesis 2) and how task-performance order also affects these probability judgments (hypothesis 3). Two experiments are necessary since the case materials for testing hypotheses 2 and 3 must include potentially disconfirmatory, or
counterexplanation, information while the case materials for testing hypothesis 1 must exclude such information. The latter exclusion is necessary to avoid the possibility of the subjects spontaneously incorporating counterexplanation evidence into their judgments and, thus, masking the effects of explanation.

II. EXPERIMENT ONE -- THE EXPLANATION EFFECT

Materials. There were three components to the case materials. Part A described a first-year manufacturing client to which subjects were assigned the task of performing attention-directing analytical review. In addition, Part A contained financial and non-financial background information about the client and its industry. Part B of the case materials contained narratives summarizing prior discussions with various client personnel. These narratives contained information pertaining to (but not perfectly diagnostic of) the non-error cause (change in sales mix) that the client subsequently would suggest led to an unexpected fluctuation in the gross margin. Two different text orders of the discussion narratives were utilized to avoid the potential limitation of a particular text order systematically affecting the results. Finally, Part C of the case materials contained either four or five questions (including a post-experimental questionnaire) depending on the experimental condition.

Subjects. Participants were 34 senior-level auditors from five international public accounting firms with experience ranging from 1.7 to 5.0 years (average experience of 2.8 years). The use of audit seniors is justified, since they typically perform the attention-directing analytical review task employed in this study. In addition, audit seniors with 2-5 years of audit experience would be expected to have reasonably well-developed knowledge
of potential (error and non-error) causes of unexpected fluctuations (Libby and Frederick forthcoming). The experiment was administered to the subjects in their offices during regular business hours. On average, subjects required about 60 minutes to complete experiment 1.

**Design and Procedure.** To test for the effects of explanation on audit judgments, a 2x1 design is used. The independent variable is when the explanation for the hypothesized non-error cause is written (i.e., prior to or after the elicitation of an initial probability for the non-error cause) and is manipulated between subjects. The dependent variable is the subject’s assessment of the probability that the non-error cause accounts for substantially all of the gross margin fluctuation. "Substantially all" was defined in the case as the non-error cause accounting for a specified dollar amount (approximately 90%) of the gross margin increase.

Figure 1 depicts the procedure for experiment 1. Initially, the subjects were given Parts A and B of the case materials and were told to read and think about the case materials. Subsequent to an initial computation task (in Part C) designed to familiarize the subjects with the financial information, all subjects were told to focus on an unexpected increase in the gross margin from the prior year (i.e., the unexpected fluctuation). After learning the client’s suggested cause (i.e., change in sales mix) for the gross margin increase, subjects assigned to the treatment condition were asked to develop a written explanation documenting how the client’s cause could have accounted for substantially all of the gross margin increase. Importantly, the subjects were told that their explanation should be regarded as a "pre-workpaper explanation." This instruction was intended to avoid potential confounding due to justification effects (cf. Gibbins 1984).
After the subjects completed their explanation, they were instructed to assess the probability that the hypothesized non-error cause suggested by the client accounted for substantially all of the gross margin increase. This probability assessment question is shown below.\(^6\)

Consider [the client's] statement about the gross margin increase. Based on your assessment of the evidence collected so far, please assess your judgment of the likelihood that a change in sales mix accounts for \textit{substantially all} of the gross margin increase. That is, how likely do you think it is that a change in sales mix (alone) accounts for \textit{at least} $3,300,000 of the $3,667,250 gross margin increase?

After learning the client’s suggested cause for the gross margin increase, subjects assigned to the control condition were asked to make a probability judgment about the non-error cause (i.e., were asked to respond to the probability-assessment question presented above) before being asked to develop a written explanation for that cause. Next, the control-condition subjects were asked to provide a written explanation and to make a second, post-explanation probability assessment. The second probability question was similar to the initial probability question. This second probability assessment provided a \textit{supplementary} within-subjects test of the explanation effect. The purpose of the latter test was to determine whether anchoring occurs when probabilities are elicited on a within-subjects basis (Tversky and Kahneman 1974).

