INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book. These are also available as one exposure on a standard 35mm slide or as a 17" x 23" black and white photographic print for an additional charge.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600
Assessing a company's exposure to fraudulent financial reporting: Implications of seemingly irrelevant evidence

Hackenbrack, Karl Edward, Ph.D.

The Ohio State University, 1988
Copyright by
Karl Edward Hackenbrack
1988
To Ma-maw
VITA

April 6, 1957 ....................... Born - Harrisburg, Pennsylvania

1979 .............................. B.S., Davis & Elkins College, West Virginia

1979-1982 .......................... Audit Senior, Laventhol & Horwath, Harrisburg, Pennsylvania

1981 .............................. Certified Public Accountant, Pennsylvania

1983 .............................. M.B.A., Shippensburg University, Pennsylvania

1983-1984 .......................... Assistant Professor, Shippensburg University, Pennsylvania

PUBLICATIONS


FIELDS OF STUDY

Major Field: Accounting

Minor Fields: Microeconomic Theory, Mathematical Statistics
TABLE OF CONTENTS

DEDICATION ......................................................... ii

ACKNOWLEDGEMENTS ................................................... iii

VITA ................................................................ iv

LIST OF TABLES ...................................................... viii

LIST OF FIGURES ...................................................... ix

CHAPTER PAGE

I. INTRODUCTION .................................................. 1

Nature of the Problem ............................................. 1
Contributions of the Study ........................................ 2
Organization of the Paper ......................................... 6

II. LITERATURE REVIEW ............................................ 8

Fraudulent Financial Reporting .................................... 8
Implications of Irrelevant Evidence for Decision-making .... 18
Applied Research on the Influence of Irrelevant Evidence ... 18
Information Load ................................................... 24
Conceptual Structures and the Processes that Operate on Them ................................................... 25
Influence of Prototypes on Perception and Processing of Evidence ............................................. 29
Confirmation Bias ................................................... 30
The Role of Confirming/Disconfirming Evidence in Rule-discovery Tasks ........................................... 33
Chapter Summary ................................................... 33

III. THEORY DEVELOPMENT ...................................... 35

Structure of the Audit Domain ................................... 35
Formal Model of the Evidence Evaluation Process .......... 39
Different Types of Nondiagnostic Evidence .................. 52
Chapter Summary ................................................... 54
## IV. DEVELOPMENT OF THE EXPERIMENTAL MATERIALS

<table>
<thead>
<tr>
<th>Overview</th>
<th>55</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
<td>55</td>
</tr>
<tr>
<td>Experimental Materials</td>
<td>56</td>
</tr>
<tr>
<td>Create the List of Potential Items Instrument</td>
<td>58</td>
</tr>
<tr>
<td>Distribute the List of Potential Items Instrument to the Consultants</td>
<td>60</td>
</tr>
<tr>
<td>Classify Items in the List of Potential Items Instrument</td>
<td>60</td>
</tr>
<tr>
<td>Edit Items Selected From the List of Potential Items Instrument</td>
<td>62</td>
</tr>
<tr>
<td>Informative Items</td>
<td>62</td>
</tr>
<tr>
<td>Nondiagnostic Items</td>
<td>63</td>
</tr>
<tr>
<td>History and Background Narrative</td>
<td>64</td>
</tr>
<tr>
<td>Preliminary Version of the Experimental Materials</td>
<td>64</td>
</tr>
<tr>
<td>Create Sets of Nondiagnostic Workpaper Excerpts</td>
<td>65</td>
</tr>
<tr>
<td>Preliminary Versions</td>
<td>65</td>
</tr>
<tr>
<td>Practitioner Review of Nondiagnostic Workpaper Excerpts</td>
<td>66</td>
</tr>
<tr>
<td>Validate Nondiagnostic Workpaper Excerpts</td>
<td>67</td>
</tr>
<tr>
<td>Select Informative Workpaper Excerpt Themes and Labels for the Rating Scale</td>
<td>69</td>
</tr>
<tr>
<td>Preliminary Selection of Informative Themes and Labels</td>
<td>70</td>
</tr>
<tr>
<td>Final Selection of Informative Themes and Labels</td>
<td>72</td>
</tr>
<tr>
<td>Chapter Summary</td>
<td>73</td>
</tr>
</tbody>
</table>

## V. DATA GATHERING AND ANALYSIS

<table>
<thead>
<tr>
<th>Method</th>
<th>74</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants and Setting</td>
<td>74</td>
</tr>
<tr>
<td>Materials</td>
<td>75</td>
</tr>
<tr>
<td>Procedure</td>
<td>76</td>
</tr>
<tr>
<td>Dependent Variable</td>
<td>79</td>
</tr>
<tr>
<td>Effects of Nondiagnostic Evidence on Diagnostic Evidence</td>
<td>80</td>
</tr>
<tr>
<td>Observations</td>
<td>80</td>
</tr>
<tr>
<td>Dilution Effect</td>
<td>83</td>
</tr>
<tr>
<td>Favorable/Neutral/Unfavorable Distinction</td>
<td>86</td>
</tr>
<tr>
<td>Conclusion for Diagnostic Evidence</td>
<td>90</td>
</tr>
<tr>
<td>Effects of Nondiagnostic Evidence on Counterdiagnostic Evidence</td>
<td>93</td>
</tr>
<tr>
<td>Observations</td>
<td>93</td>
</tr>
<tr>
<td>Dilution Effect</td>
<td>97</td>
</tr>
<tr>
<td>Conclusion for Counterdiagnostic Evidence</td>
<td>98</td>
</tr>
<tr>
<td>Relative Importance of Individual Informative Items</td>
<td>101</td>
</tr>
</tbody>
</table>

## VI. GENERAL DISCUSSION

<table>
<thead>
<tr>
<th>Conclusion</th>
<th>104</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternative Interpretations</td>
<td>106</td>
</tr>
<tr>
<td>Recommendations</td>
<td>107</td>
</tr>
</tbody>
</table>
LIST OF REFERENCES .................................................. 110

APPENDIX .............................................................. 119

General Instructions ............................................... 120
History and Background of Mr. Chip Company ................... 121
Excerpts From the 1986 Mr. Chip Company Audit Workpapers ... 125
Questions About Mr. Chip Company ................................ 134
General Questions .................................................. 144
Set of Cards Labeled 'Question One' ............................... 153
Set of Cards Labeled 'Question Two' ............................... 156
LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Reconciliation of Number of Participants with Number of Observations -- Diagnostic Version</td>
<td>82</td>
</tr>
</tbody>
</table>
| 2. Location Tests
  Diagnostic Version | 92 |
| 3. Reconciliation of Number of Participants with Number of Observations -- Counterdiagnostic Version | 96 |
| 4. Location Tests
  Counterdiagnostic Version | 100 |
<p>| 5. Concordance Tests | 103 |</p>
<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Relationship Among Sets of Features</td>
<td>38</td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

1.1 Nature of the Problem

All audit evidence is relevant to one or more of the many decisions an auditor makes in forming an opinion as to whether a client's financial statements are presented in conformity with applicable standards, but not all audit evidence is relevant to every such decision. Indeed, much of the evidence an auditor uses pertains to only one or a few of the individual decisions that influence his/her opinion.

An auditor's opinion on a set of financial statements is an expert judgment. An essential ingredient of expert judgment is the ability to make suitable use of items of evidence that vary in their degree of relevance. Nevertheless, a wealth of findings across a variety of settings demonstrate that seemingly irrelevant evidence can influence the judgments of experienced decision-makers (see Gaeth and Shanteau [1981]). The current study explores this issue in the context of an important, timely, but rarely-studied type of expert judgment -- an auditor's assessment of the risk that the financial statements a client presents for audit have been falsified (have been affected by fraudulent financial reporting). This judgment is called the "fraud-risk assessment" in this paper.
An auditor uses his/her fraud-risk assessment "... in determining the nature, timing, ... [and] extent of procedures, assigning staff, ... [and] requiring appropriate levels of supervision [AICPA, 1988a, p. 8]." In this way, decisions about the risk of financial-statement fraud influence the conduct of an auditor's examination of a client's financial statements and thus may lead to the detection of (or to a failure to detect) intentional misstatements in them.

1.2 Contributions of the Study

Persons in both the private and public sectors believe that auditors should improve their ability to detect fraud in financial reporting. Public accountants (auditors) themselves recognize that they must develop more efficient and effective audit procedures, training methods, and decision aids to help auditors detect such fraud. The ultimate goal of the stream of research initiated by this study is to find ways to increase the likelihood that auditors will identify misstated financial statements before they issue an opinion on them.

The objectives of this study are:

1. To develop a theoretical framework that explains how auditors combine different types of evidence when they assess the risk that a client's financial statements are fraudulent.

2. To investigate the role, in the fraud-risk assessment, of seemingly irrelevant evidence, that is, evidence auditors typically consider useless when they make that assessment.

3. To extend the literature on how apparently irrelevant factors affect human judgment.

The first objective is accomplished by creating an algebraic model of the evidence-evaluation process. This model draws on previous work on probabilistic information processing [Wallsten, 1972, 1976; Fischhoff
and Beyth-Marom, 1983], on conceptual structure [Smith and Medin, 1981],
on judgments of similarity [Tversky, 1977], on averaging rules in
Bayesian inference [Lopes, 1985], and on classification performance
[Forsyth, 1988]. It is consistent with the prototype view of
categorization behavior.

Audit evidence is any information used by an auditor to determine
whether the set of financial statements being audited is presented in
conformity with established criteria. Three classes of audit evidence
are considered. The classes, and the names assigned to them for the
purposes of this study, are:

1. Diagnostic evidence -- Audit evidence that describes incentives
or opportunities to commit fraud in financial reporting.

2. Counterdiagnostic evidence -- Audit evidence that describes
aspects of a company's financial reporting process that are
intended to prevent or detect such fraud.

3. Nondiagnostic evidence -- Audit evidence that does not describe
conditions that indicate whether a company's exposure to fraud
in financial reporting is high or a company's exposure to such
fraud is low. This evidence describes aspects of a company
that are useful for one or more audit decisions but not useful
for, and hence not relevant to, the fraud-risk assessment.

Shepard [1987] suggested that until the advent of the
categorization paradigm there did not seem to exist a law in psychology
comparable in generality to the laws that exist in the physical
sciences. The proliferation in the psychological literature of articles
that adopt the perspective provided by the categorization paradigm
testifies to the usefulness of the paradigm. The current study
illustrates how this perspective can be used to develop testable
hypotheses in audit-judgment research.
The algebraic model is a feature-matching model. Such a model is especially useful because it explicitly portrays the interaction between the audit evidence and the knowledge brought by an auditor to an audit task. Frederick and Libby [1986, p. 289] noted that a feature-matching process may allow auditors to recognize problems and to focus on likely solutions quickly.

The second objective of the study -- to investigate the role, in the fraud-risk assessment, of seemingly irrelevant evidence -- is accomplished in the empirical part of the study. The model of the evidence-evaluation process is used to structure both the data-gathering and the data-analysis aspects of the study. The focus is on the role nondiagnostic evidence plays in the evaluation process. Practicing auditors were consulted in developing and validating the experimental materials. They considered, among other matters, the classification of items of evidence. The experimental exercise was administered to a different group of practicing auditors, referred to in this paper as "participants", in a laboratory setting.

The relationships among the data are consistent with the implications of the algebraic model -- nondiagnostic evidence tempered the influence of evidence they considered useful for the fraud-risk assessment, and relatively weak useful evidence had no effect when nondiagnostic evidence was presented. This moderating effect was observed in connection with the participants' evaluations of both diagnostic and counterdiagnostic evidence (cf. Nisbett, Zukier, and Lemley [1981]). Not only did the participants use nondiagnostic evidence far more than they realized, but their reports of how the
nondiagnostic evidence affected their assessments contradicted its observed effect.

The author believes that these results occurred because fraud-risk assessments are mediated by similarity judgments. That is, the participants compared their mental representations of the audit evidence, including the nondiagnostic evidence, with their conceptions of the fraud-related conditions at a company that has a high risk of financial-statement fraud and with their conceptions of the fraud-related conditions at a company that has a low risk of such fraud. The presence of nondiagnostic evidence reduced the similarity between conditions at the company described in the study and each such conception, and judgments less extreme than those made in the absence of the nondiagnostic evidence resulted.

The third objective -- to extend the literature on apparently irrelevant factors -- is accomplished in several ways. First, the study demonstrates that a phenomenon observed in social psychology, the moderating effect of nondiagnostic evidence, can also be observed in an auditing context. Second, moving beyond the auditing issue, an algebraic model is formulated that explains the mechanisms that may have produced the observed results in this and previous studies. This is a general model. In an auditing context it can serve as a framework in studying decision-making behavior on issues other than the fraud-risk assessment. More important, it can serve as a framework for studying decision-making behavior in domains unrelated to auditing (or to accounting). Third, the study demonstrates that a descriptive model that explains why seemingly irrelevant evidence affects judgments, can
preserve the two cornerstones of a Bayesian framework -- the likelihood principle and the multiplying principle. Troutman and Shanteau [1977, p. 53-4] stated that it is unlikely that such a model can be developed. Even Wallsten's [1972; 1976, p. 146] very flexible formulation of a Bayesian model predicts that nondiagnostic evidence will not affect judgments.

Fourth, researchers in other disciplines have demonstrated a phenomenon that is conceptually similar to the dilution effect observed in this study. This study extends that work by using experienced, motivated decision-makers and a task they consider realistic. It also considers the relative influence of different types of nondiagnostic evidence, in contrast to the routine, unexceptional evidence manipulated in prior studies. Support for this extension was found; more interesting nondiagnostic evidence had a greater moderating effect than less interesting nondiagnostic evidence.

1.3 Organization of the Paper

The rest of the paper is organized as follows. Chapter II is a literature review. In Chapter III, an algebraic model of the evidence-evaluation process is developed. This model was used to guide the data-gathering and data-analysis aspects of the study. In Chapter IV, the process of developing the audit evidence portion of the experimental materials is described. In Chapter V, the data-gathering and data-analysis aspects of the study are discussed. Details of the experimental method, the results, and the implications of the results for the practice of auditing are presented. Chapter VI presents a
general discussion of the results of the study, alternative explanations for the results, and recommendations.
CHAPTER II
LITERATURE REVIEW

The literature review is divided into four sections. Studies that consider the fraudulent financial reporting issue are summarized in the first section. The studies described in the second section demonstrate that seemingly irrelevant evidence can influence judgments. The third section is a review of the theories that underlie models of conceptual structures and the psychological processes that operate on such structures. This review complements the description in Chapter III of the theoretical framework used in this study. The studies described in the fourth section demonstrate that when prototypes (stereotypes) are brought to mind, individuals form expectations that influence how they evaluate evidence.

2.1 Fraudulent Financial Reporting

Fraudulent financial reporting, sometimes called management fraud or financial fraud, is "... intentional or reckless conduct, whether by act or omission, that results in materially misleading financial statements [National Commission on Fraudulent Financial Reporting, 1987, p. 2]." It may entail falsifying corporate records, reporting transactions dishonestly, or misapplying accounting principles. Employees at any level in an organization may be involved.
Fraudulent financial reporting differs from other causes of misstatement in financial statements in that there must be a deliberate attempt to deceive third parties [Elliott and Willingham, 1980, p.4]. Unintentional misstatements or omissions in financial statements (errors), misappropriations of assets (employee fraud or defalcations), and corporate activities that are fraudulent under law but do not lead to intentional misstatement of financial statements are not instances of fraudulent financial reporting. Some of these activities may lead to fraudulent financial reporting but are not themselves examples of it.

Hylas and Ashton [1982] studied 152 audits by a Big-8 firm of financial statements with fiscal years ending in 1978. The firm detected 281 errors leading to adjustments of the statements. Before each audit began, the partner in charge received a letter that described the study and its information requirements. During the audits, the audit team members accumulated the requested information for each error discovered that required an adjustment of the financial statements. Ten of the 281 errors were believed by the auditors to have been intentional. These errors tended to be small and to occur in small companies. Six were initially signaled by detailed tests, two by analytical reviews, one by prior expectations, and one by a discussion with the client. The National Commission on Fraudulent Financial Reporting [1987, p. 52] suggests that auditors perform analytical review procedures in all audits.

St. Pierre and Anderson [1984] examined 129 lawsuits filed against public accountants during the 1960's and 70's. Forty-two of the cases involved deliberate misrepresentations in the financial statements;
seventeen of these involved local or regional firms. Forty-three percent of the fraud cases involved auditors with three or fewer years' experience with the client, whereas only twenty-four percent of all the cases analyzed involved auditors with three or fewer years' experience with the client. The accounting profession's concern over excessive responsibility detecting fraud is not supported by the results of this study. The study showed that the public accountants had been held responsible only if the fraud would have been discovered by an ordinary audit, properly performed. Thus, the auditors had not been required to modify normal auditing procedures.

Additional descriptive evidence was gathered by Palmrose [1987] (s.a. Palmrose [1988]). She studied 472 lawsuits against the fifteen largest auditing firms filed from 1960 to 1985, exploring the role of business failure and management fraud both in litigation brought against public accountants and in settlements of such cases. Forty-four percent of the litigation involved allegations of management fraud. Fifty-six percent of the lawsuits involving bankrupt companies involved allegations of management fraud. The management-fraud cases were the ones most often resolved through payment of damages by auditors, and they were the primary type of case with large auditor payments.

Gibbins and Wolf [1982] asked eighty partners and managers from six major public accounting firms in Canada to complete an extensive questionnaire. It was designed to gain an understanding of the decision environment in which an audit is conducted. Respondents were asked to rate the operational influence of many potential components of the environment at three stages of an audit engagement. One component was
fraud considerations. They did not reach even average importance at any stage of the audit.

Currently, elements of both the public and the private sectors have expressed a belief that the auditing profession should take a more active role in detecting and deterring fraudulent reporting. In its recent report, the National Commission on Fraudulent Financial Reporting recommended that independent public accountants "... be responsible for actively considering the potential for fraudulent financial reporting in a given engagement [1987, p. 51]." The Commission identified a number of steps to improve auditors' ability to detect fraudulent financial reporting. The Commission asked the auditing profession to: restate auditors' responsibility for detecting fraudulent reporting; require auditors to evaluate a client's control environment; strengthen the profession's quality assurance programs; clarify the profession's role and responsibilities vis-a-vis users of public accountants' reports; reorganize the auditing-standards-setting process.

In April 1988, the Auditing Standards Board of the American Institute of Certified Public Accountants (AICPA) issued several Statements on Auditing Standards that address many of the issues considered by the Commission. Statement No. 53 requires an auditor to "... assess the risk that errors and irregularities may cause the financial statements to contain a material misstatement. Based on this assessment, the auditor should design the audit to provide reasonable assurance of detecting errors and irregularities that are material to the financial statements [AICPA, 1988a, p. 2]." Irregularities include both fraudulent financial reporting and defalcations.
This risk assessment is initially made during the planning stage of an audit. If conditions or circumstances that cause an auditor to revise his/her initial risk assessment arise during the audit, he/she reconsiders the planned scope of the audit procedures. Statement No. 53 lists factors an auditor should consider in making the risk assessment.

Statement No. 55 describes the elements of an internal control structure and explains how auditors should use their understanding of a client's internal control structure in planning and performing an audit. This statement requires an auditor to obtain "... knowledge about whether [policies, procedures, and records] have been placed in operation ... [AICPA, 1988b, p. 8]."

Control risk is the risk that a material misstatement that could occur will not be prevented or detected by a client's internal control policies and procedures before financial statements are issued. The auditor's understanding of a client's internal control structure is used in assessing control risk. Statement No. 55 requires an auditor to document his/her understanding of the client's internal control structure and the basis for his/her conclusions about the assessed level of control risk.

Understanding incentives and opportunities for fraudulent financial reporting has been a matter of interest for many years. Most studies of how to detect fraudulent financial reporting have attempted to identify the individual and organizational factors, called red flags, that would help auditors focus their attention on high-risk areas. Some authors have suggested using these factors to develop client screening programs to alert auditors to the existence of a higher-than-normal risk of
financial fraud (e.g., Sorenson and Sorenson [1980] and AICPA [1979]). Once the attributes of the stereotypical fraud-prone company were identified, these systems would help auditors assess the risk of financial fraud in a given engagement by comparing the client to the stereotype [Libby, 1981].

Touche Ross [1974] designed a set of procedures to determine management's involvement in material (not necessarily fraudulent) transactions. This was the first step in the development of a systematic approach to auditors' review of management's actions [Sorenson and Sorenson, 1980, p. 197]. Coopers and Lybrand [1977] extended this approach to consider "danger signals of improper [management] practice."

Albrecht, Cherrington, Payne, Roe, and Romney [1980] (s.a. Albrecht and Romney [1980]) conducted an extensive interdisciplinary review of the fraud-related literature in a search for factors that signal fraud. They used a content-analysis of fifty fraud cases to "validate" the items identified in the literature search. The resulting list of eighty-five items (red flags) were grouped into 'personality', 'situational pressure', and 'opportunity to commit' indicators of fraud. They suggested that the decision to commit management fraud is determined by the synergistic influence of factors from each of these three groups.