**Preliminary Analysis.** Preliminary data analysis included a search for possible firm effects, text-order effects, and written responses clearly indicating an inability to successfully explain the non-error cause. No firm or text-order effects were noted for any of the
probability measures (smallest p-value=.1353, two-tailed F tests). Of the 34 written non-error explanations obtained from the subjects (17 in each condition), eight were judged to be unsuccessful attempts at explaining the non-error cause accounting for substantially all of the gross margin increase. One subject made a computational mistake and five subjects considered the wrong products, thus precluding them from explaining an amount over the judgment threshold. The remaining two subjects made logic errors which caused them not to believe that the change in sales mix could account for substantially all of the gross margin increase. The following analysis, then, is based on data from 26 subjects (13 subjects each in the two experimental conditions). The appendix contains a representative explanation from the 26 explanations.

**Test of Hypothesis 1.** Table 1 presents the mean probability judgments assessed by the auditors in the control and treatment conditions. Both the between- and within-subjects tests of the explanation effect are presented in Table 1.

![Insert Table 1 about here]

Two t-tests were performed. The first involved a between-subjects t-test for which the independent variable was the explanation condition (control versus treatment). The test compared the initial, pre-explanation probability provided by the control-condition subjects to the post-explanation probability from the treatment-condition subjects. The t-test results indicate that the final probability assessment of 76.62 provided by the treatment-condition subjects was significantly greater than the initial, pre-explanation probability of 64.62 provided by the control-condition subjects (t=2.44, p=.0112). Thus, the null version of hypothesis 1 can be rejected in favor of the alternative which predicted that auditors would
assess higher probabilities for the hypothesized non-error cause after they had developed a written explanation for that cause.

The second t-test involved a within-subjects test for which the independent variable was the timing of the probability assessment by the control-condition subjects. That is, it compared the initial, pre-explanation and final, post-explanation probability assessments provided by those subjects. The results of that paired t-test indicated that the final probability assessment of 75.62 was significantly greater than the initial, pre-explanation probability of 64.62 \((t=3.26, p=.0035)\). Thus, like the between-subjects analysis, this within-subjects test also showed a significant effect for explanation. Further, this finding suggests that the within-subjects probability elicitation did not lead to significant anchoring effects.

Consistent with psychological research but inconsistent with prior auditing research, instructing the auditor-subjects to develop a written explanation for the non-error cause led them to hold stronger beliefs for that cause, relative to the beliefs of auditors who were not asked to develop a written explanation. This finding suggests that auditors do not have stored in memory previously generated scenarios for potential causes of unexpected fluctuations. Rather, such scenarios are constructed when written explanation occurs.

III. EXPERIMENT TWO -- THE COUNTEREXPLANATION EFFECT AND THE PRIMACY ORDER EFFECT

Materials. The case materials used for the second experiment were similar to those used for the first. The primary differences pertained to the client-discussion narratives and case questions. Unlike experiment 1, the client-discussion narratives for experiment 2 contained information pertaining to other potential causes (primarily financial-statement
errors) of the gross margin increase as well as information pertaining to conditions that could cancel or negate the effects of the change in sales mix. That is, they contained potential disconfirmatory, or counterexplanation, information. Furthermore, questions pertaining to counterexplanation were incorporated into the second experiment's case materials.

Subjects. Participants in the second experiment were 53 "senior" auditors from five international public accounting firms. These auditors had an average of 3.2 years of audit experience with a range of 1.9 to 5.3 years. The experiment was administered to the subjects in their offices during regular business hours. On average, subjects required about 75 minutes to complete experiment 2.

Design and Procedure. A one-way design is employed for this experiment, for which the independent variable -- type of explanation procedure employed -- is manipulated between subjects. The three levels of this independent variable are: (1) explanation-only (E), (2) explanation-then-counterexplanation (E/CE), and (3) counterexplanation-then-explanation (CE/E). The dependent variable is the same as that used for the first experiment -- the subject's assessment of the probability that the non-error cause offered by the client accounts for substantially all of the unexpected fluctuation.

Figure 2 depicts the procedure for experiment 2. Subjects were initially asked to read the background information in Part A of the case materials. Next, subjects were instructed to move ahead to Part C of the case materials which contained the case questions. Subsequent to an initial computation task, all subjects were told to focus on an unexpected increase in the gross margin from the prior year (i.e., the unexpected fluctuation) and were told the client's suggested cause for the fluctuation -- a change in sales
mix. Next, subjects were given written narratives documenting their prior "discussions" with various client personnel (i.e., Part B of the case materials).

---

Insert Figure 2 about here
---

After reading the discussion narratives, subjects assigned to the explanation-only condition were asked to document arguments supporting the non-error cause (change in sales mix) producing substantially all of the gross margin increase. Once the change-in-sales-mix explanation was written, the explanation-only subjects were asked to assess the probability that this cause accounted for substantially all of the gross margin increase. The probability assessment question was identical to that from experiment 1. As a check on the perceived multiplicity of causation, subjects were subsequently asked to assess the probability that a material financial-statement error could have resulted because of each of seven types of errors. Because of the potential for multiple causes (Shaklee and Fischhoff 1982), the subjects were told that their non-error and error probabilities need not sum to 1.00 (cf. Einhorn and Hogarth 1985 and Robinson and Hastie 1985).

After reading the client-discussion narratives, subjects assigned to the explanation-then-counterexplanation condition were asked to document arguments supporting the change in sales mix causing substantially all of the gross margin increase. After completing this task, subjects were instructed to document arguments that did not support the non-error cause accounting for substantially all of the gross margin increase. The latter arguments were noted to include evidence about alternative potential causes (error or non-error) of the gross margin increase and/or evidence pertaining to conditions that could cancel or negate the effects of the change in sales mix. After the explanation and counterexplanation tasks,
subjects assessed the non-error cause probability and the seven financial-statement error probabilities. Subjects assigned to the counterexplanation-then-explanation condition followed essentially the same procedure, except for the reversal of the explanation and counterexplanation tasks.

Preliminary Analysis. Preliminary data analysis included a search for possible firm effects, text-order effects, and written responses clearly indicating an inability to successfully explain the non-error cause. No firm or text-order effects were noted for any of the probability measures (smallest p-value=.6347, two-tailed F tests). Using the same criteria employed in experiment 1 for judging a successful explanation, seven out of 53 responses were judged to be unsuccessful attempts at explaining the change in sales mix accounting for substantially all of the gross margin increase. Of these seven subjects, three did not identify any components to the change in sales mix nor did they perform calculations. Their written comments clearly indicated that they did not believe that the change in sales mix could account for substantially all of the gross margin increase. Two subjects excluded products from the change-in-sales-mix computation and, thus, were unable to explain an amount over the judgment threshold. The usage of an incorrect product price and a computational error precluded the remaining two subjects from successfully explaining the change in sales mix. After excluding these seven non-explainers from the sample, forty-six subjects remained with fifteen each in the E and CE/E conditions and sixteen in the E/CE condition. The appendix illustrates a representative explanation and counterexplanation from the 46 responses.

Tests of Hypotheses 2 and 3. Table 2 presents both the descriptive statistics for the non-error probabilities assessed by auditors in the three experimental conditions and the
planned comparisons for purposes of testing hypotheses 2 and 3. The planned comparison for hypothesis 2 shows that the mean probability assessment from the explanation-only auditors (77.20) was significantly higher than the mean probability assessment from the counterexplanation auditors (59.65) (F=10.00, p=.0015). Thus, the null version of hypothesis 2 can be rejected in favor of the alternative which suggests that auditors’ probability judgments will be reduced once they are explicitly requested to counterexplain.

------------------------------------------------------------------------
Insert Table 2 about here
------------------------------------------------------------------------

Further analysis indicates that subjects in both counterexplanation conditions had significantly lower non-error probabilities than the explanation-only subjects. That is, E/CE subjects assessed a significantly lower mean probability (53.25) than did the E-only subjects (77.20) (F=14.61, p=.0002) and CE/E subjects assessed a significantly lower mean probability (66.47) than did the E-only subjects (77.20) (F=2.84, p=.0496). These results suggest that the beliefs of auditors who have only written an explanation for the non-error cause are more likely to perseverc in the presence of potentially disconfirmatory information than are the beliefs of auditors who also have written a counterexplanation. Further analysis of this result is presented in the next subsection.