To date, two approaches to implementing red flags have been considered: checklists (e.g., Touche Ross's 'Management Fraud Decision Aid') and statistical prediction models (e.g., Sorensen, Grove, and Selto [1983]). The checklist approach involves generating lists of
questions auditors ask themselves about the client and its people, and lists of questions they ask key client personnel. The statistical prediction models are similar to the bankruptcy-type prediction model, e.g., Altman [1968].

Sorensen, et al. [1983] sorted nearly sixty red flags into eight factors they called major audit problems. The authors first analyzed twenty fraud cases to determine if the eight major audit problems were present. A number of major audit problems could be found in each fraud case. The authors then used a discriminant analysis approach to develop a formal fraud detection model. Despite their emphasis on the eight major audit problems, they used sixteen financial statement ratios to capture the single factor 'economic and financial pressure'. Wallace [1983] and Willingham [1983] leveled several rather harsh criticisms of the Sorensen, et al. paper.

Jones and Maher [1987] introduced a project on which they presumably are continuing to work. The goal of their research is to develop a fraud-prediction model based on characteristics that can be measured using publicly available data. Related studies in the red-flags literature have been criticized for (1) identifying characteristics that can be found both in firms with and in those without management fraud and (2) failing to identify antecedent conditions (see Wallace [1983]). Jones and Maher are explicitly addressing these issues as they develop their model.

Attempts to establish an empirical link between red flags and the occurrence of management fraud have not been particularly successful. Reasons cited include: (1) the failure of the researchers to isolate a
small number of factors; (2) the use of a large number of financial ratios oriented toward predicting financial stress, when one motivation for financial fraud is to conceal financial difficulties; and (3) the need to evaluate the personal lives and moral character of senior managers. (See Wallace [1983], Jones and Maher [1987], and Johnson, Jamal and Berryman [1988] for further discussion.)

Joyce and Biddle [1981a] studied the anchor-and-adjustment heuristic in three related experiments. The topic of the first experiment was management fraud. In it, subjects were asked two questions. The extremity of the anchor was manipulated by first asking subjects whether the incidence of executive-level management fraud was more than a specified percent of firms audited by Big Eight accounting firms. Then, subjects were asked to estimate the number of audit clients out of a thousand that have executive-level management fraud. The auditors' estimates of fraud prevalence were sensitive to the extremity of the irrelevant anchor.

A process that assigns an object to a class on the basis of how much it resembles the stereotypic class member is insensitive to prior probabilities (the representativeness heuristic). A red-flag approach is such a process [Libby, 1981]. Joyce and Biddle [1981b] considered whether auditors would fail to appreciate the low probability of fraudulent financial reporting in using a red-flag system. Auditors in their study were asked to assess the probability that a manager had been involved in fraudulent activities when a fraud prediction system signaled "fraud". Three pieces of information were provided: the positive hit rate of the system, the false positive rate, and the base
rate. They found that most auditors regressed their estimates toward the base rate, but the magnitude of the regression was insufficient. Auditors also failed to appreciate the significance of the false-positive rate. These findings suggest that use of a red-flag system may cause auditors to distrust innocent managers.

Albrecht and Romney [1986] used a survey approach to determine the predictive ability of the eighty-seven red flags in Albrecht, et al. [1980]. A questionnaire was completed by sixty-three practicing audit partners. Each was instructed to base his/her responses on a single audit engagement of which he/she is (was) in charge. Twenty-seven subjects considered engagements where fraud was found while thirty-six considered engagements where fraud was not believed to exist. If a red flag has predictive ability, the item should be present in fraud engagements and absent in non-fraud engagements. Albrecht and Romney concluded that roughly one-third of the red flags had predictive ability and one-third did not. The remaining one-third could not be evaluated because data were insufficient.

Pincus [1986] studied the relationship between individual differences in cognitive style and fraud-detection performance. The subjects' information acquisition patterns and decisions were observed as they completed a case modeled from an audit wherein a material undetected fraud existed. An individual's ability to separate a single piece of evidence from a complex background (field independence) was found to be positively related to success at the fraud detection task. Mixed results were found for predictions about an individual's ability to tolerate ambiguity.
Pincus [1987] investigated whether (1) the use of a red-flag questionnaire results in a more comprehensive and uniform information search pattern and (2) the approach is functional, i.e., raises the auditor's level of suspicion without creating excessive suspicion. Users of the red-flag questionnaire had a more comprehensive and uniform information search pattern. However, its use resulted in dysfunctional behavior when a material fraud existed -- the questionnaire users' assessment of the risk of fraud was lower than that of the nonusers.

Johnson, Jamal and Berryman [1988] examined the expertise developed by auditors to detect the presence of management fraud and errors in financial statements. The task of concurring-partner review was used to elicit this expertise. They posit that "... irregularities, including management fraud, are deliberately constructed framing manipulations [p. 3]." (See Tversky and Kahneman [1981] for a discussion of the framing effect.) Hence, their study focused on how auditors overcome framing effects in order to become effective detectors of material misstatements in financial statements.

Seven audit partners performed a concurring partner review of cases developed from annual and 10K reports of actual companies. No audit workpapers were provided. In the fraud case, the perpetrator purposely set up conditions suggesting a problem representation in which the company appeared to be a high-growth firm. This representation was incorrect -- the firm was actually in decline.

Problem representations and knowledge structures used by each subject were inferred from analysis of concurrent verbal protocols and rating judgments. Only subjects who had industry knowledge and were
able to combine cues configurally were able to overcome the initially incorrect problem representation and detect a pattern of income manipulation. Subjects who did not have industry experience attended to the same cues but were unable to combine them. This result supports a recommendation of the National Commission on Fraudulent Financial Reporting [1987, p. 55] which suggests that familiarity with the client's industry as a qualification of a concurring partner.

2.2 Implications of Irrelevant Evidence for Decision-making

A critical component of expert judgment is the ability to appropriately characterize and use information that varies in its degree of relevance. Ideally, experts should select and use only the most relevant information. In a literature review prepared by Gaeth and Shanteau [1981], approximately 125 papers were reported which found some influence of irrelevant evidence. In basic studies, irrelevant evidence was found to reduce rates of learning, increase errors, slow behavior, and increase task difficulty. In applied studies, irrelevant evidence was found to reduce accuracy, lower consistency, and impede performance. The influence on experienced decision-makers in applied settings is especially noteworthy. (See Shanteau and Gaeth [1983, p. 1-8] for a summary of the results of their literature review.)

2.21 Applied Research on the Influence of Irrelevant Evidence

Castellan [1973] reported that subjects were not able to learn to ignore irrelevant cues in a multiple-cue probability learning task--even after 400 trials. In such a task, subjects learn about the
relationship between cues and events in order to predict which event will occur on a given trial. The validity of a cue (its relevance) is measured by the binary analog to the product-moment correlation between the cue and the event. When the cue and the event are independent the validity is zero and hence should be ignored in the judgment process.

The tendency to ignore irrelevant cues was directly related to the validity of the relevant cue. The failure to ignore irrelevant cues was most serious when the relevant cue was of moderate validity. In this case, there was a large performance decrement as the number of irrelevant cue dimensions increased. Subjects had a great deal of difficulty discriminating the relevant cue from the irrelevant cue has the validity of the relevant cue diminished. Although these effects decreased over trials, they did not disappear.

Decision-making often involves accumulating evidence until uncertainty is reduced to an acceptable level. Levine [1973] studied the influence of the degree of information relevance on information-seeking behavior. The information seeking task involves gathering information in order to choose among alternatives. The task has several discrete stages. At each stage, the decision-maker must decide whether to choose an alternative or whether to gather additional, costly information. The objective is to make the correct decision is quickly as possible. This paradigm emphasizes the role of a person as a sequential information processor.

Subjects purchased more information before they made an initial decision and delayed the time of making that decision when the proportion of relevant information was low, but correct decisions were
reached in the same number of trials under all conditions. Likewise, latencies of correct choices were not influenced by the proportion of relevant information. If the trial number or the latency of a correct decision is used as a measure of performance, this result suggests that subjects were capable of integrating information regardless of its degree of relevance. In contrast, subjects in the Snapper and Peterson [1971] study were not appropriately sensitive to the relevance of the available information.

Rice [1975] found that altering biographical information that had little or no importance for effective teaching, e.g., family income (welfare versus $6200), father's occupation (bartender versus minister), influenced the degree to which school administrators were willing to grant a prospective teacher a job interview. Subjects were given one of two variations of a teacher's application form. They were asked to rate their degree of interest in seeing the teacher for a future interview. The application forms were identical except for six categories of information that were rated by a panel of administrators as related only minimally or not at all to effective teaching performance. On one form these categories were given positive responses; on the other, negative responses. Subjects used information unrelated to job performance far more than they realized.

Youngblood and Himmelfarb [1972] studied the effects of prior neutral messages on the degree to which subsequent positive or negative descriptions of an object (a country and its people) influence a subject's attitude about the object. They found that a neutral message received prior to a positive (negative) message resulted in less extreme
judgments than did exposure to only the positive (negative) message. The neutral information by itself produced a positive attitude change.

In Troutman and Shanteau [1977] (s.a. Shanteau [1975]), nondiagnostic samples of evidence were presented in a bookbags-and-pokerchips sequential inference task. Their subjects drew beads from a box that contained either 70 red, 30 white, and 50 blue beads or 30 red, 70 white, and 50 blue beads and inferred the probability that the beads were predominantly white. Unknown to the subjects, the experimenter could manipulate the drawing of the beads. The composition of the boxes made it possible to investigate three types of nondiagnostic information: (1) a sample that included two pieces of diagnostic information that canceled each other out (an equal number of red and white beads); (2) a sample of beads equally represented in the two boxes (just blue beads); and (3) a sample containing no beads. Less extreme responses were observed for each type of sample in both sequential- and aggregate-response modes. Control sequences were used to verify that subjects thought the samples were nondiagnostic. Both Troutman and Shanteau [1977] and Youngblood and Himmelfarb [1972] were working in the Anderson [1968] tradition.

Gilovich [1981] reported the results of three studies to demonstrate that (1) associations with past experiences influence people's judgments, (2) associations might be formed on the basis of representativeness, and (3) representativeness is determined by both relevant similarities and irrelevant or incidental features. The results indicated that when irrelevant associations are made available to subjects, they will use these sources of information in making
predictions. This result was obtained from subjects who had a great deal of experience in the domain studied (varsity football coaches from major universities rating hypothetical college football players in terms of their potential as professional football players).

The experimental materials were designed to lead subjects to form irrelevant associations. In debriefing interviews, subjects stated that they noticed the manipulations but unanimously denied that the manipulations had influenced their decisions. Thus, it appears that the manipulations used in these studies led subjects to form impressions that were difficult to ignore once formed. This result is similar to that of Rice [1975] and the hindsight bias [Fischhoff, 1975], other examples wherein subjects were unaware of the effect on their decisions of information they should have ignored.

Nagy [1981] studied the influence of job-relevant information (experience and recommendations) and job-irrelevant information (sex, age, and physical attractiveness) on hiring decisions made by both undergraduate subjects and experienced personnel managers. Her results indicate that (1) undergraduates made use of all three irrelevant dimensions, (2) except as to sex in the cases of the top eight applicants, experienced personnel managers did not rely on job-irrelevant information, (3) the effect of job-irrelevant information was reduced when subjects reviewed job descriptions and hiring regulations.

In a series of studies, Gaeth and Shanteau [1984] (s.a. Shanteau and Gaeth [1983]) investigated the influence of irrelevant evidence on soil judgments and personnel selection. Their goals were to (1) to determine whether experienced judges are influenced by irrelevant
information and (2) to evaluate and compare the effectiveness of two training procedures design to reduce its influence -- lecture training and interactive training. Two major conclusions can be drawn from these studies. First, irrelevant information had a detrimental effect on experienced decision makers. (While relevant evidence is more important, irrelevant evidence can play a decisive role, e.g., when making distinctions among similar job applicants [Shanteau and Gaeth, 1983, p. 42].) Second, training can reduce the influence of irrelevant factors on experts. In the soil judgment task, only the interactive training proved successful. In the more cognitively-oriented personnel selection task, lecture training was almost as effective as interactive training.

Nisbett, Zukier, and Lemley [1981] (s.a. Nisbett and Ross [1980, chapter 7]) reported the results of two studies. In the first study, subjects made predictions about the shock tolerance and movie attendance behavior of hypothetical male undergraduates. Subjects were given various items of information about these target individuals, items that had previously been rated either highly diagnostic of the behavior studied or nondiagnostic. Some subjects were given only a diagnostic item -- the target's major field of study (sciences versus humanities). Other subjects were given both the diagnostic item and a series of nondiagnostic items, e.g., the target individual's home town. A between-subjects design was used. Subjects given both nondiagnostic and diagnostic information made less extreme predictions than subjects given only the diagnostic information. Weak diagnostic information had no effect on predictions when nondiagnostic information was added.
In the second study, these authors used experienced social workers making a judgment of a sort that is common in their profession -- the likelihood that several middle-class, male social-work clients were child abusers. A within-subjects design was used; each subject was exposed to a diagnostic-information-only condition and to conditions providing varying amounts of nondiagnostic information. Increasing amounts of nondiagnostic information led to less extreme predictions when combined with one or two diagnostic items (suggesting a child abuser) but had no statistically significant influence on judgments based on counterdiagnostic evidence (suggesting not a child abuser). There was no detectable effect of increasing the amount of nondiagnostic information in the absence of diagnostic or counterdiagnostic items.

In interpreting these results, Nisbett, et al relied on Tversky's [1977] analysis of similarity judgments. They suggested that nondiagnostic information serves to make the target less similar to the judge's conception of the relevant outcome than it would be if there were no nondiagnostic information. No direct evidence was gathered to test whether similarity judgments or some other judgment mechanism produced the results. The major implication of Nisbett, et al's project is that stereotypic labels that exert a pronounced effect on judgments about abstract, undifferentiated targets may have little effect when judgments are made about concrete, individuated targets.

2.22 Information Load

studied the influence of information load in an accounting context. Information overload is the decline in performance as the decision-maker assimilates so much information that he/she exhausts his/her limited capacity to process and store information.

Information load, as operationalized in those studies, is qualitatively different from the phenomenon examined in this study. As noted in Casey, "... information load is defined as the relative amount of data judged by experts [bankers] to be useful for making decisions [about lending] [1980, p. 38]." In the current paper, evidence auditors consider useless, not evidence they consider useful, is manipulated. Iselin [1988, p. 150] criticized both Casey [1980] and Shields [1983] for confounding information load with irrelevant data -- the focus of the current study.

2.3 Conceptual Structures and the Processes that Operate on Them

The literature on object- and person-perception is the basis for the current study. Contrasted to an abstract notion such as fraudulent financial reporting, the concepts used in such studies -- artificial concepts such as geometric figures, natural concepts such as cups, and social concepts such as types of people -- are relatively concrete. Nevertheless, the categorization rules and conceptual structures used in those studies may not be fundamentally different from those used in an auditing (financial-reporting) context. To the extent that there are differences, insights can be gained by comparing analyses of auditor judgments framed as categorization decisions with results obtained in the studies of object- and person-perception.
Until recently, research on the classification of natural objects had been guided by the classical view. It describes classification performance under the following conditions: (1) membership is determined by a set of singly necessary and jointly sufficient features; (2) all members possess all the defining features and are therefore equally good examples of the category; (3) at any level of abstraction, there are distinct boundaries between categories; (4) all features are equally important in determining category membership [Cantor and Mischel, 1979, p. 9].

These conditions describe categorization performance for artificial objects such as geometric objects, but they do not describe categorization performance for most naturally occurring objects. Natural categories are "fuzzy sets" that violate the classical view's all-or-nothing rule. Rosch [1975] suggested that the categories people use are organized around the best examples of a concept and that other, less clear, examples form a continuum away from the best examples. Rips, Shonen, and Smith [1973] and Rosch [1978] reported that subjects could reliably rate whether members of a category were better or worse examples of the category.

The "fuzzy set" approach suggests that a categorization decision is probabilistic, and that borderline cases result in overlapping, unclear boundaries between categories at the same level of abstraction. Such a view deemphasizes the search for necessary and sufficient defining features and instead emphasizes operational definitions of how well objects fit an individual's conception of a category label.
Research on the rules individuals use to determine whether an object is a member of a "fuzzy" category has followed two major schools of thought. One approach suggests that concepts are learned by retaining information about known instances (exemplars) of a category. New instances are categorized on the basis of their similarity to these discrete, known instances. This approach has been labeled the exemplar view. The context theory of categorization behavior proposed by Medin and Schaffer [1978] (Nosofsky [1984]) is an exemplar model.

The other approach to the categorization of objects into "fuzzy" categories is the prototype view. Proponents of this view assert that categories at the same level of abstraction are essentially continuous. The distinctions among categories are determined by "... conceiving of each category in terms of its clear cases rather than its boundaries [Rosch, 1978, p. 36]." Rosch and Mervis [1975], Tversky [1977], and others have studied attributes of prototypical category members. They have shown that prototypical members are easier to learn, classify, name, and image than are nonprototypical members (see [Rosch, 1978]). In a priming paradigm, categorization is faster when primes are typical members of the category rather than atypical members [Rosch, 1975].

The prototype view of category structure does not itself have implications for observed behavioral effects, e.g., judgments of how good a member of a category an object is (graded typicality). It is necessary to augment the structural theory with a theory of a psychological process that operates on that structure [Forsyth, 1984].

Like the exemplar view, the prototype view emphasizes the probabilistic nature of categorization decisions and assumes that new
instances are judged members of a given category by the degree to which they possess the common characteristics of the category. It differs from the exemplar view in that new instances are compared to an abstracted image or set of features that define the central tendency of the salient, recurring properties of category members (prototype) as well as, or instead of, known instances. This prototype is not restricted to a set of necessary and sufficient features nor is it necessarily recognizable as an example of the concept.

Prototype models typically describe categorization in terms of a pattern-matching process. A pattern-matching theory describes the process individuals use to determine category membership as a comparison between the features of an object and the features of the prototype. The degree to which the two sets of features match is accumulated feature by feature. An object is recognized as an instance of the category if the match between the object and the prototype reaches some critical level.

In representing an object or concept, one can first distinguish component properties and holistic properties. A component property is one that describes something less that the entire entity. A holistic property is an isomorphic, unanalyzable description of the entity. Components can be characterized by dimensions or features. Dimensions are quantitative components. Features are qualitative components. Thus, if two objects differ with respect to a particular dimension, one object has more of the dimension than the other. If two objects differ with respect to a feature, one object has it but the other does not; a feature either exists or it does not exist. (See Smith and Medin [1981,
p. 131] for an illustration of a featural, dimensional, and holistic representation of a cup and Garner [1978] for a more complete description of the properties of objects.)

In a dimensional approach, objects are represented by a list of values they take on a set of necessary dimensions. In a featural approach, objects are represented by lists of features, and relations among objects are captured in terms of common and distinctive features. The featural approach is the most commonly used, because (1) rated category membership judgments (similarity judgments) violate the metric assumptions of distance-based prototype models, (2) quantitative variations can be expressed by sets of nested features [Gati and Tversky, 1982], and (3) a feature-matching process is not metric in nature [Tversky, 1977].

2.4 Influence of Prototypes on Perception and Processing of Evidence

In the current context, two sources of information are available: (1) prior, structured knowledge about the audit domain and (2) specific facts about a client. The research described in this section demonstrates that prior, abstract forms of knowledge (prototypes) serve to structure and reduce the flow of incoming information, preventing cognitive chaos [Cohen, 1981]. The same processes may influence the way auditors structure and use audit evidence.

The auditor's mental representation of audit evidence is called an impression (a term used by Hamilton, Kratz, and Leirer [1980]). Forming an impression is an active process in which the auditor imposes structure on the available audit evidence in an effort to develop a
coherent mental representation. This mental representation, not the individual facts made available to the auditor, is the basis for subsequent judgments about the client.

Prototypes could conceivably influence how individuals process information about an object in several ways. For example, prototypes may focus one's attention on a particular aspect of an object, thereby making that aspect more salient; or they may cause the selective retrieval of information from memory; or they may influence the way individuals interpret certain aspects of an object. Any combination is possible. Such effects need not be limited to one process [Hamilton, 1979, p. 68].

A number of studies in the area of person-perception have been concerned with these kinds of processes. They have consistently found that when prototypes (stereotypes) are brought to mind, individuals form expectations. Once formed, expectations influence how individuals interpret information. The influence is in the direction of confirming expectations; individuals "see" information that confirms their expectations and "do not see" incongruent information. The consistency of these findings across a variety of settings suggests that this phenomenon may have widespread implications.

2.41 Confirmation Bias

In Duncan's [1976] study, white college students watched a videotape of a discussion between two males. The race (white or black) of the actors was manipulated. Each time the experimenter signaled, the subjects classified the behavior that had just occurred. The last
signal was given immediately after the discussion had become heated and one actor had shoved the other. Four coding categories were used in classifying the shoving incident — playing around, dramatization, aggressive behavior, and violent behavior. When the shover was white, the shove was viewed as playful. When the shover was black, it was interpreted as aggressive or violent behavior.

Snyder and Uranowitz [1978] had subjects read a long narrative about a female named Betty K. The narrative included information about Betty's childhood, education, and adult career. Subjects returned a week later and were given a recognition memory test on various details of the narrative. Before the recognition test, half the subjects were told Betty was by then living a lesbian lifestyle, while the other half were told she was by then living a heterosexual lifestyle. The subjects' memories of the details of Betty's life were biased in the direction of the lifestyle-set. It appears that subjects reconstructed the information they had received in a manner consistent with the prototype which had been activated prior to the recognition task.

Subjects in the Rothbart, Evans, and Fulero [1979] study were given a list of fifty behaviors associated with a group of males. Half the subjects were told that the men were "more friendly than average" while the other half were told that the men were "more intelligent than average". Friendly and intelligent behaviors were equally represented in the list. Subjects given the friendly-set thought the list contained more friendly than intelligent behaviors, and vice versa. In a free-recall task, subjects were able to recall more behaviors consistent with their expectations.
Rothbart, et al. suspected that subjects expecting an intelligent group might be more inclined to see items originally pretested as unrelated to intelligence as confirming intelligent behavior, and vice versa. To test this possibility, the subjects were given a list of fifty behaviors from the original list and asked to place each in the most appropriate of five categories. No evidence for their recategorization hypothesis was found.

Cohen [1981] had subjects view a videotape of a woman having dinner with her husband. Half the subjects were told that the woman was a librarian. The other half were told that she was a waitress. Equal numbers of characteristics stereotypic of librarians and waitresses were included in the tape. Following the video, subjects were asked a series of two-alternative, forced-choice questions about the woman in the tape. In each question, one choice was stereotypic of a librarian, the other of a waitress. Only one choice was correct. If knowledge of the woman's occupation facilitated the processing of prototype-consistent features, librarian-set subjects should have recognized librarian features more accurately, and vice versa. The results generally supported Cohen's predictions. Ambiguous information appeared to have been interpreted as prototype-consistent.

Nisbett, et al. [1981] suggested that just as ambiguous information is sometimes interpreted in terms of stereotypes and other knowledge structures (e.g., Cantor and Mischel [1979] and Hamilton [1979]), nondiagnostic information might be reinterpreted to seem supportive of evidence that is clearly relevant to the decision at hand. The relevant information may act as an initial hypothesis about the target, which may
be confirmed by the nondiagnostic evidence. This proposition was not substantiated when they manipulated "routine, neutral, unexceptional facts about a person [p. 271]."

2.42 The Role of Confirming/Disconfirming Evidence in Rule-discovery Tasks

The role of confirming and disconfirming evidence has been studied as it relates to information search strategies in a hypothesis-testing task (e.g., Wason [1960], Kida [1984], and Klayman and Ha [1986]). These researchers focused on the process of deciding where to look for information and found that individuals tend to test those cases that have the best chance of verifying current beliefs, rather than falsifying them. This has been labeled a "confirmation bias" or "positive-test strategy". In the research cited in Section 2.41, researchers have asked questions about how individuals interpret what they find. It is this latter literature that is germane to the current study.

2.5 Chapter Summary

The auditing profession is responding to the public and private sectors' desire to accept more responsibility for the detection and deterrence of fraudulent financial reporting. A number of researchers have attempted to identify the attributes, called red flags, that are uniquely associated with the stereotypical fraud-prone company. Other researchers have used these red flags to develop discriminant functions that classify clients into, for example, a high-risk or low-risk group.
These statistical models (1) presume the existence of a stereotypical fraud-prone company and (2) are essentially pattern-matching functions that compare facts about a client with the attributes of the stereotypical fraud-prone company.

The model developed in the next Chapter adopts this similarity-to-a-stereotype perspective. The model describes the role played by seemingly irrelevant evidence, thus extending the literature on the implications of irrelevant evidence for decision-making. The model is consistent with the featural, prototype view of categorization behavior discussed in the third section of this Chapter.
3.1 Structure of the Audit Domain

Consider companies of two types; call them "type-h companies" and "type-1 companies". Type-h companies have a high risk of financial-statement fraud. Type-1 companies have a low risk of financial-statement fraud. An auditor's mental representation of the fraud-related conditions at a type-h company and his/her representation of the fraud-related conditions at a type-1 company are rich in content and complex in form. When faced with the task of deciding whether a particular client, denoted $t_1$, is a type-h or a type-1 company, auditors compile from these representations a list of features they believe are relevant to the task [Tversky, 1977, p. 329].

A feature describes any aspect, property, or characteristic of an object [Gati and Tversky, 1984, p. 346]. Features may correspond to aspects of a company such as the existence of related-party transactions; they may represent concrete properties like the dollar amount of such transactions; they may reflect abstract attributes such as the complexity of a company's related-party transactions (see Tversky [1977, p. 329]). Let $H$ and $L$ represent those features of an auditor's conception of the fraud-related conditions at a type-h company and the
fraud-related conditions at a type-1 company, respectively, that he/she judges relevant to the fraud-risk assessment.

In some situations, features are easy to identify. In other situations, features are hard to enumerate or describe. The same object can be characterized in terms of different sets of features that correspond to different descriptions or to different levels of analysis [Gati and Tversky, 1984, p. 346]. Featural models can often be tested without specifying all the features associated with the objects under study [Tversky and Gati, 1982, p. 126-7].

Let denote the set of features that make up an auditor's mental representation of the audit evidence for client i. When an auditor is deciding whether client i is a type-h company or a type-1 company, his/her mental representation of the audit evidence can be characterized by three mutually exclusive feature sets. Let where i=1,2,3,... is a client designation and j=1,2,3 identifies the type of feature set. The Venn diagram in Figure 1 illustrates the relationship among H, L, and T_i.

The first feature set (j=1) describes incentives or opportunities to commit fraud in financial reporting at client i. These are called diagnostic features. The term diagnostic is used generically. It may refer to diagnostic, causal, incidental, or indicational data as operationalized by Tversky and Kahneman [1980].

The second feature set (j=2) describes aspects of client i's financial reporting process that are intended to prevent or detect fraud in financial reporting. These are called counterdiagnostic features.
The third feature set (j-3) describes aspects of client i that neither support nor contradict the auditor's conception of the fraud-related conditions at a type-h company or the auditor's conception of the fraud-related conditions at a type-l company. These features describe aspects of a company an auditor deems useful for some audit decisions but not useful for, and hence not relevant to, the fraud-risk assessment. These features are called nondiagnostic features. The focus of the empirical aspect of this study is the role such features play in the evidence-evaluation process.

The auditor's mental representation of the audit evidence may not include all the features of H and L, either because (1) they do not describe the company or (2) the auditor has not yet discovered them. Let H* (L*) represent those features of an auditor's conception of the fraud-related conditions at a company that has a high risk (low risk) of financial-statement fraud that are not included in his/her mental representation of the audit evidence.
T₁ Audit's Mental Representation of the Audit Evidence for Client i

H Auditor's Conception of the Fraud-related Conditions at a Company That Has a High Risk of Financial-statement Fraud

L Auditor's Conception of the Fraud-related Conditions at a Company That Has a Low Risk of Financial-statement Fraud

T₁₁ Diagnostic Audit Evidence

T₁₂ Counterdiagnostic Audit Evidence

T₁₃ Nondiagnostic Audit Evidence

H* Features of the Auditor's Conception of the Fraud-related Conditions at a Company That Has a High Risk of Financial-statement Fraud That are Not Included in the Auditor's Mental Representation of the Audit Evidence

L* Features of the Auditor's Conception of the Fraud-related Conditions at a Company That Has a Low Risk of Financial-statement Fraud That are Not Included in the Auditor's Mental Representation of the Audit Evidence.

Figure 1

Relationship Among Sets of Features
3.2 Formal Model of the Evidence Evaluation Process

"To say that concepts have a categorization function is to acknowledge that concepts are essentially pattern-recognition devices [Smith and Medin, 1981, p. 8]." When mental representations of evidence and possible outcomes are described in the same terms, it seems natural for individuals to use a pattern-matching process to evaluate the evidence [Tversky and Kahneman, 1983, pp. 235-6]. The pattern-matching theory used in the framework for the current study is a feature-matching process (s.a. Frederick and Libby [1986]).

Tversky's [1977] contrast model is used to characterize an auditor's assessment of the strength of the match between his/her mental representation of the audit evidence for client i and (1) his/her conception of fraud-related conditions at a company that has a high risk of financial-statement fraud and (2) his/her conception of fraud-related conditions at a company that has a low risk of such fraud. This feature-matching process is formulated as a compensatory combining rule in the form of a three-variable linear function.

Each of the three variables is a measure of the salience of a set of features. Salience is determined by two types of factors: intensive factors and diagnostic factors [Tversky, 1977, p. 342]. Intensive factors refer to the prominence or recognizability of the feature, such as brightness, loudness, and clarity. Diagnostic factors refer to the classificatory significance of the feature, which can depend on context or task effects, familiarity, or information content.

Given a set of qualitative assumptions, Tversky [1977, pp. 351-2]
has shown that there exists a scale $S$ and a nonnegative scale $f$ such that,

$$S(t_{1, h}) = \theta f(T_{11}) - \alpha f(T_{12} \cup T_{13}) - \beta f(\emptyset) \quad (1a)$$

where $S$ and $f$ are interval scales and $\theta, \alpha, \beta$ are nonnegative. The scale value $S(t_{1, h})$ represents an auditor's assessment of the strength of the match between the audit evidence for client $i$ and his/her conception of fraud-related conditions at a company that has a high risk of financial-statement fraud.

The scale value $f(\bullet)$ represents the salience of a set of features. The scale $f$ satisfies feature additivity among feature sets, i.e., $f(T_{12} \cup T_{13}) = f(T_{12}) + f(T_{13})$. The parameters $\theta, \alpha, \beta$ are weightings that reflect the relative influence of the three feature sets. The scale $f$ and the parameters $\theta, \alpha, \beta$ are task- and context-dependent. Einhorn and Hogarth [1981, pp. 63-4] provided a general overview of the contrast model.

The first term of equation (1a) describes how incentives or opportunities to commit fraud in financial reporting, if an auditor recognizes them, influence his/her assessment of the strength of the match between the auditor's mental representation of the audit evidence and his/her conception of fraud-related conditions at a type-$h$ company. The more features shared by the auditor's mental representation of the audit evidence and his/her conception of fraud-related condition at a company that has a high risk of financial-statement fraud, the more the client resembles a type-$h$ company.

The second and third terms of equation (1a) describe the influence of features not shared by the auditor's mental representation of the
audit evidence and his/her conception of fraud-related conditions at a type-h company. The second term describes the influence of distinctive features of the audit evidence -- including nondiagnostic evidence and counterdiagnostic evidence. The third term describes the influence of features of the auditor's conception of fraud-related conditions at a type-h company that he/she thinks are relevant to fraud-risk assessment but whose existence is not evidenced in the audit workpapers. The greater the number of features not shared by the auditor's mental representation of the audit evidence and his/her conception of fraud-related conditions at a type-h company, the less the client resembles a type-h company.

Similarly, an auditor's assessment of the strength of the match between the his/her mental representation of the audit evidence for client i and his/her conception of the fraud-related conditions at a company that has a low risk of financial-statement fraud is expressed

\[ S(t_1,1) = \delta f(T_{12}) - \alpha f(T_{11} U T_{13}) - \beta f(L^k) \]  

(1b)

where the definitions of the terms are easily derived from those used in equation (1a).

Nisbett, et al. [1981] applied Tversky's theory of similarity judgments to situations wherein individuals must predict, using incomplete information, the degree to which one of two competing linguistic labels better describes a person. They assumed that individuals form a subjective likelihood ratio whose numerator and denominator are assessed separately via judged similarity between the person and each label [Fischhoff and Beyth-Maram, 1983, p. 254]. Thus, the contrast model was used to describe the process by which individuals
assess the subjective correlates of the conditional probabilities in the numerator and denominator of a Bayesian likelihood ratio. In this way, the contrast model is a formalization of Kahneman and Tversky's [1972] representativeness heuristic (cf. Einhorn and Hogarth [1981, pp. 63]).

From a psychological perspective, the likelihood principle is a composition rule specifying both (1) the manner in which an individual combines subjective probabilities and (2) the relative contribution of each probability to the decision at hand. In the current paper, the likelihood principle is modeled as an averaging rule [Lopes, 1985; Shanteau, 1975]. Averaging is hypothesized to occur because individuals adopt a strategy in which a composite judgment is obtained by adjusting the implications of the evidence with respect to one outcome (e.g., the client is a type-h company) toward the implications of the evidence with respect to the competing outcome (e.g., the client is a type-1 company). Such a process is qualitatively equivalent to averaging, even though subjects would not average in any arithmetical sense [Lopes, 1985, p. 510]. The averaging rule is operationalized as an additive combining rule with weights reflecting the relative contribution of each component [Goldstein and Einhorn, 1987, p. 241].

When an auditor interprets a collection of evidence, one outcome may be more salient than the others. The most salient outcome is likely to control the individual's behavior [Wallsten, 1977, p. 90]. A variety of studies in the hypothesis-evaluation literature suggest individuals place more weight on the numerator than the denominator of a likelihood ratio [Wallsten, 1972; Wallsten, 1976; Fischhoff and Beyth-Marom, 1983, p. 247]. Although a ratio formulation is not used here, the cited
studies do document that the implication of evidence with respect to one outcome can exert a greater influence on the final decision than the implication of the same body of evidence for the competing outcome.

This possibility is represented in the current model by the decision weights $\Phi_1$ and $\Phi_2$. Three scenarios are possible. When an auditor interprets the audit evidence in order to make the fraud-risk assessment, he/she may (1) focus on the possibility that the client has a high risk of financial-statement fraud, (2) focus on the possibility that the client has a low risk of such fraud, or (3) give equal weight to the implication of the evidence (strength of the match) toward each outcome. The choice of focus depends on context and task effects as well as on the auditor's propensity to guard against type-I or type-II errors.

Assume that an auditor first evaluates the implication of the audit evidence for the more salient hypothesis. This assessment is then modified to account for the plausibility of the evidence with respect to the competing outcome, reminiscent of Tversky and Kahneman's [1974] anchor-and-adjustment strategy. When an auditor focuses on the possibility that client $i$ has a high exposure, this strategy is characterized by

$$D(T_i|h) - \Phi_1\, S(t_{i},h) - \Phi_2\, S(t_{i},l)$$

(2a)

where $D(T_i|h)$ represents an auditor's net assessment of the strength of the match between the his/her mental representation of the audit evidence for client $i$ and his/her conceptions of fraud-related conditions at (1) a company that has a high risk of financial-statement fraud and (2) a company that has a low risk of such fraud. The
parameters $\phi_1$ and $\phi_2$ are nonnegative weights that reflect how much the disparity in the salience of the outcomes influences the decision at hand. Decision weight $\phi_1$ is greater than decision weight $\phi_2$.

When $S(t_{1,h}) > S(t_{1,1})$, $D(T_{1|h})$ is always positive. If $S(t_{1,1}) > S(t_{1,h})$, $D(T_{1|h})$ may be either positive or negative depending on how much larger $\phi_1$ is than $\phi_2$. Hence, an auditor's net assessment of the strength of the match between his/her mental representation of the audit evidence for client $i$ and her/his conceptions of fraud-related conditions at type-$h$ and type-1 companies depends on both (1) how much more salient one outcome is than the other and (2) his/her independent assessments of how much the client resembles a company with a high risk of financial-statement fraud and a company that has a low risk of such fraud.

An auditor's net assessment of the strength of the match between his/her mental representation of the audit evidence for client $i$ and her/his conceptions of fraud-related conditions at type-$h$ and type-1 companies, when he/she focuses on the possibility that client $i$ has a low risk of financial-statement fraud, is expressed

\[ D(T_{i|1}) = \phi_1 S(t_{1,1}) - \phi_2 S(t_{1,h}). \]  

(2b)

When an auditor gives equal weight to each outcome, her/his net assessment is characterized by ($\phi_1 = \phi_2$)

\[ D(T_{i|\#}) = S(t_{1,h}) - S(t_{1,1}). \]  

(2c)

The definitions of the terms are easily derived from those used in equation (2a).

Equations (2a), (2b), and (2c) describe how audit evidence is combined and evaluated in order to make a decision. The result of the
modeled process is not an observable decision. Rather, the result is a subjective notion of the net implication of the audit evidence.

Since any observable behavior that reflects the processing of evidence also reflects a decision based on the process, an assumption about the relationship between the process and the decision must be made. Assume that an auditor can provide a rating response which is monotonically related to his/her evaluation of the evidence [Wallsten, 1977, p. 88]. Then, an auditor's rating of how well client i resembles a company that has a high risk of financial-statement fraud, when the auditor focuses on that possibility, is expressed

\[ M(t_i, h|h) = R( D(T_i|h) ) \]  

The response function R maps the auditor's subjective notion of the net implication of the audit evidence to her/his rating of that notion. R is strictly increasing in D(T_i|h) when D(T_i|h) is greater than the value the auditor associates with the base-rate risk of financial-statement fraud. M(t_i, h|h) is zero when D(T_i|h) is less than or equal to the base-rate assessment.

Equation (3) is one of six expressions relating an auditor's rating of how well a client resembles a company that has a high risk of financial-statement fraud to the manner in which s/he evaluated the audit evidence -- two outcomes times three anchor-and-adjustment strategies. Equation (3) is not used directly in this study. It is presented because it (1) describes how base-rate information is represented in this framework and (2) is the basis for the expression that is used to guide the data-gathering and data-analysis aspects of the study. The other five expressions can be obtained by combining
equations (2a), (2b), and (2c) with suitable response functions. They are not discussed further.

A version of equation (3) that is easier to interpret is obtained by substituting equation (2a) into equation (3)

\[ M(t_1, h|h) = R[ \Phi_1 S(t_1, h) - \Phi_2 S(t_1, 1) ] \]  

and substituting equations (1a) and (1b) into equation (4) and simplifying as follows:

\[
M(t_1, h|h) = R[ \Phi_1 \theta f(T_{11}) - \alpha f(T_{12} U T_{13}) - \beta f(H^*) ] - \\
\Phi_2 \theta f(T_{12}) - \alpha f(T_{11} U T_{13}) - \beta f(L^*) ] \\
- R[ (\Phi_1 \theta + \Phi_2 \alpha) f(T_{11}) + (-\Phi_1 \alpha - \Phi_2 \theta) f(T_{12}) + \\
(\alpha(-\Phi_1 + \Phi_2)) f(T_{13}) + (-\Phi_1 \theta) f(H^*) + (\Phi_2 \beta) f(L^*) ] \\
- R[ \Gamma_1 f(T_{11}) + \Gamma_2 f(T_{12}) + \Gamma_3 f(T_{13}) + \Gamma_4 f(H^*) + \Gamma_5 f(L^*) ] \]  

(5)

where \( \Gamma_1, \Gamma_5 \geq 0 \) and \( \Gamma_2, \Gamma_3, \Gamma_4 \leq 0 \).

Equation (5) describes how an auditor's assessment of the risk that client i's financial statements are affected by fraudulent financial reporting (the fraud-risk assessment) is influenced by both the audit evidence and the knowledge brought by the auditor to the task. This expression states that an auditor's fraud-risk assessment is an increasing function of evidence that describes incentives or opportunities to commit fraud in financial reporting \( T_{11} \) and a decreasing function of evidence that describes aspects of the client's financial reporting process that are intended to prevent or detect such fraud \( T_{12} \).

One might expect the fraud-risk assessment to be based solely on diagnostic and counterdiagnostic evidence. The third term in equation (5) states that an auditor's fraud-risk assessment is a decreasing
function of the salience of audit evidence that neither supports nor contradicts his/her conceptions of fraud-related conditions at type-h and type-1 companies -- nondiagnostic evidence ($T_{13}$). The focus of the empirical aspect of this study is the role such evidence play in the evidence-evaluation process.

The last two terms of equation (5) state that those aspects of the auditor’s conception of fraud-related conditions at type-h and type-1 companies that were not included in her/his mental representation of the audit evidence will influence his/her fraud-risk assessment. An auditor’s fraud-risk assessment is a decreasing function of his/her conception of fraud-related conditions at a company that has a high risk of financial-statement fraud whose existence is not evidenced in the audit workpapers. An auditor’s fraud-risk assessment is a increasing function of his/her conception of fraud-related conditions at a firm that has a low risk of such fraud whose existence is not evidenced in the audit workpapers.