Table 2 also presents the descriptive and inferential statistics for the non-error probabilities assessed by the auditors in the two counterexplanation conditions – the two conditions used to test for the primacy order effect predicted by hypothesis 3. A comparison of the sample means indicates that subjects assigned to the E/CE condition assessed lower probabilities for the non-error cause (53.25) than did subjects assigned to the CE/E condition (66.47). These sample means are consistent with a recency order effect, rather
than the hypothesized primacy order effect. Thus, the null version of hypothesis 3 cannot be rejected in favor of the alternative which predicted that a primacy order effect should occur (p=.9796). Further analysis of this result also is presented in the next section.

Other Analyses. Other data collected as part of the experimental procedure are useful in explaining the second experiment's results. Recall that subjects in all three experimental conditions assessed the probability of occurrence for seven financial-statement errors. Table 3 presents the auditors' assessments of financial-statement error -- averaged across the seven error types. As expected, the counterexplanation-condition subjects (separately or combined) considered the occurrence of a financial-statement error to be more probable than did the explanation-only subjects (all p-values < .05). That is, the likelihood of financial-statement error was greater when explicit counterexplanation occurred.

---

To rule out the possibility that the explanation-only subjects' non-error (average error) probability assessments were systematically higher (lower) than those rendered by the counterexplanation subjects because of more-elaborate explanations provided by the former subjects, a content analysis was performed to determine the mean number of explanation arguments provided by each subject. The content analysis revealed no significant difference in the total number of explanation arguments generated between the explanation and the two counterexplanation conditions combined (F=1.34, p=.2530).

When comparing the average error probability between the two counterexplanation conditions, no significant difference in probability emerges (F=0.88, p=.3536). The CE/E
subjects were expected to assess a larger error probability than the E/CE subjects since the theoretical development for the primacy order effect suggested that the multiplicity of causation would be greater when counterexplanation occurs initially (CE/E) rather than after explanation (E/CE). Even though there was no difference in the financial-statement error probabilities between the two counterexplanation conditions, post-experimental data reveal that the E/CE subjects perceived the counterexplanation evidence to be stronger than did the CE/E subjects (F=4.17, p=.0472). This result, in conjunction with the finding that the E/CE subjects judged the explanation evidence to be weaker than did the CE/E subjects (F=5.53, p=.0234), is consistent with the experimental results for the non-error probability data -- that is, consistent with a recency order effect. The evidence considered last (first) in the two-task sequence was deemed to be stronger (weaker) than if it were considered first (last).

A possible explanatory variable for the unexpected recency results pertains to the subjects’ perceptions of task difficulty. Post-experimental data indicate that the CE/E subjects found it significantly more difficult to counterexplain than did the E/CE subjects (F=6.13, p=.0173).\textsuperscript{12} Since prior research indicates that there is an inverse relation between task difficulty and importance weights assigned to evidence in the judgment-formulation process (Sherman, et al. 1985), these results suggest that counterexplanation evidence, when generated initially, has a smaller impact on judgments about the non-error cause than if generated after explanation. Indeed, these results may indicate that auditors must comprehend (i.e., explain) an hypothesized cause before they can disconfirm it via counterexplanation (Gilbert 1989).
Finally, to rule out the possibility that task-difficulty affected the number of causal arguments comprising the explanations and counterexplanations, several additional content-analysis comparisons were made. The content analysis revealed no significant difference in the total number of counterexplanation arguments generated by the two counterexplanation conditions (F=1.83, p=.1833). The content analysis did, however, indicate a significant difference in the total number of explanation arguments. Specifically, the CE/E subjects assessed a smaller number of explanation arguments than did E/CE subjects (F=6.47, p=.0146). A more-detailed analysis of the written explanations reveals that many of the CE/E subjects documented no more than a change-in-sales-mix computation for their written explanation. Most E/CE subjects, in contrast, elaborated to a greater extent by also documenting other supporting arguments from the case materials. This content analysis result is consistent with research in which primacy, rather than recency, order effects are observed (Hoch 1984, Moscr 1989). Since the non-error sample means from the two counterexplanation conditions are in the direction of a recency order effect, it must be concluded that differential weighting, rather than differential argument production, produced the recency results.