Every auditor brings to an audit a different set of experiences. These experiences influence the auditor’s (1) conception of fraud-related conditions at type-h and type-1 companies, (2) judgment of the base-rate risk of fraudulent financial reporting, and (3) salience of audit evidence. Because of these individual differences, one auditor’s fraud-risk assessment as described by equation (5) need not necessarily agree with another auditor’s, given the same facts and circumstances--$H$, $L$, $R$ and $f(\cdot)$ may differ. In a between-subjects design, these individual differences are potential sources of variability that can inflate the type-II error rate.
As an audit proceeds, and the auditors learn more about the client, they may learn of conditions or circumstances that cause them to revise their fraud-risk assessments; auditor's must continually reexamine the likelihood of management misrepresentation [Carmichael, 1988, pp. 42, 46; AICPA, 1988a]. The revisions are influenced by nondiagnostic evidence, but not by the first two potential sources of variability described in the preceding paragraph. Consequently, the empirical aspect of the study focuses on how auditors revise their fraud-risk assessments as a result of evaluating additional audit evidence. The rest of this section describes the revision process.

The revision, essentially a difference judgment, is modeled by a subtractive model (see Forsyth [1988, p. 23, 47-8]). When the possibility that client 1 has a high risk of financial-statement fraud is the more salient outcome,

\[
C(t_1|h,h) = R[ D(T_1|h) - D(T_1^*|h) ] \tag{6a}
\]

where \( C(t_1|h,h) \) is an auditor's rating of how his/her fraud-risk assessment changed when additional audit evidence became available. \( D(\cdot) \) is defined by equation (2a). All audit evidence available when an auditor provides the rating response is \( T_1 \). The initial set of audit evidence, a subset of \( T_1 \), is designated \( T_1^* \). The response function \( R \) is a positive scalar.

When \( C(t_1|h,h) \) is positive, the auditor's fraud-risk assessment has increased. When \( C(t_1|h,h) \) is negative, the auditor's fraud-risk assessment has decreased. A rating of zero means the additional evidence does not cause the auditor's risk assessment to change.
For the purpose of discussion, let $T_1 = ((T_{11}^*, T_{11}^-), T_{12}^*, T_{13}^*)$ where $T_{11}^-$ is a new set of diagnostic audit evidence. Then, an auditor's rating of how much her/his fraud-risk assessment increased when additional diagnostic evidence becomes available is obtained by using equation (5) to characterize in equation (6a) the auditor's subjective evaluation of the initial audit evidence and his/her subjective evaluation of all the audit evidence available when the rating response is made as follows:

$$C(t_1|h,h) = R\left\{ \sum \left[ (\Gamma_1 f(T_{11}) + \Gamma_2 f(T_{12}) + \Gamma_3 f(T_{13}) + \Gamma_4 f(H^*) + \Gamma_5 f(L^*)) - (\Gamma_1 f(T_{11}^*) + \Gamma_2 f(T_{12}^*) + \Gamma_3 f(T_{13}^*) + \Gamma_4 f(H^*+T_{11}^-) + \Gamma_5 f(L^*)) \right] \right\}$$

This expression states that two factors cause an auditor's fraud-risk assessment to increase when he/she recognizes an incentive or opportunity to commit fraud in financial reporting: (1) the increase in the salience of diagnostic evidence and (2) the decrease in the salience of fraud-related conditions an auditor expects to find at a company that has a high risk of financial-statement fraud whose existence had not previously been evidenced in the audit workpapers.

Now, let $T_1 = ((T_{11}^*, T_{11}^-), T_{12}^*, (T_{13}^*, T_{13}^-))$ where $T_{11}^- = \{T_{11}^-, T_{13}^-\}$ is a set of additional diagnostic and additional nondiagnostic audit evidence. Then, an auditor's rating of how much her/his fraud-risk assessment changes when additional diagnostic evidence and additional nondiagnostic evidence becomes available is obtained by using equation (5) to characterize in equation (6a) the auditor's subjective evaluation of the initial audit evidence and his/her subjective
This expression states that three factors influence how the auditor's fraud-risk assessment changes when diagnostic evidence and nondiagnostic evidence become available. The first and third term of equation (7b) reflect how the additional diagnostic evidence causes the auditor's fraud-risk assessment to increase. The second term describes how the additional nondiagnostic audit evidence causes the auditor's fraud-risk assessment to decrease.

The difference in the ratings described by equations (7a) and (7b), i.e., $\Gamma_3 f(T_{13}^-)$, is a measure of how much the additional nondiagnostic evidence influenced the auditor's fraud-risk assessment. This measure is not subject to the potential sources of variability associated with the rating described by equation (5). This strategy, which is used in the empirical aspect of the study, is discussed in more detail when the dependent variable is described in section 5.2 of Chapter V.

When the possibility that client $i$ has a low risk is the more salient outcome,

$$ C(t_i|1,1) = R[ D(T_i|1) - D(T_i^*|1) ] $$

(6b)

where $C(t_i|1,1)$ is an auditor's rating of how his/her fraud-risk assessment changed when additional audit evidence became available.

$D(\cdot)$ is defined by equation (2b). When $C(t_i|1,1)$ is positive...
(negative), the auditor's fraud-risk assessment has decreased (increased). A rating of zero means the additional evidence does not cause the auditor's fraud-risk assessment to change.

When an auditor gives equal weight to each outcome,

$$C(t_1|\#) = 0$$

where $$C(t_1|\#)$$ is an auditor's rating of the change in her/his fraud-risk assessment when additional audit evidence becomes available. $$D(\#)$$ is defined by equation (2c). The interpretation of the signs of $$C(t_1|\#)$$ is the same as the interpretation of the signs of $$C(t_1|h,h)$$ in equation (6a).

Equations (6a), (6b), and (6c) describe how an auditor rates the change in her/his fraud-risk assessment when the same outcome is the more salient throughout the decision process. It is algebraically possible to combine pairs of equations (2a), (2b), and (2c) by a subtractive rule to generate six additional expressions. These describe how auditors rate the change in their fraud-risk assessment when the more salient outcome changes during the decision process. This scenario may describe situations wherein the influence of nondiagnostic evidence is measured by taking the difference between two independent judgments, each described by equation (5). However, it is unlikely that the more salient outcome could shift in situations where individuals explicitly rate the change -- the strategy used in this study. Consequently, all analyses are interpreted in light of equations (6a), (6b), and (6c).
3.3 Different Types of Nondiagnostic Evidence

Rice [1975] found that manipulating biographical information that has little or no relevance for effective teaching influenced the likelihood that school administrators would grant a prospective teacher a job interview. Each irrelevant item depicted either (1) an aspect of an applicant that administrators associate with a desirable family background or (2) an aspect that they generally associate with a less-desirable family background -- for example, whether the applicant's father was a minister or a bartender.

It is clear that out-of-role biographical information did influence the likelihood that an administrator would select an applicant for an interview. Further, the seemingly irrelevant evidence that cast a less favorable light on the applicant's family background had a different influence than the information that cast a more favorable light. There was no control group in Rice's study; thus it is impossible to compare the incremental effects of the desirable biographical information with those of the less desirable information.

The out-of-role biographical information in Rice's [1975] study is analogous to nondiagnostic audit evidence in this study. One is inclined to speculate (1) whether different types of nondiagnostic audit evidence exist and, if so, (2) how each type of nondiagnostic evidence influences an auditor's judgment. Three types of nondiagnostic evidence were considered: favorable evidence, unfavorable evidence, and neutral evidence.

An auditor is likely to find some aspects of any company desirable. For example, a company may have an innovative sales force tuned to the
market or a company may be willing to help the auditor by preparing schedules and reconciling confirmations returned with exceptions. These items cast a favorable light on the company. In this study, these conditions would be called favorable nondiagnostic evidence.

An auditor is likely to find some aspects of any company undesirable. For example, a company may be facing isolated instances of increased competition that are causing a modest crimp in margins at a few locations or a company may be recording a particular type of recurring transaction effectively but not efficiently. These items cast an unfavorable light on the company. In this study, these conditions would be called unfavorable nondiagnostic evidence.

An auditor is also exposed to nondiagnostic evidence he/she would not characterize as either favorable or unfavorable. For example, a general description of an order-entry/order-fulfillment process or a list of common account characteristics. In this study, these routine or unexceptional items would be called neutral nondiagnostic evidence.

Two ways to represent the distinction among these three types of nondiagnostic evidence in the current framework are: (1) the informative evidence may act to call-up an initial hypothesis with nondiagnostic evidence interpreted in such a way as to make it consistent with the informative evidence (a "confirmation effect") or (2) some types of nondiagnostic evidence may be more interesting and hence more salient than other types (a "salience effect"). These representations lead to different predictions that are tested in this study. These possibilities are discussed in connection with the analysis of the data in section 5.32 of Chapter V.
3.4 Chapter Summary

Two sources of information are available to an auditor: (1) prior, structured knowledge about the domain and (2) specific facts about a client. The theoretical framework developed in this Chapter relies on a similarity-based process to describe how auditors mentally integrate these two sources of information in order to make what are, in essence, category membership rating responses. The implications of this process are consistent with the results observed in this study, namely, that nondiagnostic evidence tempers the influence of evidence auditors consider useful for the fraud-risk assessment, producing judgments less extreme than those made in the absence of the nondiagnostic evidence. Three types of nondiagnostic evidence were defined.

For experimental convenience, auditors did not rate the absolute level of the risk that the study company's financial statements were fraudulent. Instead, they rated how what they learned about the company by reading workpaper excerpts caused them to revise their fraud-risk assessments. This latter rating was modeled by a subtractive rule.
4.1 Overview

4.11 Subjects

Practicing auditors took part in the study as consultants, as members of the validation group, and as participants.

Six members of the audit staff of the Columbus, Ohio, office of a national public accounting firm volunteered to provide guidance and advice and to perform a series of tasks that would help the author develop the experimental materials. Three were audit seniors; the others were audit managers. One audit senior was able to take part in the task described in Section 4.3 but was unable to otherwise take part in developing the experimental materials. These auditors are called "consultants".

Forty-two members of the audit staff of the Columbus, Ohio, office of another national public accounting firm took part in an exercise to validate the experimental materials the consultants had help to developed. Sixteen were experienced seniors and managers; the other twenty-six were junior staff and light seniors. None had previously been involved in the study. These auditors are called "members of the validation group".
One-hundred audit seniors participated in the experiment. Ninety-five were from the same public accounting firm that provided the six consultants. None had previously been involved in the study. These auditors are called "participants".

In developing the experimental materials, the author assumed that the participants would consist of equal numbers of heavy audit staff members and audit managers. Consequently, heavy staff members and managers were chosen to help create a set of materials that both groups, and presumably all those whose level of experience was within that range, could use. As it turned out, only audit seniors with an average of 40 months' auditing experience participated in the experiment. Their experience level was within the stated range and was only five months greater than that of the members of the validation group.

4.12 Experimental Materials

Each participant received a single page of general instructions and six items for use in the experiment. (A version of the experimental materials is reproduced in the Appendix.) Three items were used to measure how an auditor revises his/her fraud-risk assessment when he/she evaluates one of six combinations of evidence (two types of informative evidence -- diagnostic and counterdiagnostic -- times three types of nondiagnostic evidence -- favorable, neutral, and unfavorable). These items were: (1) a folder labeled "History and Background of Mr. Chip Company", (2) a folder labeled "Excerpts from the 1986 Mr. Chip Company Audit Workpapers", and (3) a folder labeled "Questions about Mr. Chip Company".
A narrative in the history and background folder described activities before 1986 at Mr. Chip Company, a hypothetical manufacturer and distributor of snack-food products. While reading this narrative, the participants formed initial impressions on various matters an auditor considers as he/she becomes acquainted with a client. These impressions, while tentative, later served as a baseline from which to evaluate the excerpts from the 1986 audit workpapers. The workpaper excerpts describe activities at the Company during 1986. Six versions of the audit workpapers were used -- one for each combination of evidence. The participants rated how their initial impressions of Mr. Chip Company's exposure to fraudulent financial reporting changed as they read the 1986 workpaper excerpts.

The other three items were used to measure how an auditor revises his/her fraud-risk assessment when he/she evaluates an individual piece of informative evidence. Two of these were sets of five paragraphs reproduced on separate 3 1/4 x 4 3/4 inch white stock cards. One set of cards presented descriptions of an incentive or opportunity to commit fraud in financial reporting (diagnostic evidence). The other set presented descriptions of an aspect of a company's financial reporting process intended to prevent or detect such fraud (counterdiagnostic evidence). The remaining item was a folder labeled "General Questions". It presented (1) instructions on how to use the cards and (2) a series of debriefing questions.
4.2 Create the List of Potential Items Instrument

The starting point for developing the experimental materials was a "list of potential items instrument". This list was used (1) to tentatively classify individual items of audit evidence into the categories diagnostic evidence, counterdiagnostic evidence, and nondiagnostic evidence and (2) to structure detailed discussions with the consultants. The audit evidence portion of the experimental materials evolved from this instrument.

The author used the following sources to create the items of evidence to include in the list of potential items instrument: Value Line [1988], annual reports of Pepsico (1986), Nabisco (1987), Lance (1985), and Borden (1985), a national public accounting firm's professional preparation problem, Albrecht and Romney [1986], Elliott and Willingham [1980], National Commission on Fraudulent Financial Reporting [1987], and Wallace [1984, pp. 186-94]. Care was taken to choose as many items of each type of informative evidence and each type of nondiagnostic evidence as would be needed.

Forty-eight individual items of audit evidence were included in the list. Forty-three items (the "short items) were from one to several sentences in length; five items (the "long items) were from one to several pages in length. The instructions covered such topics as a definition of fraudulent financial reporting, descriptions of the five types of audit evidence, a frame of reference to use in rating the individual items, and instructions on how to rate each item.

The consultants used the list of potential items document to rate how each item would influence their fraud-risk assessment. First, they
read a one-page narrative titled 'History and Background of Mr. Chip Company'. They used this narrative to form a general impression of Mr. Chip Company's exposure to fraudulent financial reporting. They then reviewed the forty-three short items and selected (1) the item describing a condition that creates the greatest incentive or opportunity to commit fraud in financial reporting at Mr. Chip Company and (2) the item describing an aspect of Mr. Chip Company's financial reporting process that would most likely prevent or detect fraud in financial reporting.

These two items were assigned the values -7 and 7, respectively, on the following scale:

\[
\begin{array}{cccccccccccccc}
\text{decreases skepticism} & \text{increases skepticism} \\
\text{-7} & \text{-6} & \text{-5} & \text{-4} & \text{-3} & \text{-2} & \text{-1} & \text{0} & \text{1} & \text{2} & \text{3} & \text{4} & \text{5} & \text{6} & \text{7}
\end{array}
\]

strongly moderately no moderately strongly change

The auditors then rated how strongly each of the other forty-six items (both short items and long items), taken individually, influenced their sense of professional skepticism regarding Mr. Chip Company's exposure to fraudulent financial reporting. Ratings were made on the -7 to 7 scale in relation to the effects of the items they assigned to the values -7 and 7. They recorded their judgments by circling the number that best reflected their opinions on the scale following each item.

The phase "professional skepticism regarding the Company's exposure to fraudulent financial reporting" was used only in the list of potential items instrument. The phase "risk that the Company's financial statements are affected by fraudulent financial reporting" (the fraud-risk assessment) was used in all discussions and all
subsequent materials. The consultants were aware of the change; their ratings of items in the list of potential items instrument would not have changed had the term "skepticism" been replaced by the term "risk" on the -7 to 7 scale.

4.3 Distribute the List of Potential Items Instrument to the Consultants

One-hour meetings were scheduled with each consultant. The list of potential items instrument was delivered to their offices a week before these meetings so that they could read the instructions and review the audit-related materials.

The purpose of the meetings was to introduce the author, explain the objective of the dissertation, review the data-gathering process emphasizing aspects in which they would take part, and answer any questions they had about the list of potential items instrument. Few questions were asked; all thought the instructions were clear. They completed the potential-items task alone during the following week.

4.4 Classify Items in the List of Potential Items Instrument

The absolute ratings of the items in the list of potential items instrument were used to identify those items that are diagnostic evidence, counterdiagnostic evidence, and nondiagnostic evidence. For each consultant, items rated in the range -7 to -2 were classified counterdiagnostic, items rated in the range -1 to 1 were classified nondiagnostic, and items rated in the range 2 to 7 were classified diagnostic.
At least five of the six consultants agreed that eleven short items are examples of informative evidence. They judged five diagnostic and six counterdiagnostic. Two of the counterdiagnostic items were combined to form a single counterdiagnostic item. The resulting ten items (two sets of five) were reproduced on individual index cards. Edited versions of these items were (1) were used in the experiment to measure how an auditor revises his/her fraud-risk assessment when he/she evaluates an individual piece of informative evidence and (2) were candidates for the informative workpaper excerpts. The editing process is described in Section 4.51.

Four consultants agreed that thirteen short items were nondiagnostic or only slightly informative. Of the five long items, five consultants judged one nondiagnostic and two counterdiagnostic; they disagreed on the other two items. Edited versions of these items were candidates for the nondiagnostic workpaper excerpts. The editing process is described in Section 4.52.

At least five of the six consultants agreed that nine short items are examples of nondiagnostic evidence. These were considered unambiguous examples of nondiagnostic evidence. There was considerable disagreement about the other ten short items; they did not receive further attention.

When asked to criticize the list of potential items instrument, several consultants mentioned that it was difficult for them to treat the forty-eight items separately; an auditor is inclined to integrate all the available evidence. This suggests that interactions among the individual items of audit evidence may have influenced some of the
ratings. Even if that happened, the data gathered by using the instrument are not thereby suspect. The instrument was used simply as a starting point to (1) tentatively classify individual items and (2) structure detailed discussions with the consultants. If the consultants had not in fact treated the individual items separately, the resulting interactions would have become evident as the workpaper excerpts that evolved from the instrument were tested. The consultants' criticism does highlight the importance of the steps taken to ensure that alternative interpretations of the study's results can be ruled out.

4.5 Edit Items Selected From the List of Potential Items Instrument

Two-hour meetings were held with each consultant. The purpose of these discussions was to (1) understand how they had interpreted individual items in the list of potential items instrument, (2) edit items selected from the instrument, and (3) consider the viability of a preliminary version of the workpaper excerpts.

4.5.1 Informative Items

The consultants ordered (1) the five cards that described either an incentive or an opportunity to commit fraud in financial reporting (diagnostic items) and (2) the five cards that described an aspect of a company's financial reporting process intended to prevent or detect such fraud (counterdiagnostic items) in relation to how much the situation described would increase or decrease their fraud-risk assessment, respectively. They then rated how much the situation on each card affected their risk assessment on a one-hundred point scale similar to
the scale reproduced on page 147 or page 149 in the Appendix. The scale value zero was labeled 'no influence'. The scale value one-hundred represented the amount the item they considered the most diagnostic/counterdiagnostic would increase/decrease their risk assessment. They verbalized as they completed the task and followed a flexible give-and-take strategy until they were satisfied with their ratings. Occasionally, the wording an item was changed to strengthen or weaken the strength of the item or remove an ambiguity.

The results of this task indicated that: (1) auditors generally agree as to what audit evidence is useful in assessing a company’s exposure to fraudulent financial reporting; (2) five auditors can have that many different opinions about the relative influence of items of evidence that all agree are useful for the fraud-risk assessment; (3) there is a much higher degree of consensus about the relative influence of counterdiagnostic items than about the relative influence of diagnostic items, and (4) there is substantial disagreement about whether incentives or opportunities play the larger role in increasing a company’s exposure to fraudulent financial reporting, even though all agree that a combination of an incentive and an opportunity is an explosive mixture. The analyses described in section 5.5 of Chapter V shed additional light on these results.

4.52 Nondiagnostic Items

Each consultant was asked to comment on each of (1) the thirteen short items that he/she had judged diagnostic or counterdiagnostic but that others had judged nondiagnostic and (2) the five long items. In
most cases, the apparent lack of consensus among the consultants was resolved with by making minor editorial changes.

4.53 History and Background Narrative

Several of the consultants noted that their firm’s audit philosophy stresses an evaluation of a client’s internal control environment. They suggested that a discussion of Mr. Chip Company’s control environment and a discussion of the audit firm’s prior relationship with the Company be part of the experimental materials. The last paragraph of the history and background narrative reproduced on Page 121 in the Appendix was added to address this concern. In subsequent discussions, several of the consultants mentioned that they were pleased with the change. This suggested that they were using the history and background narrative to form a baseline from which to evaluate conditions described in the items of audit evidence.

4.54 Preliminary Version of the Experimental Materials

Before these meetings, the context for the experiment was to be a review by an audit manager of the compliance-test workpapers prepared by audit staff members assigned to the Mr. Chip Company audit. The nondiagnostic evidence was to appear in notes or schedules referenced in the those workpapers.

Without exception, the consultants were skeptical of this context. A reason cited was that their firm does not use compliance tests unless the internal controls are deemed ineffective and the auditors wish to
estimate the extent of error. This suggests that the proposed context was not the neutral background sought, but instead was informative.

To remedy this situation, the context was changed from the compliance-test phase of the audit to the pre-field-work phase. In subsequent discussions, several consultants mentioned that this change was appropriate.

4.6 Create Sets of Nondiagnostic Workpaper Excerpts

4.61 Preliminary Versions

Three sets of nondiagnostic audit workpaper excerpts were created. Each set contained one of the three types of nondiagnostic evidence (favorable, neutral, unfavorable). Four sources were consulted in developing these excerpts: (1) the nine short items that were judged to be an unambiguous example of nondiagnostic evidence in Section 4.4, (2) feedback from the meetings described in Section 4.5, (3) the auditing manual of a national public accounting firm, and (4) Robert Morris Associates [1985, p. 71].

Some of the analyses described in Chapter V compare the relative influence of the three types of nondiagnostic evidence. These comparisons are meaningful only if the quantity of evidence in each set of nondiagnostic workpaper excerpts is roughly the same. Since there is no method to determine the psychological unit of analysis in this context, page space was used as a surrogate for the quantity of audit evidence.
4.62 Practitioner Review of Nondiagnostic Workpaper Excerpts

The three preliminary sets of nondiagnostic workpaper excerpts, the history and background narrative, and a series of questions about each set of workpapers were delivered to the offices of the consultants. The materials were collected a week and a half later.

Working alone, the consultants (1) rated the degree to which the evidence in each set of workpaper excerpts influenced their assessment of the risk that Mr. Chip Company's financial statements are affected by fraudulent financial reporting, (2) explained why they had chosen that response to the rating question, (3) explained how the workpaper excerpts had influenced their impression of Mr. Chip Company, (4) noted whether they had found the materials realistic, and (5) made suggestions for improving the workpaper excerpts. They completed this task three times, once for each version of the nondiagnostic workpaper excerpts. They were instructed to treat the materials in each set of workpaper excerpts independently of the materials in the two other sets and were encouraged to take a break between sets to avoid carryover effects.

Each consultant considered the unfavorable workpaper excerpts nondiagnostic. One judged the neutral excerpts diagnostic because she discovered an unintended segregation-of-duties problem. This internal-control weakness was eliminated. Two consultants judged the set of favorable workpaper excerpts slightly counterdiagnostic. They suggested several ways to modify the excerpts; their suggestions were incorporated in the materials.

A former partner in the accounting firm that furnished the consultants and most of the participants reviewed preliminary versions
of the workpaper excerpts and the instructions and questions that would accompany them. Most of his suggestions were incorporated in the experimental materials.

4.63 Validate Nondiagnostic Workpaper Excerpts

Combinations of items that are judged nondiagnostic individually may be judged informative when combined. To guard against this effect, the members of the validation group took part in an exercise to provide assurance that the three separate sets of workpaper excerpts were nondiagnostic when presented in the format later used in the experiment.

The validation exercise took place in a group setting during an in-house training program conducted by the accounting firm. The author administered the exercise and was available to answer questions as they arose; few questions were asked. The format and instructions were virtually identical to those later used in the experiment.

One-half of the members of the validation group were allowed to refer to the workpaper excerpts as they answered the questions relating to the excerpts; the other half were not allowed to refer to the excerpts. Consequently, there were six versions of the validation-group materials -- three separate sets of nondiagnostic workpaper excerpts times two sets of instructions. Each member of the validation group was randomly assigned to one of the six versions; they were reminded that they could ask questions at any time during the exercise.

They performed two tasks. One was a card-sorting/rating task. The responses from this task were used to select (1) the theme of the informative workpaper excerpts and (2) the labels for the rating scales.
reproduced on pages 138 and 141 of the Appendix. The analysis of the responses is discussed in Section 4.7.1. The other task was used to verify that the three sets of nondiagnostic workpaper excerpts were nondiagnostic when presented in the format later used in the experiment. This latter task is discussed in this Section.

The instructions required the members of the validation group to consider what they had learned about Mr. Chip Company. They were reminded that they had read a brief description of the Company's activities before 1985 and that the general impression they had formed while reading this narrative may have changed as they learned more about the Company from the 1986 Mr. Chip Company audit workpaper excerpts. They were then asked whether what they had learned from reading the workpaper excerpts had caused them to change their initial assessments of the risk that Mr. Chip Company's 1986 financial statements are affected by fraudulent financial reporting. To answer this question, they placing a check mark on a line following the answer "yes" or the answer "no".

Of the forty-two members of the validation group, one did not return the folder with her answer and two did not follow the directions (but did respond "no"). Of the thirty-nine who provided usable responses, thirty-eight responded "no". This result meant that none of the three sets of nondiagnostic workpapers suggested a change during 1986 in Mr. Chip Company's exposure to fraudulent financial reporting. Thus, it is unlikely that combining the individually nondiagnostic items of evidence had made the items (or the combinations) informative. A few
typographical errors where corrected. The resulting nondiagnostic workpaper excerpts were used in the experiment.

The average time to complete the validation-group exercise was forty-seven minutes; less experienced auditors took slightly longer than more experienced auditors. The members of the validation group indicated that (1) the task was realistic, (2) they did not feel rushed, and (3) they preferred being permitted to refer to the workpaper excerpts when answering questions about them. At the end of the session the author gave a short talk about the dissertation and explained the purpose of the validation exercise.

4.7 Select Informative Workpaper Excerpt Themes and Labels for the Rating Scale

Six sets of workpaper excerpts were used in the experiment -- two types of informative workpaper excerpts (diagnostic and counterdiagnostic) times three nondiagnostic types (favorable, neutral, and unfavorable). The selection of the three sets of nondiagnostic workpapers was described in Section 4.6. The selection of (1) the theme of the diagnostic workpaper excerpt (the "diagnostic theme"), (2) the theme of the counterdiagnostic workpaper excerpt (the "counterdiagnostic theme"), (3) the label of the value ten on the rating scale reproduced on page 138 of the Appendix, and (4) the label of the value ten on the rating scale reproduced on page 141 of the Appendix is described in this Section.
4.71 Preliminary Selection of Informative Themes and Labels

The members of the validation group completed two card-sorting/rating tasks -- one for diagnostic items and one for counterdiagnostic items. The responses of those who had two or more years of audit experience were used. The contents of the cards, the instructions, and the method of administration were the same (except as noted in this Section) as those later used in the experiment. The instructions for the card sorting/rating task are reproduced on pages 146-147 of the Appendix. The contents of the cards are reproduced on pages 153-158 of the Appendix. There were twenty-five sets of responses for the diagnostic card-sorting/rating task, and twenty-seven sets for the counterdiagnostic task.

The following method was used to select the diagnostic (counterdiagnostic) theme and label to the diagnostic (counterdiagnostic) rating scale: (1) assign the value ten to one card, (2) rescale the value assigned another card preserving the relative position the two cards have on the one-hundred point scale reproduced on Page 147 (Page 149) of the Appendix, and (3) count the number of rescaled ratings outside the range four to sixteen on the diagnostic (counterdiagnostic) rating scale. The pair of items that resulted in the fewest rescaled values outside the range four to sixteen on the corresponding twenty-point scale was chosen.

The item assigned the value ten became the label for the value ten on the diagnostic (counterdiagnostic) rating scale and the other item became the diagnostic (counterdiagnostic) theme. This method of selection insured that the greatest number of participants have ample
room to move on the twenty-point rating scales used in the experiment. In other words, this method minimizes floor effects and ceiling effects.

For the counterdiagnostic items, when the description of the change in the internal audit function (card A) was assigned the value ten and the description of the program to assess the Company's exposure to fraudulent financial reporting (card B) was rescaled, only one of twenty-seven rescaled ratings (4%) fell outside the range four to sixteen. This high degree of consensus about the relative importance of counterdiagnostic items is consistent with the discussion in Section 4.51.

For the diagnostic items, when the description of the Company's narrow margin of compliance with debt covenants (card D) was assigned the value ten and the description of the change in the management compensation program (card A) was rescaled, ten of twenty-five rescaled ratings (40%) fell outside the range four to sixteen. Five were on the high side and three were on the low side; an equal number audit seniors and audit managers fell outside the target range. Two members of the validation group, each having two years' experience, did not consider the compensation theme informative. No other combination of diagnostic items resulted in so few rescaled ratings outside the range four to sixteen. This relatively low degree of consensus about the relative importance of diagnostic items is consistent with the discussion in Section 4.51.

To bring more individuals into the target range, the debt theme was strengthened by adding the sentence "Management is convinced that this is a serious threat." This change was expected to move those on the
high side toward the target range and was expected to have a relatively small effect on those on the low side, since they already appreciate the seriousness of the lender's threat. The second paragraph of the diagnostic workpaper excerpt (reproduced on page 130 of the Appendix) was expected to help most participants to appreciate the relationship between the compensation theme and fraudulent financial reporting. The desired effect was obtained.

4.72 Final Selection of Informative Themes and Labels

One-hour meetings with the consultants were arranged. Approximately a week before these meetings, the following materials were delivered to their offices: (1) one complete version of the experimental materials, (2) the three sets of nondiagnostic workpaper excerpts, and (3) the two informative workpaper excerpts. Each consultant received a different version of the experimental materials. A cover letter explained the purpose of the meeting. Working alone, each consultant completed their version of the experimental materials and reviewed the other materials before the meetings.

The meetings were in two parts. During the first part, each consultant expressed his/her view on the following matters: (1) the clarity and conciseness of the instructions and questions, (2) the transparency of the tasks as they related to the measurement of the dependent variable as described in Section 5.2 of Chapter V, (3) the strength of the informative memos, and (4) the order of the workpaper excerpts. They found the materials realistic and easy to use. Minor editorial changes were made to the instructions, several questions, and
the diagnostic workpaper excerpt. One consultant suggested categories to used in several debriefing questions; the suggestions were used.

Combinations of items that are judged nondiagnostic may be judged informative when combined with an item that is judged informative. Nisbett, et al. [1981] called this "interactive diagnosticity". To guard against this effect, each consultant was asked whether the information conveyed in either informative workpaper excerpt, when combined with either set of nondiagnostic workpaper excerpts, caused any aspect of the nondiagnostic workpapers to suggest that Mr. Chip Company's exposure to fraudulent financial reporting had changed during 1986. None could find a pattern of interactive diagnosticity. Hence, this phenomenon is not a credible explanation for the effects observed in this study.

4.7 Chapter Summary

The experimental materials were developed with the assistance of practicing auditors whose experience levels were on average the same as those of the participants in the experiment. The goal was to create an instrument that (1) was as realistic as possible, (2) was suited to auditors of varying levels of experience, and (3) would take about fifty minutes to complete. The workpaper-excerpts portion of the experimental materials and the related cards for the sorting/rating task grew from a single list of individual items of audit evidence. As the workpaper excerpts evolved, they passed through a series of screens to rule out unintended effects. The diagnostic-unfavorable version of the experimental instrument is reproduced in the Appendix.
Algebraic models can guide empirical research by means of global tests and specific predictions. Global tests, of which conjoint measurement is an example, focus on the descriptive accuracy of the model as a whole. Alternatively, the researcher can investigate the predictive accuracy of the model by deriving specific predictions from the model and searching for systematic violations of them [Wallsten, 1979, p. 4-5]. Typically, the more immediate task is to demonstrate the predictive ability of the model. This study took that direction.

The objective of the experiment was to study the role nondiagnostic evidence plays when participants assess the risk that a company's financial statements are fraudulent. The model developed in Chapter III was used to define the dependent variable and to interpret the relationships among the data.

5.1 Method

5.11 Participants and Setting

One hundred practicing senior auditors participated in the experiment. The experiment was administered in two groups. Ninety-five auditors completed the tasks during a one-week development seminar of a large public accounting firm. Five auditors from a second large
public accounting firm completed the tasks in their local office. All had ample room in a quiet, comfortable setting. None of the participants had previously been involved in the study. The author was present throughout the two sessions to monitor the exercise, debrief the participants, and answer questions as they arose. Few questions were asked.

5.12 Materials

Each participant received a single page of general instructions and six items for use in the experiment. (A version of the experimental materials is reproduced in the Appendix.) Three items were used to measure how an auditor revises his/her fraud-risk assessments when he/she evaluates a combination of informative evidence and uninformative evidence. These items were: (1) a folder labeled "History and Background of Mr. Chip Company", (2) a folder labeled "Excerpts from the 1986 Mr. Chip Company Audit Workpapers", and (3) a folder labeled "Questions about Mr. Chip Company". Each folder included instructions for using the materials in it. Each folder was a collection of 8 1/2 x 11 sheets of paper.

The other three items were used to measure how an auditor revises his/her fraud-risk assessment when he/she evaluates an individual piece of informative evidence. Two of these were sets of five paragraphs reproduced on separate 3 1/4 x 4 3/4 inch white stock cards. One set of cards presented descriptions of an incentive or opportunity to commit fraud in financial reporting (diagnostic evidence). The other set presented descriptions of an aspect of a company's financial reporting
process intended to prevent or detect such fraud (counterdiagnostic evidence). The remaining item was a folder labeled "General Questions". It presented (1) instructions on how to use the cards and (2) a series of debriefing questions.

5.13 Procedure

Each participant received one of the six versions of the experimental materials, randomly assigned. The only difference among the six was in the materials in the folder labeled "Excerpts From the 1986 Mr. Chip Company Audit Workpapers". The excerpts included either (1) a workpaper describing a change in Mr. Chip Company's management compensation policy to reflect the company's increased emphasis on achieving budgeted targets (the diagnostic theme) or (2) a workpaper describing a new program whereby the management of Mr. Chip Company formally assesses the Company's exposure to fraudulent financial reporting (the counterdiagnostic theme). In addition, the workpaper excerpts included nondiagnostic evidence that was favorable, neutral, or unfavorable. A between-subjects design was judged necessary due to potential carryover effects (cf. Nisbett, et al. [1981]) and a desire to use a task that was as realistic as possible but could be completed in an hour.

Before the participants began the exercise the author made a few introductory remarks, including a review of the general instructions. The participants were reminded that they could ask questions at any time during the session. They were given an hour to complete the exercise.
The participants completed two tasks. They began the first by reading a one-page description of Mr. Chip Company's activities before 1986, which was in the folder labeled "History and Background of Mr. Chip Company". While reading this narrative, they formed initial impressions on various matters an auditor considers as he/she becomes acquainted with a client. These impressions, while tentative, later served as a baseline from which to evaluate the excerpts from the 1986 audit workpapers. They were instructed to consider how their initial impressions of Mr. Chip Company changed as they read the excerpts from the 1986 workpaper excerpts.

They were then asked "Did what you learned about Mr. Chip Company while reading ... the 1986 audit workpaper excerpts cause you to change your assessment of the risk that Mr. Chip Company's 1986 financial statements are affected by fraudulent financial reporting?" If they answered no, they described their impression of the Company. If they answered yes, they rated how much their assessment had changed. This rating was made on the twenty-point scale reproduced on Page 138 or Page 141 of the Appendix (depending on whether their fraud-risk assessment increased or decreased). They then explained their choice of rating. This first task is called the "case-study task"; the rating scale is called the "case-study scale"; the rating response is called the "context-dependent rating".

In the second task, the participants rated how much (1) five separate incentives or opportunities to commit fraud in financial reporting "... would cause you to increase your assessment of the risk that a typical company's financial statements are affected by fraudulent
financial reporting" and (2) five separate aspects of a company's financial reporting process that are intended to prevent or detect such fraud would decrease their assessment of that risk. Each of these ten situations (two sets of five) was described in a single paragraph reproduced on a white stock card. For each set of five cards, the participants (1) ordered the cards in relation to how much the situation posed on each card would change their fraud-risk assessment, (2) assigned the value one-hundred to the situation that would change their fraud-risk assessment the most, and (3) rated the other four situations on a one-hundred point scale by comparing them with the one assigned the value one-hundred and the label, having a scale value of zero, "does not increase (decrease) your risk assessment". The one-hundred-point scales are reproduced on page 147 and page 149 of the Appendix. This second task is called the "card-sorting/rating task"; the rating scale is called the "card-sorting/rating scale"; the rating responses are called the "context-free ratings".

The average time to complete the exercise was 49 minutes. At the end of each session, the author discussed the dissertation. The participants asked questions indicating that they were interested in the topic; most said they would like a copy of the results. When asked, none recognized the relationship between the card-sorting/rating task and the case-study task; this suggested that the exercise was not transparent. Eighty-two percent thought the task was realistic.
5.2 Dependent Variable

The context-free ratings are a measure of how an individual item of informative evidence affects an auditor's fraud-risk assessment. The context-dependent rating is a measure of how an item of nondiagnostic evidence and an item of informative evidence, taken together, affect an auditor's assessment of that risk. A measure of how nondiagnostic evidence affects such an assessment is obtained by using appropriate context-free ratings to remove the contribution of the informative evidence from the context-dependent rating.

The participants had seen the situations described on two of the cards used in the card-sorting/rating task in the case-study task. The situation described on one card was the same as the situation described in the informative workpaper excerpt (the "informative theme"). The situation on another card was the description of the value ten on the case-study scale (the "label theme"). The other eight cards were filler cards. The context-free ratings are described by equation (7a). The context-dependent rating is described by equation (7b).

A common unit of measure is necessary to compare a context-dependent rating with a context-free rating. This common unit was the scalar that when multiplied by the context-free rating of the label theme yields the product ten -- the value of the label theme on the case-study scale. This scalar was used to rescale the context-free rating of the informative theme to a value on the case-study scale. This strategy preserved the relative positions of the label theme and the informative theme between the card-sorting/rating scale and the case-study scale. The difference between the context-dependent rating
of the informative theme and the rescaled value of the context-free rating of the informative theme is a measure of the effect of the nondiagnostic evidence on the context-dependent rating of the informative theme, i.e., the term $T_3f(T_{13^-})$ in equation (7b).

The dependent variable is expressed

$$D_S = x_1 - \frac{10}{x_2}(x_3)$$

where $D_S$ is a measure of the influence of the nondiagnostic evidence on participant $s$'s context-dependent rating of the informative theme, $x_1$ is his/her context-dependent rating of the informative theme, $x_2$ is his/her context-free rating of the label theme, and $x_3$ is the context-free rating of informative theme.

When $D_S$ was positive, the nondiagnostic evidence caused the participant's fraud-risk assessment to be greater than it would have been had such evidence not been available. This is called an "enhancement effect". When $D_S$ was zero, the nondiagnostic evidence did not affect the participant's risk assessment. When $D_S$ was negative, the nondiagnostic evidence caused the participant's risk assessment to be less than it would have been had such evidence not been available. This is called a "dilution effect".

5.3 Effects of Nondiagnostic Evidence on Diagnostic Evidence

5.3.1 Observations

Fifty-three auditors with an average of 42 months' auditing experience received a diagnostic version of the instrument; forty-eight were from one auditing firm and five from another firm. Table 1 reconciles the number of participants who received a diagnostic version
with the number of observations included in the statistical analyses. Of these fifty-three participants, thirty-nine (82%) provided responses that were used in the statistical analyses. The responses of fourteen participants were not used because they did not answer all the questions necessary to compute the dependent variable; two participants did not complete the card-sorting/rating task, and twelve did not answer the context-dependent rating question because what they learned about Mr. Chip Company while reading the 1986 workpaper excerpts did not cause them to revise their fraud-risk assessments.

Of the twelve who did not answer the context-dependent question, one considered the diagnostic theme counterdiagnostic. Two of the twelve participants rated the diagnostic theme zero in the card-sorting/rating task, indicating that they did not appreciate the significance of that piece of diagnostic evidence. Four participants clearly did not understand the task or gave an explanation for their 'no-change' response that is uninterpretable. Five of the twelve participants rated the diagnostic theme marginally indicative of fraud in financial reporting in the card-sorting/rating task (mean rating of 31; median rank of 4). Here, the relatively weak diagnostic evidence had no effect on the participants' fraud-risk assessments when nondiagnostic evidence was presented. The effect of the nondiagnostic evidence was sufficiently strong to overcome completely the effect of the diagnostic evidence. This result is consistent with Nisbett, et al. [1981, pp. 261-3]. The cumulative effect of this phenomenon is particularly troublesome if auditors neglect the implications of diagnostic evidence that is singly "weak".
Table 1
Reconciliation of Number of Participants with Number of Observations
Diagnostic Versions

<table>
<thead>
<tr>
<th>Participants</th>
<th>53</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did Not Complete The Card-Sorting/Rating Task</td>
<td>2</td>
</tr>
<tr>
<td><strong>Risk Assessment Did Not Change:</strong></td>
<td></td>
</tr>
<tr>
<td>Judged Theme Counterdiagnostic</td>
<td>1</td>
</tr>
<tr>
<td>Rated Theme Nondiagnostic in the</td>
<td></td>
</tr>
<tr>
<td>Card-Sorting/Rating Task</td>
<td>2</td>
</tr>
<tr>
<td>Did Not Understand Task or Gave</td>
<td></td>
</tr>
<tr>
<td>Uninterpretable Write-up</td>
<td>4</td>
</tr>
<tr>
<td>Rated Theme Marginally Informative</td>
<td></td>
</tr>
<tr>
<td>in the Card-Sorting/Rating Task</td>
<td>5</td>
</tr>
<tr>
<td><strong>Observations Used in Statistical Analyses</strong></td>
<td>39</td>
</tr>
</tbody>
</table>
5.32 Dilution Effect

Statistical tests were performed to determine whether nondiagnostic evidence affects the degree to which auditors revise their fraud-risk assessments upward when they encounter additional diagnostic evidence. The analyses were two-sided, one-sample location tests under the null hypothesis that the nondiagnostic evidence did not affect the participants' fraud-risk assessments.