IV. CONCLUDING COMMENTS

This study provides mixed support for belief-perseverance theory and its predictions regarding sequential belief revision in the context of audit analytical review. Experiment 1 demonstrates that auditors asked to develop and document explanations supporting the validity of a non-error cause consider its likelihood of occurrence to be significantly greater than do auditors not asked to develop and document such explanations. This explanation-
effect finding is consistent with belief-perseverance theory. It is important from a theoretical perspective, since it is the first empirical demonstration of the explanation effect in a diagnostic-inference domain as well as the first to demonstrate the explanation phenomenon using subjects who have significant prior experience with the experimental task.

As shown by the second experiment, the explanation-effect result may have significant potential audit effectiveness implications. That is, once an auditor has developed a written explanation for the non-error cause, his/her belief in the validity of that cause appears to persevere unless he/she is explicitly requested to counterexplain. Since current practice does not require auditors to engage in counterexplanation when performing analytical review, these results suggest that auditors who do not explicitly consider why an hypothesized cause may not be true will be more prone to accepting a plausible, but potentially incorrect, cause.

The counterexplanation effect demonstrated in this study is especially important in light of the fact that disconfirmatory evidence was provided to the subjects in the client-discussion narratives, thus alleviating the necessity to recall such information from long-term memory. Although subjects had to recognize the readily available disconfirmatory evidence as such, it appears that they had little trouble with this recognition task given that the average number of opposing arguments generated by the counterexplanation-condition subjects was 16.74. That is, this result may imply that even if counterexplanation evidence is readily available in the audit workpapers or other documentation, the auditor may not incorporate its impact into his/her analytical review judgments unless explicitly directed to do so.
Other results from experiment 2, however, do not support belief-perseverance theory. Specifically, a primacy order effect is not indicated when explanation and counterexplanation activities are sequentially performed in the two possible task-performance orders. Rather, the experimental data are in the direction of a recency order effect. Since post-experimental data indicate that CE/E auditors had greater difficulty generating opposing arguments than did E/CE auditors, it may be that auditors must comprehend (via explanation) a non-error cause before they can disconfirm it (Gilbert 1989). That is, the "natural order" of comprehension may be to explain a belief and then (possibly) to reject it.

There are several limitations of this study, many of which are common to most experimental auditing studies. First, the auditors in this study may have guessed the experimental hypotheses when asked to develop written explanations and/or counterexplanations during completion of the case. Such experimental demand effects would have increased the likelihood of finding significant results. However, the use of a between-subjects design should have reduced the potential for demand effects (Pany and Reckers 1987). Another potential limitation pertains to the auditors' level of motivation. Although informal feedback from the study participants indicated that they found the study quite realistic and interesting, it still could be that the participants were not sufficiently motivated to spontaneously generate explanations (experiment 1) and counterexplanations (experiment 2) when completing the experimental tasks. Again, this possibility would have increased the likelihood of finding significant results.

The present study's results may have important practical implications for auditing firms. Of course, additional research is necessary before any changes in practice are in order. To the extent that auditing firms are more concerned with incorrect acceptance --
rather than with incorrect rejection -- of a non-error explanation when the "correct" explanation involves a financial-statement error, several ways are suggested for structuring analytical review. First, auditors could be required to document reasons why a client's suggested non-error cause may not be valid. This study demonstrated that once auditors document an explanation for a client-provided non-error cause, they are less likely to consider counterexplanation evidence than are auditors who are explicitly requested to counterexplain. Thus, it may be that simply requiring auditors to write down reasons that do not support an explained non-error cause may be an effective way to avoid compromising audit effectiveness. Second, since other results from experiment 2 suggest that counterexplanation has the greater impact when considered after a written explanation for the non-error cause has been developed, counterexplanation activities could be advised after a causal explanation has been developed (or possibly after a thorough comprehension of the non-error cause).
ENDNOTES

1. As used herein, error causes refer to intentional and unintentional financial-statement misstatements, while non-error causes refer to unusual transactions or events, accounting changes, or random fluctuations.