Summary statistics and the results of the individual location tests are presented in Panel A of Table 2. Four sets of responses were analyzed: responses to the diagnostic versions of the experimental materials wherein the nondiagnostic evidence was in turn favorable, neutral, and unfavorable, and responses to all diagnostic versions. Each type of nondiagnostic evidence had a pronounced effect on the participants' judgments. The average mean shift was -2.85; the range was from -1.61 for the neutral-treatment group to -4.30 for the favorable-treatment group. All p-values were less than .005, whether a signed-rank procedure or student's-t procedure was used. This is strong evidence that each type of nondiagnostic evidence had a moderating effect on the degree to which diagnostic evidence caused an upward revision in the participants' fraud-risk assessments -- a dilution effect.

With a dilution effect, the auditor's absolute level of sensitivity to the risk of financial-statement fraud is less than it would be in the absence of nondiagnostic evidence. When an auditor's sensitivity to such fraud is significantly less than it should be, he/she may select audit procedures that fail to give adequate consideration to the risk of
such fraud. When this happens, there is an increased probability that the auditor will fail to detect fraud in financial reporting when it exists.

The consequences of failing to detect financial-statement fraud are significant. The public may be affected: shareholders and creditors; those harmed when investor confidence in the financial markets is shaken; "... employees who suffer job loss or diminished pension fund value; depositors with financial institutions; the company's underwriters, auditors, attorneys, and insurers; and even honest competitors whose reputations suffer by association [National Commission of Fraudulent Financial Reporting, 1987, p. 4]." If the fraud is detected after the financial statements have been issued, the auditors may suffer the cost of damages, imprisonment, court costs, fines, rising liability insurance rates, and adverse effects on their reputations.

The study's results demonstrate that each of the three types of nondiagnostic evidence considered tempers the degree to which diagnostic evidence causes auditors to revise their fraud-risk assessments. The question remains: Does each type have the same affect? In Section 3.3 of Chapter III, two possibilities were proposed: a confirmation effect and a salience effect. These two interpretations are distinguishable in the current study.

The confirmation interpretation states that there will be a tendency to interpret nondiagnostic evidence in such a way as to make it consistent with the informative evidence. The informative evidence will call-up an initial hypothesis that is confirmed by the nondiagnostic evidence. This interpretation is consistent with the theoretical
framework developed in Chapter III, since the model describes how audit evidence is evaluated once it has been classified as diagnostic evidence ($T_1$), counterdiagnostic evidence ($T_2$), or nondiagnostic evidence ($T_3$)—the model does not describe how an item of evidence is judged a member of one of these three feature sets.

The confirmation interpretation leads to the expectation that nondiagnostic evidence will cause auditors to make more extreme revisions than warranted, especially when it is unfavorable. Positive values of the dependent variable (an enhancement effect) are consistent with this interpretation. Of the thirty-nine participants who provided usable responses, the value of the dependent variable was positive for only three (mean = .627). Of the fourteen participants in the unfavorable-treatment group, this value was positive (.500) for only one. Hence, the confirmation interpretation is not a credible explanation.

The salience interpretation states that some types of nondiagnostic evidence are inherently more interesting than other types. In particular, neutral items, such as flowcharts or general descriptions of order-entry/fulfillment processes, will attract less attention than either the favorable or the unfavorable items. This distinction was initially made by several consultants during the editing activities described in Section 4.4 of Chapter IV. It is captured in the model through the scale $f$ (an intensive factor). The salience interpretation leads to the expectation that the dependent variable will be negative, and that its magnitude will be greater (more negative) for the more salient types of nondiagnostic evidence.
This relationship is present in the data. All treatment means are negative; both the favorable and the unfavorable treatment means are smaller than the neutral treatment mean. This U-shaped relationship (seen when the treatment means are ordered favorable/neutral/unfavorable) is the focus of the analysis in the next section.

5.33 Favorable/Neutral/Unfavorable Distinction

The analysis of the U-shaped relationship is a one-way location test with three levels of the factor "type of nondiagnostic evidence". Certain assumptions about the distribution of scores within treatment groups must be met to use the analysis-of-variance technique [Keppel, 1982, pp. 85-87]. The practical consequences of failing to meet these assumptions have been determined; the general conclusion is that the analysis-of-variance technique is insensitive to moderate violations of the assumptions.

One assumption requires that the treatment scores be randomly sampled from individual treatment populations that are normally distributed. Violations of the normality assumption "... do not constitute a serious problem, except if the violations are especially severe ... [Keppel, 1982, p. 86]" (s.a. Norton [1952] as reported by Linquist [1953]). The violation is more serious with unequal sample sizes. The sample size, Shapiro-Wilk statistic, and a measure of skewness for each treatment group are presented in Panel A of Table 2. Plots of the scores in each cell are reasonably symmetric. Hence, violations of normality are not a serious problem.
Another assumption requires that the within-group variances be equal. Traditional statistical procedures to test this assumption are sensitive to departures from normality as well as to the presence of heterogeneity. To avoid this dual-sensitivity problem, Rogan and Keselman [1977] have suggested a simple heuristic: researchers should be concerned about violations of the homogeneity assumption if the ratio of the largest to smallest within-group variance exceeds ten (reminiscent of the Hartley test [Ott, 1984, p. 340]). Dividing the variance of the observations in the favorable treatment group by the variance of the observations in the unfavorable treatment group yields a value of 6.65, indicating that the heterogeneity is within acceptable limits. (This insensitivity of F to violations of homogeneity does not extend to analytical comparisons [Keppel, 1982, p. 116].)

A third assumption, independence of the error terms, was met through the random assignment of participants to treatments. Thus, the violations of the statistical assumptions inherent in the diagnostic data set are not so extreme as to limit the applicability of the analysis-of-variance technique.

The results of the omnibus F and analytical comparisons are presented in Panel B of Table 2. The effect of "type of nondiagnostic evidence" is statistically significant; F(2,36) = 3.24 (p = .0509). (A chi-square approximation of the Kruskal-Wallis statistic is \(X^2(2) = 3.44\) (p = .1794).) Thus, the magnitude of the moderating effect of at least two types of nondiagnostic evidence was different.

A statistically significant association does not indicate that the treatment effect is strong. The index omega-squared is a measure of the
magnitude of the treatment effect, reflecting the proportional amount of
the population variance accounted for by the independent variable
scale to interpret values of omega-squared in the behavioral sciences.
He suggests that a "large" effect is indicated by a value of .15 or
more, a "medium" effect by .06, and a "small" effect by .01. The value
obtained in this analysis is .103.

Which differences are real, that is, cannot be explained by
experimental error? The results of three analytical comparisons are
presented in Panel B of Table 2. Corrections for the violation of the
homogeneity assumption were made using a pooled estimate of the mean
square error based solely on the treatment groups in the comparison
[Keppel, 1982, pp. 116-7]. The favorable treatment mean is
statistically different from the neutral treatment mean (F(1,36) - 5.14,
p-value -.031) but not statistically different from the unfavorable
treatment mean (F(1,36) - 1.545, p-value -.232). The difference between
the unfavorable treatment mean and the neutral treatment mean is
marginally significant (F(1,36) - 2.75, p-value -.111) at a relatively
low level of power (β = .38 using the mean square error term from the
analytical comparison as an estimate of the treatment population
variance, a type-1 error rate of .05, and a sample size of thirteen
[Keppel, 1982, p. 118]). While there are no established conventions for
selecting a level of power, Cohen [1977, pp. 53-56] recommends that beta
equal .20. On that basis, the relatively low power of the
unfavorable/neutral comparison suggests that the difference between the
two treatment means cannot be explained solely by experimental error.
Thus, the degree to which nondiagnostic evidence dilutes the influence of diagnostic evidence is directly related both to the quantity of the nondiagnostic evidence [Nisbett, et al., 1981, p.270] and to its salience. In the current study, remarks of the consultants indicated that the routine, unexceptional nature of the neutral workpapers makes them less interesting (less salient) than either the favorable or the unfavorable workpapers. This belief is consistent with the relationship among the treatment groups; the more salient nondiagnostic evidence had a greater moderating effect on the degree to which diagnostic evidence caused the participants to revise their fraud-risk assessments than the less salient nondiagnostic evidence had.

One explanation of the relatively large dilution effect for the favorable treatment group that is inconsistent with the salience interpretation (and also with the similarity interpretation) is that one or more aspects of the nondiagnostic workpapers were seen as counterdiagnostic. Of the twelve participants in the favorable treatment group, six (50%) referred to only the diagnostic theme in explaining the magnitude of their context-dependent rating responses and four (33%) thought items in the nondiagnostic workpapers reinforced their decision to increase their fraud-risk assessment. No participant mentioned that he/she had interpreted any aspect of the nondiagnostic evidence as counterdiagnostic. Further, the scores in the favorable treatment group are not statistically different from the scores in the unfavorable treatment group. This result does not suggest an interpretation of nondiagnostic evidence as counterdiagnostic. Thus, the counterdiagnostic interpretation is not a credible explanation.
5.34 Conclusion for Diagnostic Evidence

The statistical analyses reveal that: (1) relatively weak diagnostic evidence had no effect on revisions of participants fraud-risk assessments when nondiagnostic evidence was presented, (2) each type of nondiagnostic evidence had a pronounced moderating effect on such revisions, and (3) more salient nondiagnostic evidence had a greater moderating effect on such revisions than less salient nondiagnostic evidence. These results are consistent with the similarity interpretation developed in Chapter III of how auditors judge a company's exposure to fraudulent financial reporting, when class membership is judged by similarity to the prototypical class member.

This moderating effect is described by the term $\Gamma_3 f(T_1^T)$ in equation (7a), which is negative if and only if (1) the nondiagnostic evidence is salient, i.e. $f(T_1^T)$ is positive and (2) the decision weight $\Gamma_3 - \alpha(-\phi_1 + \phi_2)$ is negative. The effect of varying the salience of the nondiagnostic evidence was demonstrated in the one-way analysis.

The decision weight $\Gamma_3$ is negative if and only if auditors (1) use the salient distinctive features of the audit evidence ($\alpha$ is positive) and (2) the implications of the evidence with respect to the possibility that the company has a high risk of financial-statement fraud control his/her behavior ($\phi_1 > \phi_2$). This latter conclusion is consistent with the results of previous research which suggest that individuals are more interested in how consistent the evidence is with the more salient outcome and less interested in how consistent the evidence is with the alternative outcome.
Not only did the participants use the nondiagnostic evidence far more than they realized, but their reports of how it influenced their ratings contradicted its actual affect. Nineteen of the thirty-nine participants that provided usable responses to the diagnostic version of the experimental materials (49%) stated that only the diagnostic theme affected their context-dependent ratings. Sixteen participants (41%) interpreted the nondiagnostic evidence as support for increasing their fraud-risk assessment. This result is further evidence of the inability of individuals to explain their decision processes (e.g., lack of self-awareness in policy-capturing studies), suggesting that techniques based solely on retrospective introspection may be of limited value in studying judgment and choice [Nisbett and Wilson, 1977; Einhorn, Kleinmuntz, and Kleinmuntz, 1979].
Table 2
Location Tests
Diagnostic Versions

Panel A: One Sample

| Version       | # of Observations | Mean | Std. Dev. | $|T|_1^1$ | skewness | $u^2$ | Median | $|S|_2^2$ |
|---------------|-------------------|------|-----------|---------|----------|-------|--------|---------|
| favorable     | 12                | -4.30| 3.92      | .0029   | -.072    | .1164 | -3.17  | .0020   |
| neutral       | 13                | -1.61| 1.52      | .0025   | .153     | .6190 | -1.76  | .0049   |
| unfavorable   | 14                | -2.77| 2.08      | .0003   | -.240    | .7177 | -2.42  | .0005   |
| combined      | 39                | -2.85| 2.80      | .0001   | -1.320   | .0005 | -2.00  | .0001   |

1 Student's t statistic; p-value under the null that the mean is zero
2 Shapiro-Wilk statistic; probability of obtaining a smaller value of the test statistic under the null that values of the dependent variable are a random sample from a normal distribution
3 Signed Rank statistic; p-value under the null that the mean is zero

Panel B: One-Way Layout

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>F Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>2</td>
<td>45.572</td>
<td>3.24</td>
<td>.051</td>
</tr>
<tr>
<td>fav to neu</td>
<td>1</td>
<td>(45.417)</td>
<td>5.14*</td>
<td>.031</td>
</tr>
<tr>
<td>fav to unfav</td>
<td>1</td>
<td>(15.211)</td>
<td>1.55*</td>
<td>.232</td>
</tr>
<tr>
<td>unfav to neu</td>
<td>1</td>
<td>(9.126)</td>
<td>2.75*</td>
<td>.111</td>
</tr>
<tr>
<td>Within</td>
<td>36</td>
<td>253.336</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>38</td>
<td>298.909</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* corrected for heterogeneity of variance
5.4 Effects of Nondiagnostic Evidence on Counterdiagnostic Evidence

5.41 Observations

Forty-seven auditors with an average of 39 months' auditing experience received a counterdiagnostic version of the instrument. Table 3 reconciles the number of participants who received a counterdiagnostic version with the number of observations included in the statistical analyses. Thirty-eight participants (81%) did not answer all the questions necessary to compute the dependent variable; seven participants revised their fraud-risk assessment upward and thirty-one did not revise their risk assessments.

The counterdiagnostic theme was a description of a new program, initiated by management, to formally assess Mr. Chip Company's exposure to fraudulent financial reporting. Ten participants believed that management, by instituting this program, evidenced a belief that the Company had a higher-than-average exposure to such fraud. These participants did not recognize the preventive nature of the program. This interpretation caused seven of these participants to revise their fraud-risk assessments upward. Three did not revise their risk assessments.

Seven of the thirty-one participants who did not revise their fraud-risk assessments rated the counterdiagnostic theme marginally informative in the card-sorting/rating task (mean rating of 10.4; median rank of 4). One interpretation of this result is that the relatively weak counterdiagnostic evidence had no effect on the participants' risk assessments when nondiagnostic evidence was presented; the effect of the
nondiagnostic evidence was sufficiently strong to overcome completely the effect of the counterdiagnostic evidence. This interpretation is consistent with the dilution effect observed in connection with relatively weak diagnostic evidence (Section 5.31).

Twenty-one of the thirty-one participants who did not revise their fraud-risk assessments stated that the evidence in the 1986 workpaper excerpts is consistent with the idea that the risk of financial-statement fraud at Mr. Chip Company is low. For these twenty-one participants, the mean rating of the counterdiagnostic theme was 75 (median rank of 2) in the card-sorting/rating task, suggesting that they thought the theme was highly informative relative to the other situations used in the counterdiagnostic version of the card-sorting/rating task.

Why would participants who thought the evidence in the 1986 workpaper excerpts consistent with the idea of a low-risk firm not revise their fraud-risk assessments downward? Three explanations were considered. First, the effect of the nondiagnostic evidence was sufficiently strong to overcome completely the effect of the counterdiagnostic evidence, suggesting that the moderating effect of nondiagnostic evidence is much stronger on counterdiagnostic evidence than on diagnostic evidence (cf. Nisbett, et al. [1981]). Second, the description of the Company's activities before 1986 was such a good example of a low-risk company that the participants could not revise their assessments downward and retain a minimum level of professional skepticism. Participants who received a diagnostic version of the instrument had plenty of room to revise their assessments upward; hence
the difference in the percentage of participants who chose to answer the respective context-dependent rating question.

Third, it is possible that the situations used in the counterdiagnostic version of the card-sorting/rating task were less informative than the situations used in the diagnostic version of the card-sorting/rating task. Thus, just as weak diagnostic evidence is swamped by the nondiagnostic evidence, the weak counterdiagnostic theme was also completely swamped by the nondiagnostic evidence. If this explanation were well-founded, one would expect to see a lack of consensus in the rankings of the counterdiagnostic cards. The analysis in Section 5.5 reveals that this is not the case. Hence, this explanation seems unlikely.

The participants' reluctance to revise their fraud-risk assessments in light of the counterdiagnostic theme is an interesting issue, one that will be pursued in subsequent studies. Fortunately, the effect of the nondiagnostic evidence is so strong that a sample size of nine was sufficient to demonstrate the implications of nondiagnostic evidence—the focus of the empirical part of the study—when auditors evaluate counterdiagnostic evidence.
### Table 3

Reconciliation of Number of Participants with Number of Observations

Counterdiagnostic Versions

<table>
<thead>
<tr>
<th>Description</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>47</td>
</tr>
<tr>
<td>Risk Assessment Increased</td>
<td>7</td>
</tr>
<tr>
<td>Risk Assessment Did Not Change:</td>
<td></td>
</tr>
<tr>
<td>Judged Theme Diagnostic</td>
<td>3</td>
</tr>
<tr>
<td>Rated Theme Marginally Informative in the Card-Sorting/Rating Task</td>
<td>7</td>
</tr>
<tr>
<td>Judged Evidence Consistent with a Low-risk Firm</td>
<td>21</td>
</tr>
<tr>
<td>Observations Used in Statistical Analyses</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>9</td>
</tr>
</tbody>
</table>
5.42 Dilution Effect

Statistical tests were performed to determine whether nondiagnostic evidence affects the degree to which auditors revise their fraud-risk assessments downward when they encounter additional counterdiagnostic evidence. The analyses were two-sided, one-sample location tests under the null hypothesis that the nondiagnostic evidence did not affect the participants' fraud-risk assessment.

Summary statistics and the results of the individual location tests are presented in Table 4. Two sets of responses could be analyzed: (1) responses to the neutral version and (2) responses to all counterdiagnostic versions. Sample sizes were too small to perform separate analyses of the favorable and the unfavorable versions. The nondiagnostic evidence had a pronounced effect on the participants' judgments. The average mean shift was -2.43. The mean shift for the neutral treatment group was -3.02. The value of the dependent variable was positive for two of the nine participants (mean of .368). All p-values were less than .04 using the student's t procedure; the signed rank statistic for the neutral treatment group (a sample size of five) was not significant. This is evidence that the nondiagnostic evidence had a moderating effect on the degree to which counterdiagnostic evidence caused a downward revision in the participants' fraud-risk assessments -- a dilution effect.

With a dilution effect, the auditor's absolute level of sensitivity to the risk of financial-statement fraud is more than it would be in the absence of nondiagnostic evidence. When an auditor's sensitivity to such fraud is significantly more than it should be, he/she may select
audit procedures that give undue consideration to the risk of such fraud. When this happens, there is an increased probability that the audit will be inefficient.

5.43 Conclusion for Counterdiagnostic Evidence

The statistical analyses reveal that nondiagnostic evidence had a pronounced moderating effect on the degree to which counterdiagnostic evidence caused downward revisions in the participants' fraud-risk assessments (cf. Nisbett, et al. [1981]). This result is consistent with the similarity interpretation developed in Chapter III of how auditors judge a company's exposure to fraudulent financial reporting, when class membership is judged by similarity to the prototypical class member.

This moderating effect is described by the term \( \Gamma_3 f(T_{13^-}) \) in equation (5b), which is negative if and only if (1) the nondiagnostic evidence is salient, i.e., \( f(T_{13^-}) \) is positive and (2) the decision weight \( \Gamma_3 = \alpha(\Phi_1 + \Phi_2) \) is negative. The salience of the nondiagnostic evidence and the implications of the sign of the decision weight \( \Gamma_3 \) were discussed in connection with the results of the diagnostic manipulation in Section 5.34.

If the participants in both the counterdiagnostic and the diagnostic treatment groups had focused on the same outcome, an enhancement effect would have been observed in one group and a dilution effect in the other. This was not the case; a dilution effect was observed in both conditions. This result suggests that (1) auditors focus on the outcome that is consistent with the informative evidence
and (2) this outcome controls auditors' behavior. Therefore, the distinction between equation (2a) and equation (2b) is real.