2. If planned tests of details are decreased based on the results of the attention-directing analytical review, the analytical review not only serves to assist the auditor in planning the nature, timing and extent of other auditing procedures but serves as substantive-testing evidence. If, in contrast, planned tests of details are increased, the analytical review serves only to assist the auditor in planning the nature, timing and extent of other auditing procedures.

3. Prior research has shown that merely thinking about a target event can lead to an explanation effect (Carroll 1978). Thus, it is not documenting the explanation per se that leads to the explanation effect.

4. Predictive inference involves reasoning from observed cues or causes to potential effects. This is in contrast to diagnostic inference which entails reasoning from observed effects to potential causes (Einhorn and Hogarth 1986).

5. Although auditors are not required by current auditing standards to consider counterexplanation evidence when performing analytical review, investigating the effect of counterexplanation and the order of explanation and counterexplanation on auditors' judgments is nevertheless important from a scholarly viewpoint. Specifically, if it is shown
that auditors do not incorporate (or that they differentially incorporate -- depending on task-performance order) the impact of readily available counterexplanation evidence into their judgments, our understanding of the potential for acceptance of a plausible, but incorrect, cause would be enhanced.

6. To avoid a potential experimental confound, subjects (in both experiments) were not allowed to refer to their prior responses or to the case materials when rendering their probability assessments. However, subjects were free to refer to Parts A and B of the case materials when formulating their explanations and (for experiment 2) written counterexplanations.

7. The criteria used for judging a successful target-hypothesis explanation were twofold. First, if the subject identified products, quantities, or selling prices in his/her explanation, then correct identification of such factors was required. Second, if the subject attempted to quantify the magnitude of the change in sales mix, then his/her calculations were required to free of error -- both in terms of logic and mathematical accuracy.

8. For both experiments 1 and 2, none of the hypothesis-test results change if the "non-explainers" are left in the analysis.

9. Materiality was quantified at $367,250 which was the difference between the gross margin increase ($3,667,250) and the dollar threshold used for explaining the change in sales mix ($3,300,000).

10. Although the mean probabilities for each error type could be analyzed separately, this approach is problematic since there is significant variability in the number of subjects
explaining each error type. Thus, to provide a less-variable measure, the average error measure (across the seven financial-statement errors) was used.

11. The content analysis was performed independently by the researcher and by a second paid coder. Intraclass correlations for agreement on counts of total arguments were separately computed for the explanation and, where applicable, counterexplanation responses. All statistics indicated statistically significant agreement between the two coders (p < .0001). Any differences in coding were jointly resolved.

12. Although the E/CE subjects found it more difficult to generate their written explanations than the CE/E subjects, this difference was not statistically significant (F=0.80, p=.3750).
REFERENCES


Pei, K., Moeckel, C., and H. Reneau. 1990. The relations between auditors’ memory and judgment. Working paper. Arizona State University, Tempe, AZ.


**TABLE 1**

RESULTS OF THE BETWEEN- AND WITHIN-SUBJECTS TESTS
OF THE EXPLANATION EFFECT -- EXPERIMENT 1

<table>
<thead>
<tr>
<th>BETWEEN-SUBJECTS TEST</th>
<th>Mean Probability Assessment For Non-Error Cause (Std. Dev.)</th>
<th>t-Statistic</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control-Condition Subjects (Pre-explanation assessment)</td>
<td>64.62 (12.21)</td>
<td>2.44</td>
<td>.0112*</td>
</tr>
<tr>
<td>(n=13)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment-Condition Subjects (Post-explanation assessment)</td>
<td>76.62 (12.86)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n=13)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>WITHIN-SUBJECTS TEST</th>
<th>Mean Probability Assessment For Non-Error Cause (Std. Dev.)</th>
<th>t-Statistic</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control-Condition Subjects (Pre-explanation assessment)</td>
<td>64.62 (12.21)</td>
<td>3.26</td>
<td>.0035*</td>
</tr>
<tr>
<td>(n=13)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control-Condition Subjects (Post-explanation assessment)</td>
<td>75.62 (14.90)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n=13)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*One-tailed test.
### TABLE 2