As in the diagnostic treatment groups, the participants used the nondiagnostic evidence far more than they realized. Here, eight of the nine participants whose responses were included in the analysis stated that only the counterdiagnostic theme affected their context-dependent ratings.
Table 4

Location Tests
Counterdiagnostic Versions

| Version        | # of Obs. | Mean | Std. Dev. | |T| \(^1\) | Skewness | W^2 \(^2\) | Median | |S| \(^3\) |
|----------------|-----------|------|-----------|--------|-----------|----------|--------|-------|-------|
| favorable      | 1         | .50  | ...       | ...    | ...       | ...      | ...    | ...   | ...   |
| neutral        | 5         | -3.02| 2.47      | .0346  | -.2351    | .5334    | -3.00  | .1250 |
| unfavorable     | 3         | -2.42| ...       | ...    | ...       | ...      | ...    | ...   | ...   |
| combined       | 9         | -2.43| 2.49      | .0019  | .4971     | .0391    | -2.00  | .0391 |

\(^1\) Student's t statistic; p-value under the null that the mean is zero
\(^2\) Shapiro-Wilk statistic; probability of obtaining a smaller value of the test statistic under the null that values of the dependent variable are a random sample from a normal distribution
\(^3\) Signed Rank statistic; p-value under the null that the mean is zero
5.5 Relative Importance of Individual Informative Items

Kendal's coefficient of concordance \( W \) expresses the over-all agreement among rankings of independent sets of objects. In the current study, \( W \) is used to measure the extent of agreement among the context-free ratings (transformed to ranks) for all participants who completed the respective card-sorting/rating tasks. The value of \( W \) for each set of ratings can be computed from the data presented in Table 5 (see Siegel [1956, pp. 229-238]).

Under the null hypothesis that the sets of rankings are independent, \( W \) is 0.336 \( (X^2(4) = 131.9) \) for the diagnostic items and 0.576 \( (X^2(4) = 221.26) \) for the counterdiagnostic items. The probability under the null of a value as large as the observed \( W \) is less than 0.0001 in each case. This is evidence that the participants were applying essentially the same standard in ranking the five diagnostic items and in ranking the five counterdiagnostic items.

The following two observations from the activities described in Section 4.41 of Chapter IV are relevant here: (1) the degree of consensus about relative importance was much higher for counterdiagnostic evidence than for diagnostic evidence and (2) there was substantial disagreement about whether incentives or opportunities play the larger role in increasing a company's exposure to fraudulent financial reporting. If one accepts the criterion that the magnitude of \( W \) (not simply its statistical significance) is a measure of the extent of agreement, there was a higher degree of consensus about the rank order of the counterdiagnostic items than there was about the rank order of the diagnostic items. If one accepts the criterion that when \( W \) is
significant the sum of the ranks for each item is an indication of its relative importance, there is no evidence to suggest that incentives are either more or less important than opportunities.

For audits of financial statements for periods beginning on or after January 1, 1989, auditors must explicitly assess the risk that the financial statements they are auditing contain material misrepresentations [AICPA, 1988a]. The results of this analysis suggest that a bootstrapping technique would yield relatively stable estimates of weights to attach to various indicators of fraud if firms choose to use an improper linear model as a decision-aid. (Dawes [1971] and Ashton [1975] are examples of the technique.)
Table 5
Concordance Tests

<table>
<thead>
<tr>
<th>Diagnostic Items</th>
<th>Counterdiagnostic Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Sets of Ranks</td>
<td>98</td>
</tr>
<tr>
<td>Sums of Ranks:</td>
<td></td>
</tr>
<tr>
<td>Card A</td>
<td>257.5</td>
</tr>
<tr>
<td>Card B</td>
<td>349.5</td>
</tr>
<tr>
<td>Card C</td>
<td>222.5</td>
</tr>
<tr>
<td>Card D</td>
<td>218.5</td>
</tr>
<tr>
<td>Card E</td>
<td>422</td>
</tr>
<tr>
<td>Correction for Ties</td>
<td>9</td>
</tr>
</tbody>
</table>
6.1 Conclusion

This study found, as prior studies did, that seemingly irrelevant evidence can affect human judgment. The current study provided an initial demonstration of the role seemingly irrelevant evidence plays when auditors assess the risk that a client's financial statements are fraudulent. In all cases, the nondiagnostic evidence tempered the influence of evidence auditors consider useful for the fraud-risk assessment -- a dilution effect. Not only did the participants use the nondiagnostic evidence far more than they realized, but many reported that such evidence reinforced the implications of the informative evidence when, in fact, a dilution effect was observed. The practical implication of these results is that the selection of audit procedures may be influenced by audit evidence not thought to sway such decisions.

A word of caution is in order. Studies of audit judgment have been popular because they may have policy implications for such professional practice areas as development and modification of auditing methods and standards, approaches to training and supervision, and creation of decision aids [Boritz, 1986, p. 335]. Practitioners have expressed concern that biases (as opposed to heuristics) that auditor subjects have demonstrated in many of these studies may adversely affect the
quality of real-world decisions. For example, in Peat, Marwick, Mitchell and Company's manager-training program, managers are sensitized to potential biases [Libby, 1981, p. 68].

Funder [1987] suggests that the implications of such studies may be illusory because an "error" is not the same as a "mistake". An error is a judgment, based on experimental materials, that departs from a model of the judgment process. A mistake is an incorrect real-world judgment. While errors are an important and valuable tool for studying the process of human judgment, they do not necessarily imply corresponding inaccuracies in real-world judgments. Errors in a laboratory setting may not be mistakes in a broader, more realistic frame of reference; the processes that produce such errors may lead to correct decisions in the real world [Hogarth, 1981; Ebbesen and Konecni, 1980; Marcus and Zajonc, 1985, pp. 196-7].

Thus, the results this study cannot be construed as yet another example of how cognitively inept individuals seem to be. Rather, this study highlights an aspect of auditor judgment not previously identified. A definitive assertion is not made about the normative or counternormative nature of how auditors weight and mentally integrate evidence. Some might say that evidence that is used is, by definition, relevant. Determining whether the behavior evidenced in this study implies costly mistakes in real-world judgments will require converging lines of evidence that further research can provide.
6.2 Alternative Interpretations

The current results are consistent with the similarity interpretation of the dilution effect. While there is no direct evidence that similarity judgments produce that effect, a number of alternative explanations have been ruled out. Three additional explanations are worth considering.

First, the participants may have been so overwhelmed by the nondiagnostic evidence that they did not remember the informative evidence. Of the forty-eight participants who provided usable responses, forty-seven (98%) explicitly mentioned the informative theme when explaining their context-dependent rating responses. Hence, this explanation is not credible.

Second, after recognizing an item of informative evidence, participants may have developed expectations about the experimental company's exposure to fraudulent financial reporting. Instead of seeming supportive of the informative evidence (the confirmation effect discussed in Section 5.32 of Chapter V), the nondiagnostic evidence may have failed to confirm the participants' expectations about the experimental company's exposure to such fraud, making the participant less certain about the implications of the informative evidence. Three of the fifty-three participants (5.7%) who received a diagnostic version of the instrument indicated they used this strategy (one was included in the statistical analyses -- the others did not revise their fraud-risk assessments). Hence, this explanation is at least partially valid for this study.
A third alternative suggests that "... new judgments are a weighted average of previous judgments and the scale value of new evidence [Shanteau, 1975, p. 84]." This serial adjustment strategy has been used to explain phenomena conceptually similar to the dilution effect when the task required subjects to make a series of ratings, each dependent on the previous rating. This strategy will produce a dilution effect across a series of ratings if individuals ignore the weights, which presumably are zero, that they attached to nondiagnostic evidence. The model developed in Chapter III uses this basic strategy, but at the level of an individual judgment instead of across a series of judgments. Here, similarity judgments produce the scale values that are averaged by an additive combining rule. The decision-maker is less aware of role played by nondiagnostic evidence in the framework developed in this study than in the framework described by Shanteau.

6.3 Recommendations

Two general recommendations are proposed. The first concerns the interpretation of judgment research. As Gibbins [1984] and Waller and Felix [1984] point out, auditors perform a series of subtasks in forming an audit opinion. Frederick and Libby [1986, p.289] note that efforts to understand the audit process require disaggregation and detailed analyses of these subtasks. The results of this study suggest that care should be taken in interpreting judgment research when the experimental design ignores the influence of evidence that is available when a decision is made, even if it is typically considered to be of little or no value for the decision. For example, auditors may be sufficiently
regressive when they have a richer frame of reference or a more realistic conception of the client than those provided in the Joyce and Biddle [1981b] experimental materials.

The second recommendation concerns the need for more research on the role that seemingly irrelevant evidence plays in decision-making and on what to do about its affect on real-world decisions [Shanteau and Gaeth, 1983, p. iii]. There are many unanswered questions: (1) What role does seemingly irrelevant evidence play in predecisional processes, such as generating alternative hypotheses? (2) Are there individual differences in susceptibility to seemingly irrelevant evidence? (3) What influence do environmental conditions, such as time pressure, have on the use of seemingly irrelevant evidence? (4) Is the similarity interpretation convincing? (5) Is the similarity-based process dysfunctional, that is, a heuristic or a bias [Zukier, 1981]? (6) Can (should) training procedures to mitigate the influence of seemingly irrelevant information be developed (or improved)?, (7) Can the influence of training be modeled at a psychological processing level? (8) What is the relative influence of seemingly irrelevant evidence across tasks? (9) How much nondiagnostic evidence does it take to dilute some specified unit of informative evidence? (10) Since expertise has been found to mitigate the dysfunctional aspects of heuristics in a number of studies [Joyce and Biddle, 1981a, 1981b; Kinney and Uecker, 1982; Libby, Artman, and Willingham, 1985; Anderson and Wright, 1988], will domain-specific experience diminish or even eliminate the dilution effect? The model developed in Chapter III predicts that domain-specific experience will lessen the extent to which
nondiagnostic evidence influences the fraud-risk assessment. (11) What is the relationship between the influence of seemingly irrelevant evidence on an auditor's fraud-risk assessment and its influence on the selection of audit procedures or on the subsequent interpretation of evidence? (12) What is the net effect of seemingly irrelevant evidence when both diagnostic and counterdiagnostic evidence are available?
LIST OF REFERENCES


Gati, I. and Tversky, A., "Representations of Qualitative and


Nagy, G. F., "How are Personnel Selection Decisions Made?: An Analysis


Snyder, M. and Uranowitz, S., "Reconstructing the Past: Some Cognitive


APPENDIX

Diagnostic-Unfavorable Version of the Experimental Materials

In order to comply with the Graduate School's guidelines for preparing dissertations, the experimental materials were adapted as follows: (1) each 8 1/2 x 11 inch page of the instrument was reduced to 90% of its original size to adhere to the 6 x 9 inch area restriction, (2) page numbers were deleted, (3) the 'Questions about Mr. Chip Company' and 'General Questions' folders, which were presented in a two-sided format, were reproduced in a one-sided format, (4) pages labeled "page intentionally left blank" are excluded, and (5) the individual 3 1/4 x 4 3/4 inch white stock cards are reproduced two per page with rectangles to represent the edges of the cards. The text of the instrument was not modified to accommodate these format changes.
GENERAL INSTRUCTIONS

August 25, 1988

Materials:

There are four folders and two sets of cards inside this envelope. Take a moment to make sure you have each item, but please do not open these items until told to do so.

One folder is labeled 'History and Background of Mr. Chip Company'. Another folder is labeled 'Excerpts From the 1986 Mr. Chip Company Audit Workpapers'. In this folder you will find several memos and permanent file schedules from the 1986 audit workpaper file of Mr. Chip Company. Another folder is labeled 'Questions About Mr. Chip Company'. In this folder you will find a series of questions relating to issues raised in the Mr. Chip Company workpapers. The other folder is labeled 'General Questions'. You will use the two sets of cards to answer some of the general questions.

Tasks:

There are several parts to this exercise. The materials and specific instructions on how to complete each task are included in the separate folders described above.

Please Remember:

If you have a question at any time during this exercise, please ask. You may write on any portion of the materials. There are no correct answers. I am interested in your judgment.

Please stay seated until everyone has had an opportunity to finish the exercise. It should take approximately one hour.

To Begin:

Do not begin this exercise until I have had an opportunity to make a few introductory remarks.

Thank you for your cooperation.

Karl Hackenbrack
HISTORY AND BACKGROUND
OF MR. CHIP COMPANY

please do not open this folder until told to do so
Frame of Reference:

Imagine you are the audit manager assigned to the 1986 audit of Mr. Chip Company. This is the first time you have been involved with this client. The audit manager that had been assigned to the previous audits of Mr. Chip Company is no longer with your firm.

The narrative on the following page was prepared by Doug Blouch, the in-charge senior. This is the third year Doug has been assigned to the Mr. Chip Company audit.

Task:

Please read the narrative on the following page. The purpose of this narrative is to acquaint you, in a preliminary way, with Mr. Chip Company. As you read the narrative, form general impressions on various matters an auditor considers as he/she begins the process of becoming acquainted with a new client. Among these matters would be: the client's size and the industry it is in; the client's history and where it is likely to go; who are the users of the client's financial statements; the nature of any special accounting or auditing problems likely to be encountered; the risk that the client's financial statements, as presented for audit, will not be in conformity with generally accepted accounting principles.

Your impressions on these matters will be based partly on the accompanying narrative and partly on your previous auditing experience. At this early stage in an audit, your impressions would be highly tentative. Nevertheless, they will be important in this experiment because they will serve as a baseline from which to evaluate the audit workpapers you will read during the next phase of this exercise.
Mr. Chip Company, a Delaware Corporation whose principal office is in Harrisburg, Pennsylvania, is a manufacturer and distributor of several lines of snack foods. Its potato chip sales are strongest in New York, New Jersey, Florida, the Middle Atlantic States, and the Southeast. The company manufactures several varieties of potato chips, cheese curls, and pretzels which account for approximately seventy percent of its net sales. The company also purchases other snack food products from various producers. These products are repackaged under the Mr. Chip label and distributed with Mr. Chip Company's manufactured products. The company also licenses Brotherly restaurants and baked goods, and subleases about a dozen valuable properties in New York City.

Mr. Chip Company distributes its snack food products through its own sales organization and through brokers to service stations, drug stores, wholesale grocers and supermarket chains, small food retailers, and national retail merchandising chains. National grocery store chains account for 5% of snack food sales; other grocery stores 55%; food service outlets 40%.

Mr. Chip Company operates three manufacturing facilities and thirty-six distribution centers. The average plant age is six years. There are approximately 500 employees.

The company was organized in Harrisburg, Pennsylvania on May 14, 1947 as a sole proprietorship by Herbert Dunley, the company's current president and C.E.O. In 1953, Herbert opened his first modern production facility on the site of today's corporate offices and primary production facility. This expansion was made possible by debt financing through regional banks. In order to obtain financing to expand the plant and operations, Herb incorporated the business in 1959.

The corporation is authorized to issue 100,000,000 shares of $2 par value common stock. Of the shares authorized, Herb received 500,000 in exchange for his interest in the sole proprietorship. Another 300,000 shares were issued in a private offering. In May 1968, a public offering of 4,000,000 shares was made. To date, this has been the only public offering. The stock is occasionally traded on the over-the-counter market and is registered with the Securities and Exchange Commission pursuant to Section 12(b) of the Securities Exchange Act of 1934. Today, thirty percent of the company's capital structure is long-term debt.

This is the third year your firm has been engaged by Mr. Chip Company to provide a standard audit. Mr. Chip Company's overall internal control environment has been evaluated as effective; management has established an appropriate level of control consciousness among the officers and employees and has instituted a number of worthwhile general control mechanisms. There is an internal audit department that reports to the audit committee of the Board of Directors. Few audit adjustments have been necessary in prior engagements. The fiscal year for both accounting and tax purposes is December 31.
Now that you have finished reading the narrative, open only the folder labeled 'Excerpts From the 1986 Mr. Chip Company Workpapers' and follow the instructions on the second page of that folder. You may refer to this history and background document as often as you like while reading the audit workpaper excerpts.
please do not open this folder until you have read the narrative in the folder labeled 'History and Background of Mr. Chip Company'.
In this folder, you will find several memos taken from the 1986 audit workpaper files of Mr. Chip Company. Each memo was either written by or updated by the in-charge senior, Doug Blouch, during the pre-field work phase of the audit.

These memos generally relate to the sales/collections cycle. For the purpose of this exercise, many different sets of workpapers could have been used; the workpapers are simply a convenient format to present information about Mr. Chip Company. I chose to emphasize the credit sales/accounts receivable cycle.

Please use the next fifteen minutes to thoroughly acquaint yourself with the materials in this folder. You should read these memos with an eye toward revising/enriching the general impressions of Mr. Chip Company you formed while reading the history and background document. Focus on what the memos tell you about Mr. Chip Company. Do not focus on the audit process.

When you feel reasonably comfortable with your mastery of the materials in this folder, you will answer specific questions about how the information in these memos influenced the preliminary impressions you formed while reading the history and background document. More detailed instructions on what to do after you have read the materials in this folder are on the last page (page 9) of this folder.
Market Perspective:

Retail sales of snack foods are expected to reach $24 billion in 1986. Salty snacks, one of the largest segments of the snack food market, will account for more than $6 billion in sales. In the last five years, salty snack sales have increased at an average annual rate of nine percent.

Results of Operations—1986 compared to 1985:

Sales were up at an annual rate of nine percent over 1985. A few regional competitors are continuing to vie for shelf space and promote deeply discounted products. Mr. Chip Company has refrained from entering the price-slashing game. Instead, it is spending heavily on television and radio advertising and is targeting these areas with intensive, innovative promotions. Because Mr. Chip Company has a large distribution area, a modest crimp in the margins at a few locations is unlikely to significantly affect the overall profit picture.

Part of the 1986 sales growth is attributable to the introduction of two new snack food products during 1985. Consumer response to one of these new products (chocolate chip cookies) did not meet the company's expectations. The item is outside Mr. Chip Company's traditional salty snack category. The relatively modest sales increase this new product generated simply could not offset the significantly higher promotional spending necessary to support it. The company has withdrawn the underperforming product, eliminating the incremental promotional expense.

Cost of goods sold as a percentage of sales is up slightly. Increases in operating efficiency have not quite offset higher raw material costs, particularly potatoes and packaging materials. The company expects the price of potatoes to retreat in 1987, but they doubt potato prices will return to prior levels.

(continued on the next page)
Financial Position:

Working capital has increased slightly. The established line of credit was used to meet seasonal working capital requirements, but was unused at September 30, 1986. There are no immediate plans to use it.

During 1986, over $6 million will be spent on capital improvements including approximately $500 thousand on new vending machines, $2.3 million on a new pretzel line and replacement of a potato chip line, and $1 million on expansion and renovation of facilities. The continuing addition of sales routes will boost volume, but will also put pressure on margins. Capital expenditures are currently projected to be slightly less in 1987 than in 1986.

An acquisition might seem in order based on Mr. Chip Company's strategy for growth based largely on its ability to acquire regional brands with a market niche. Duplicating the company's historical return on total capital is unlikely as the current going rate for acquisition candidates in this industry is relatively high. Thus, Mr. Chip Company's efforts to expand through strategic acquisitions may be frustrated until prices retreat somewhat.
MR. CHIP COMPANY
MENOS ON PRELIMINARY REVIEW OF
SALES/COLLECTIONS CYCLE
for the year ending December 31, 1986
prepared by Doug Blouch (in-charge senior)

As part of the preliminary study and evaluation of the internal accounting controls in the credit sales/accounts receivable cycle, I (a) made inquiries of client personnel, (b) observed the processing methods and procedures used, (c) reviewed client manuals and other written instructions, and (d) performed walkthroughs of selected transactions. The purpose of these steps was to update/confirm (1) our understanding of the flow of the significant transactions and (2) the basic structure of the accounting controls in the credit sales/accounts receivable cycle. No updates of our permanent file documentation of this cycle are necessary; the relevant systems are unchanged from the prior audit period. The following items were noted:

- Interest is not typically charged on accounts receivable unless the credit manager deems it helpful in motivating payment. Even though the amount is immaterial, it is booked to accounts receivable when billed. It is rarely collected, resulting in write-offs in later months. A management letter comment has been proposed to record this immaterial amount of interest on the cash, rather than the accrual, basis to avoid unnecessary journal entries.

- Sales representatives have the authority to grant their customers price discounts. These discounts are not well documented; sometimes they are simply verbally agreed on with no written documentation. If the customer's recollection of the terms does not match that of the sales representative, the customer is given the benefit of the doubt. The amount of revenue lost is not determinable at this time, but is unlikely to be significant. A management letter comment has been proposed to require discount terms to be in writing and approved by a company official.

- The monthly printout of the accounts receivable subsidiary ledger shows, for each customer that has a month-end balance, the month's beginning and ending balances and each transaction during the month. This printout is sometimes awkward to use: in order to trace a transaction of a previous month, employees must refer to that month's printout. This is not a weakness in internal control, but it does make resolving questions more difficult than need be. The issue here is not effectiveness, but efficiency. A management letter comment suggesting that the monthly printouts show open items rather than beginning balances has been proposed. This would not be difficult to change. Almost all customers almost always pay specific invoices, and the Data-entry Department already does the necessary coding of cash receipts so that the weekly printout of delinquent accounts, run for the Credit Department, can show open items.
Effective the first day of the current audit period, the management compensation package was changed to reflect the company's increased emphasis on achieving budgeted targets. Basically, the mix of salary and bonus has been changed to place a much greater emphasis on bonuses tied to responsibility-center financial performance.