DESCRIPTIVE STATISTICS AND PLANNED-COMPARISON TESTS FOR THE NON-ERROR PROBABILITIES -- EXPERIMENT 2

#### PANEL A: DESCRIPTIVE STATISTICS

<table>
<thead>
<tr>
<th>Mean Probability Judgment</th>
<th>(A) Explanation-Only</th>
<th>(B) Explanation-then-Counter-explanation</th>
<th>(C) Counterexplanation-then-Explanation</th>
<th>(D) Columns B &amp; C Averaged</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>77.20</td>
<td>53.25</td>
<td>66.47</td>
<td>59.65</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>10.93</td>
<td>19.01</td>
<td>20.67</td>
<td>20.62</td>
</tr>
<tr>
<td>Number of Subjects</td>
<td>n=15</td>
<td>n=16</td>
<td>n=15</td>
<td>n=31</td>
</tr>
</tbody>
</table>

#### PANEL B: PLANNED-COMPARISON TESTS

<table>
<thead>
<tr>
<th>PLANNED COMPARISON</th>
<th>MEAN SQUARE</th>
<th>F-STATISTIC</th>
<th>PROB.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. EXPLANATION-ONLY VERSUS BOTH COUNTEREXPLANATION CONDITIONS</td>
<td>3038.99</td>
<td>10.00</td>
<td>.0015*</td>
</tr>
<tr>
<td>1A. EXPLANATION-ONLY VERSUS EXPLANATION-THEN-COUNTEREXPLANATION</td>
<td>4440.99</td>
<td>14.61</td>
<td>.0002*</td>
</tr>
<tr>
<td>1B. EXPLANATION-ONLY VERSUS COUNTEREXPLANATION-THEN-EXPLANATION</td>
<td>864.03</td>
<td>2.84</td>
<td>.0496*</td>
</tr>
<tr>
<td>2. EXPLANATION-THEN-COUNTEREXPLANATION VERSUS COUNTEREXPLANATION-THEN-EXPLANATION</td>
<td>1352.36</td>
<td>4.45</td>
<td>.9796*</td>
</tr>
</tbody>
</table>

*One-tailed test.
### TABLE 3

DESCRIPTIVE STATISTICS AND PLANNED-COMPARISON TESTS FOR THE AVERAGE FINANCIAL-STATEMENT ERROR PROBABILITIES -- EXPERIMENT 2

#### PANEL A: DESCRIPTIVE STATISTICS

<table>
<thead>
<tr>
<th></th>
<th>(A) Explanation-Only</th>
<th>(B) Explanation-then-Counterexplanation</th>
<th>(C) Counterexplanation-then-Explanation</th>
<th>(D) Columns B &amp; C Averaged</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Probability</td>
<td>35.28</td>
<td>49.12</td>
<td>45.47</td>
<td>47.35</td>
</tr>
<tr>
<td>Judgment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>12.80</td>
<td>10.39</td>
<td>8.98</td>
<td>9.75</td>
</tr>
<tr>
<td>Number of Subjects</td>
<td>n=15</td>
<td>n=16</td>
<td>n=15</td>
<td>n=31</td>
</tr>
</tbody>
</table>

#### PANEL B: PLANNED-COMPARISON TESTS

<table>
<thead>
<tr>
<th>PLANNED COMPARISON</th>
<th>MEAN SQUARE</th>
<th>F-STATISTIC</th>
<th>PROB.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. EXPLANATION-ONLY VERSUS BOTH COUNTEREXPLANATION CONDITIONS</td>
<td>1458.84</td>
<td>12.45</td>
<td>.0010*</td>
</tr>
<tr>
<td>1A. EXPLANATION-ONLY VERSUS EXPLANATION-THEN-COUNTEREXPLANATION</td>
<td>1482.91</td>
<td>12.65</td>
<td>.0010*</td>
</tr>
<tr>
<td>1B. EXPLANATION-ONLY VERSUS COUNTEREXPLANATION-THEN-EXPLANATION</td>
<td>778.84</td>
<td>6.64</td>
<td>.0136*</td>
</tr>
<tr>
<td>2. EXPLANATION-THEN-COUNTEREXPLANATION VERSUS COUNTEREXPLANATION-THEN-EXPLANATION</td>
<td>103.11</td>
<td>0.88</td>
<td>.3536*</td>
</tr>
</tbody>
</table>

*One-tailed test.
APPENDIX

EXAMPLES OF A REPRESENTATIVE EXPLANATION
AND COUNTEREXPLANATION

EXPERIMENT 1:

Explanation

The reason for the fluctuation is due to the change in sales mix. Approximately 2,000 color monitors with a gross margin of approximately 35% replaced the monochrome monitors at about 29%. In addition, laser writers which have a gross margin of approximately 34% have just hit the market in the current year. Approximately 6,000 were sold in the current year at a gross margin of 34%. These printers are replacing the letter-quality and dot-matrix printers which earn approximately 33% and 20% gross margins, respectively.

EXPERIMENT 2:

Explanation

| Laser printers | 6,000 * $1,660 * .34 | $3,386,400 |
| Dot matrix     | (6,100) * 230 * .20  | (280,600)  |
| Color monitors | 2,200 * 425 * .35    | 327,250    |
| Monochrome monitors | (2,300) * 149 * .29 | (99,383)   |

$3,333,667

Note that the design work on the laser printer was performed in 87 but not in 88.

Counterexplanation

Sales cutoff on the 2 computer manufacturer order of letter-quality printers. Sales cutoff problems could have occurred by recording the sale in 88 and also counting the printers as inventory.

1,500 * 542 = $813,000
2,000 * 425 = $850,000

Improper cutoff of purchases. Improper cutoff of purchases could have occurred by counting the December shipment but not recording the related purchase of $600,000.

Other potential errors:

Improper cutoff of sales returns and related accounting for returns of sales.
Improper costing of inventory.
Improper cutoff and related accounting for returns of purchases.
Improper LIFO calculation.
Improper physical count.

There is no information on the change in mix between contract and non-contract sales. Furthermore, the gross margins are not available for 1988 or by contract and non-contract. Sales also are not available by contract and non-contract. R&D expenditure would appear to be up in 1988.
FIGURE 1
PROCEDURE FOR EXPERIMENT ONE

TREATMENT CONDITION

READ BACKGROUND CASE INFORMATION AND CLIENT DISCUSSION NARRATIVES (Parts A & B)

PERFORM FAMILIARITY TASK (Part C)

FOCUS ON UNUSUAL INCREASE IN MARGIN

LEARN CLIENT'S REASON (NON-ERROR CAUSE) FOR MARGIN INCREASE

WRITE EXPLANATION FOR NON-ERROR CAUSE

ASSESS PROBABILITY FOR NON-ERROR CAUSE

COMPLETE POST-EXPERIMENTAL QUESTIONNAIRE

CONTROL CONDITION

ASSESS PROBABILITY FOR NON-ERROR CAUSE

COMPLETE POST-EXPERIMENTAL QUESTIONNAIRE

* To equate the time spent on the task between the control and treatment groups and to provide a supplemental, within-subjects test of the explanation effect, the control-condition subjects also were asked to write an explanation for the non-error cause and to reassess the non-error probability. The development of the written explanation and the reassessment of the non-error probability were performed after their initial probability assessment but before completion of the post-experimental questionnaire.
FIGURE 2

PROCEDURE FOR EXPERIMENT TWO

LEGEND:
A -- Explanation-Only Condition
B -- Explanation-then-Counterexplanation Condition
C -- Counterexplanation-then-Explanation Condition