By instituting this new compensation program, management may be motivating personnel to meet legitimate economic goals. The new program may also have created an atmosphere conducive to undisciplined or unethical behavior, increasing the company's exposure to fraudulent financial reporting.

(documentation of the specific changes to the management compensation program is not included in the case materials)
Except for the following matters, the controls in the accounting applications that affect accounts receivable have been preliminarily evaluated as adequate and will be tested to establish a basis for reliance.

. Controls to ensure that totals in the sales journal are correctly posted to the general ledger are not adequate. The reconciliation of the accounts receivables subsidiary ledger to the general ledger is performed by a person who posts sales to the general ledger. The potential amount of error that could result is not determinable at this time. We will trace each month's postings from the relevant source journals to the general ledger. This test will enable us to determine if errors occurred in this area. In prior audits, no adjustments have been made as a result of lack of control mechanisms in this area.

. Controls to ensure that sales and cash receipts are recorded in the proper period are not adequate. The amount of error that would result will not involve amounts greater than one week's transactions. Extensive sales and cash receipts cutoff procedures will be performed at both 11/30/86 (date of accounts receivable confirmation and observation of inventory) and 12/31/86 (balance sheet date). These tests will enable us to determine whether errors have occurred in this area. In prior audits, no adjustments have been made as a result of lack of control mechanisms in this area.
MR. CHIP COMPANY
ACCOUNTS RECEIVABLE AND RELATED ALLOWANCES
SELECTED ACCOUNT CHARACTERISTICS
for the year ending December 31, 1986

prepared by Doug Blouch (in-charge senior)

<table>
<thead>
<tr>
<th>Date</th>
<th>Total</th>
<th>Allowance</th>
<th>Sales</th>
<th>W/O</th>
</tr>
</thead>
<tbody>
<tr>
<td>12/31/84</td>
<td>$3050</td>
<td>$150</td>
<td>$55,000</td>
<td>$420</td>
</tr>
<tr>
<td>12/31/85</td>
<td>$3460</td>
<td>$170</td>
<td>$60,000</td>
<td>$480</td>
</tr>
<tr>
<td>08/30/86</td>
<td>$4160</td>
<td>$210</td>
<td>$43,600</td>
<td>$390</td>
</tr>
<tr>
<td>12/31/86 (budget)</td>
<td>$4190</td>
<td>$220</td>
<td>$65,900</td>
<td>$590</td>
</tr>
</tbody>
</table>

(all dollar amounts in thousands)

- 30,000 active customer accounts; primarily in the $750—$1500 range; the 40 largest accounts typically amount to 10% of the outstanding balance.

- The cash receipts and sales/accounts receivable applications are the primary data applications affecting the accounts receivable general ledger account.

- The only accounting judgment affecting the account is the estimate of the allowance for doubtful accounts. The estimate is made jointly by the assistant controller and the credit manager and is reviewed by the controller. An audit adjustment to increase the allowance account was made in 1984.

- Historically, the percentage of customer accounts over ninety days old is between 4.5% to 5%. This relationship is expected to continue.

- The account has historically been between 15% and 17% of total assets. This relationship is expected to continue.

- As a result of management’s relaxation of credit policies (a strategic move aimed at market penetration), the average number of days’ sales outstanding has risen from 20 to 22 days. Write-offs of accounts as uncollectible have increased slightly as a percentage of sales.

- During the second quarter of the 1986, one of Mr. Chip Company’s larger customers signed a contract with a competitor of Mr. Chip Company. This customer accounted for only one-half of one percent of sales. This was an isolated instance.
When you feel reasonably comfortable with your mastery of the materials in this folder, you should (1) open only the folder labeled 'Questions About Mr. Chip Company' and (2) follow the instructions on the third page of that folder. You may refer to the materials in this folder and the history and background narrative as often as you like while answering the questions about Mr. Chip Company.
QUESTIONS ABOUT
MR. CHIP COMPANY

please do not open this folder until you have read the materials
in the folder labeled 'Excerpts From the 1986 Mr. Chip Company Workpapers'
QUESTIONS ABOUT
MR. CHIP COMPANY

SPECIFIC INSTRUCTIONS

Consider what you have learned about Mr. Chip Company. First you read a brief description of the company's activities through 1985 titled the 'History and Background of Mr. Chip Company.' While reading that narrative, you formed a general impression of Mr. Chip Company. This tentative general impression served as a baseline from which to evaluate the excerpts from the 1986 workpapers of Mr. Chip Company.

That first impression of the company may have changed as you learned more about the company while reading excerpts from the 1986 audit workpapers. I would like to learn how your baseline impression of the company changed as a result of reading the materials in the folder labeled 'Excerpts From the 1986 Mr. Chip Company Audit Workpapers'.

A common 'complaint' among professionals that participate in this type of study is the lack of realism; particularly, the lack of information that they would like to have available and that probably would be available in a typical engagement. Since time is a limiting factor, I cannot give you everything you would like to have. Please do the best you can using only the information made available.

Specific instructions are provided with each question. Answer the questions in the order given. Once you feel comfortable with your response to a question, the instructions will tell you which other question to turn to; you will not answer every question.
FIRST QUESTION

Some participants in this study received excerpts from the 1986 audit workpaper file of Mr. Chip Company intended to suggest that the company's exposure to fraudulent financial reporting increased during 1986. Fraudulent financial reporting is material and intentional misstatements in or omissions from a set of financial statements. Other participants received excerpts from the 1986 audit workpaper file intended to suggest that Mr. Chip Company's exposure to fraudulent financial reporting decreased during 1986. Still others received excerpts from the audit workpapers describing only activities that had little or no influence on Mr. Chip Company's exposure to fraudulent financial reporting during 1986.

You were randomly assigned to one of these three groups. The purpose of this question is to determine which version of the case study you have and, on that basis, tell you which other questions to answer.

Please answer the following question(s). To record your answer(s), place a check mark on the line immediately after the response that best describes your opinion. The instruction in parentheses immediately after your answer tells you which question you should answer next.

A. Did what you learned about Mr. Chip Company while reading your version of the 1986 audit workpaper excerpts cause you to change your assessment of the risk that Mr. Chip Company's 1986 financial statements are affected by fraudulent financial reporting?

   no _____ (turn to question six on page 15 of this folder; do not answer question B on this page)

   yes _____ (go to question B on this page)

B. If your answer to question A is 'yes', did your assessment of the risk that Mr. Chip Company's 1986 financial statements are affected by fraudulent financial reporting:

   increase ____ (turn to question 2 on page 6 of this folder)

   decrease ____ (turn to question 4 on page 10 of this folder)
SECOND QUESTION

You should answer this question only if you indicated in the first question of this folder that your version of the audit workpapers caused you to increase your assessment of the risk that Mr. Chip Company's 1986 financial statements are affected by fraudulent financial reporting (hereafter referred to as your risk assessment).

The purpose of this question is to determine how much your risk assessment increased as a result of reading the materials in the folder labeled 'Excerpts From The 1986 Mr. Chip Company Audit Workpapers'. Currently, the amount by which your risk assessment increased is a feeling, not some hard and fast number. The following instructions will explain how to use the scale on the opposite page to express this feeling as a numerical value (hereafter referred to as a rating).

The endpoint zero (0) and the value ten (10) of this scale are described on the opposite page immediately below the scale. The reason for labeling two points on this scale is to give you a reference from which to make your rating. Take a moment to familiarize yourself with the description of the scale values zero (0) and ten (10) before continuing to read these instructions.

As indicated by the scale, use a low number if everything you learned while reading the 1986 Mr. Chip Company audit workpaper excerpts caused your risk assessment to increase little relative to the increase in risk represented by the description of the scale value ten (10). Use an intermediate value if the audit workpapers caused your risk assessment to increase as much as the increase in risk represented by the description of the scale value ten (10). Use a high value if the audit workpapers caused your risk assessment to increase much more than the increase in risk represented by the description of the scale value ten (10). If you want to use a number greater than twenty, simply write that value in the right margin next to the scale.

To record your judgment, circle the number on the scale that best reflects your opinion. If you change your mind, erase your old response completely and make a new selection. Please reread the question or ask me if you are unsure of what you are being asked to do.
SECOND QUESTION
(continued)

Ten (10) represents the amount your risk assessment would increase in the following situation:

Consider your initial assessment of the risk that Mr. Chip Company's 1986 financial statements are affected by fraudulent financial reporting, i.e., your baseline risk assessment after reading only the history and background document. Now, imagine that another company has that same initial level of risk.

Your initial assessment of the risk of such fraud at this other company increases by ten (10) units when you learn that (1) its management is predicting a narrow margin of compliance with certain restrictive debt covenants for 1986, (2) even though a waiver has always been obtained, the lender has threatened to not be so lenient in 1986, (3) if a waiver is needed and is not granted, a significant amount of long-term debt will become current, and (4) management is taking the lender's threat seriously.

Zero (0) means that the excerpts from the 1986 audit workpaper file of Mr. Chip Company, taken in their entirety, did not cause you to change your risk assessment.

Once you are satisfied with your response to this question, turn to question 3 on page 9 of this folder and follow the directions.
THIRD QUESTION

You should answer this question only if you have completed your answer to the second question.

Please explain, in the space below, why you chose the response you did in the second question. Be as specific as possible.

When you are satisfied with your answer to this question, please lay aside all materials you have used, open only the folder labeled 'General Questions', and follow the specific instructions on the third page of that folder.
FOURTH QUESTION

You should answer this question only if you indicated in the first question of this folder that your version of the audit workpapers cause you to decrease your assessment of the risk that Mr. Chip Company's 1986 financial statements are affected by fraudulent financial reporting (hereafter referred to as your risk assessment).

The purpose of this question is to determine how much your risk assessment decreased as a result of reading the materials in the folder labeled 'Excerpts From The 1986 Mr. Chip Company Audit Workpapers'. Currently, the amount by which your risk assessment decreased is a feeling, not some hard and fast number. The following instructions will explain how to use the scale on the opposite page to express this feeling as a numerical value (hereafter referred to as a rating).

The endpoint zero (0) and the value ten (10) of this scale are described on the opposite page immediately below the scale. The reason for labeling two points on this scale is to give you a reference from which to make your rating. Take a moment to familiarize yourself with the description of the scale values zero (0) and ten (10) before continuing to read these instructions.

As indicated by the scale, use a low number if everything you learned while reading the 1986 Mr. Chip Company audit workpaper excerpts caused your risk assessment to decrease little relative to the decrease in risk represented by the description of the scale value ten (10). Use an intermediate value if the audit workpapers caused your risk assessment to decrease as much as the decrease in risk represented by the description of the scale value ten (10). Use a high value if the audit workpapers caused your risk assessment to decrease much more than the decrease in risk represented by the description of the scale value ten (10). If you want to use a number greater than twenty, simply write that value in the right margin next to the scale.

To record your judgment, simply circle the number on the scale that best reflects your opinion. If you change your mind, erase your old response completely and make a new selection. Please reread the question or ask me if you are unsure of what you are being asked to do.
FOURTH QUESTION
(continued)

Ten (10) represents the amount your risk assessment would decrease in the following situation:

Consider your initial assessment of the risk that Mr. Chip Company's 1986 financial statements are affected by fraudulent financial reporting, i.e., your baseline risk assessment after reading only the history and background document. Now, imagine that another company has that same initial level of risk.

Your initial assessment of the risk of such fraud at this other company decreases by ten (10) units when you learn that early in the current audit period, this company (1) replaced its chief internal auditor with a better-trained, more aggressive individual with five years' experience on the audit staff of a Big-8 accounting firm and six years' experience in the internal audit department of a fortune 500 company and (2) approved resources to hire additional internal audit personnel. By the beginning of the second quarter, all additional personnel were hired. The effectiveness of the internal audit function has significantly improved.

Zero (0) means that the excerpts from the 1986 audit workpaper file of Mr. Chip Company, taken in their entirety, did not cause you to change your risk assessment.

Once you are satisfied with your response to this question, turn to question 5 on page 13 of this folder and follow the directions.
FIFTH QUESTION

You should answer this question only if you have completed your answer to the fourth question.

Please explain, in the space below, why you chose the response you did in the fourth question. Be as specific as possible.

When you are satisfied with your answer to this question, please lay aside all materials you have used, open only the folder labeled 'General Questions', and follow the specific instructions on the third page of that folder.
SIXTH QUESTION

You should answer this question only if you indicated in the first question of this folder that your version of the audit workpapers did not cause you to change your assessment of the risk that Mr. Chip Company's 1986 financial statements are affected by fraudulent financial reporting.

Please explain, in the space below, how the materials in the folder labeled 'Excerpts From the 1986 Mr. Chip Company Audit Workpapers' influenced your impression of Mr. Chip Company? What type of picture did I paint? How would you describe this company? I have intentionally made this an open-ended question.

When you are satisfied with your answer to this question, please lay aside all materials you have used, open only the folder labeled 'General Questions', and follow the specific instructions on the third page of that folder.
GENERAL QUESTIONS

Please do not open this folder until you have completed the questions you were asked to answer in the folder labeled 'Questions About Mr. Chip Company'.
GENERAL QUESTIONS

SPECIFIC INSTRUCTIONS

Specific instructions are provided with each question. Answer the questions in the order given. You will answer every question in this folder.

Do not refer to previous materials, questions, or answers as a guide to completing questions in this folder. I would like you to answer these questions based on your general knowledge of factors that influence a typical client's exposure to fraudulent financial reporting. Your answers should not be based on a particular client.
In your envelope you will find a set of cards labeled 'Question One'. On each card you will find a brief description of a situation that other auditors have classified as either an incentive or an opportunity to commit fraudulent financial reporting. Take a moment to familiarize yourself with these five cards. Depending on which version of the Mr. Chip Company workpapers you read, you may recognize the situation described on two of the five cards.

Your task is rate the degree to which the situation represented on each card would cause you to increase your assessment of the risk that a typical company's financial statements are affected by fraudulent financial reporting. The situation presented on each card should be evaluated separately from the situations presented on all other cards. All ratings will be made on the one-hundred point scale on the opposite page. The following instructions explain how to use this scale.

**Step 1:** Sort the cards from high to low.
Carefully read each card. Then, order the cards in relation to the degree to which the situation posed on each card would increase your assessment of the risk that a company's financial statements are affected by fraudulent financial reporting. This ordering should be based on a general case rather than a specific client. When you have finished sorting the cards, the card on the top of the stack should describe the situation that most increases your risk assessment. The next most influential card should be placed second in the stack, the next most third, and so on until the card with the description of the situation that increases your risk assessment less than the situations on all the other cards is on the bottom of the stack. You may rearrange the cards as often as you like. Feel free to change the position of the cards in the stack at any time during this task.

**Step 2:** Assign the top card the value one-hundred.
Assign the scale value one-hundred (100) to the top card on your stack by placing the letter in the upper right corner of the card next to the value one-hundred (100) on the scale.

**Step 3:** Rate the remaining cards.
The reason for labeling the two endpoints of the scale is to give you a reference from which to rate the situations represented on each of the other cards. Use zero (0) if a situation would not increase your risk assessment. Use one hundred (100) if a situation would increase your risk assessment as much as the situation represented on the top card of your stack. Use an intermediate value if a situation would increase your risk assessment but not as much as the situation represented on the top card of your stack.

Assign a scale value to each card by placing the letter in the upper right corner of the card next to the value on the scale that best reflects its relative influence on your risk assessment. Please clearly indicate which value you assign to each card.
QUESTION ONE
(continued)

This entire task should be a flexible give-and-take process until you are satisfied with the relative positions of all the cards on the scale. Do not be afraid to change your answers several times before moving on to the next question. If you do change your mind, please erase your initial rating(s) completely.

Please use this scale to record your answers to question one:

```
--- 100 (___ letter of top card)
  95 |--- 90
   |----- 80
  75 |----- 70
   |----- 60
  65 |----- 50
   |----- 40
  55 |----- 30
   |----- 20
  45 |----- 10
   |----- 0 (does not increase your risk assessment)
```

When you are satisfied with your answers to this questions, please turn to question two on page 6 of this folder and follow the instructions.
In your envelope you will find a set of cards labeled 'Question Two'. On each card you will find a brief description of a situation that other auditors have classified as an aspect of a company's financial reporting structure that is intended to prevent or detect instances of fraudulent financial reporting. Depending on which version of the Mr. Chip Company workpapers you read, you may recognize the situation described on two of the five cards.

With one exception, this task is procedurally equivalent to the task in question one of this folder. The only difference is the scale is reversed. Instead of rating how much the situations represented on each card would cause you to increase your assessment of the risk that a typical company's financial statements are affected by fraudulent financial reporting, you will rate how much the situations cause you to decrease your assessment of the risk that a typical company's financial statements are affected by such fraud.

All ratings will be made on the one-hundred point scale on the opposite page. If the following instructions do not provide enough detail, please refer to the instructions to question one of this folder.

**Step 1:** Order the cards.
Order the cards in relation to the degree to which the situation posed on each card would decrease your assessment of the risk that a typical company's financial statements are affected by fraudulent financial reporting. The card on the top of the stack should describe the situation that most decreases your risk assessment. The card with the description of the situation that decreases your risk assessment less than the situations on the other four cards should be on the bottom of the stack.

**Step 2:** Assign the top card the value one-hundred.

**Step 3:** Rate the remaining four cards on the one-hundred point scale.
QUESTION TWO
(continued)

Please use this scale to record your answers to question two:

- 0 (does not decrease your risk assessment)
- 5 - 10
- 15 - 20
- 25 - 30
- 35 - 40
- 45 - 50
- 55 - 60
- 65 - 70
- 75 - 80
- 85 - 90
- 95 - 100 (____ letter of top card)

When you are satisfied with your answers to this question, please turn to the debriefing questions starting on page 9 of this folder.
DEBRIEFING QUESTIONS

1. In what geographic area is your office? ____________________________
   (be as specific as you wish, e.g., the city)

2. How many months of experience have you had in
a. a CPA firm? _______ months
b. the audit department of a CPA firm? ________ months
c. business not including a CPA firm? ________ months

3. What is your present department?
   a. audit ______
   b. tax ______
   c. MCG ______
   d. ESG ______
   e. other (please specify) ______________________

4. What is your highest level of education?
   a. Bachelor's degree ______
   b. some graduate work ______
   c. MBA ______
   d. MAcc ______
   e. JD ______
   f. other (please specify) ______________________

5. What percentage of your public accounting experience has been spent
   in the following areas?
   a. Manufacturing _____
   b. Health Care _____
   c. High Tech _____
   d. Retailing _____
   e. Government _____
   f. Construction _____
   g. Extractive Industries _____
   h. Banking and Finance _____
   i. Insurance _____
   j. Service Industries _____
   k. other (please specify) _____

6. What percentage of your clients have gross sales in the following
   ranges?
   a. less than 5 million  _____
   b. between 5 million and 30 million _____
   c. between 30 million and 50 million _____
   d. between 50 million and 100 million _____
   e. between 100 million and 500 million _____
   f. greater than 500 million _____

7. What is your best guess of the average annual rate of fraudulent
   financial reporting (both detected and undetected) per 1000
   companies
   a. in companies whose exposure to such frauds is the same as your
      assessment of Mr. Chip Company's exposure after reading
      the history and background narrative? _______ per thousand
   b. in all types of companies? _______ per thousand
DEBRIEFING QUESTIONS
(continued)

8. Have you been involved in an engagement where
   a. fraudulent financial reporting was suspected? yes ___ no ___
   b. an instance of fraudulent financial reporting
      was detected? yes ___ no ___

9. Did you explicitly consider the risk that Mr. Chip Company's 1986
   financial statements are affected by fraudulent financial
   reporting
   a. while reading the history and background document?
      yes ____ no ____
   b. while reading the audit workpapers?
      yes ____ no ____

10. Did you find this task realistic? yes _____ no _____
    If no, please explain

11. What do you think is the purpose of this study?

12a. How long did it take you to reach this point? ______ minutes
    b. Did you feel rushed? yes ____ no ____
Please insert all materials in your envelope. Once I have had an opportunity to make a few closing remarks, bring your envelope to the front desk. Please do not leave early as this will disturb others that have not finished. If you do finish earlier than others, do not change your answers.

Please do not talk about this project with anyone not attending this conference as this may jeopardize subsequent analyses should anyone you talk to be involved in a later phase of this study. All responses are strictly confidential.

Thank you for your cooperation. If you would like to know the results of this study, leave one of your business cards with me. Thanks!
Effective the first day of the current audit period, the management compensation package was changed to reflect the company's increased emphasis on achieving budgeted targets. Basically, the mix of salary and bonus has been changed to place a much greater emphasis on bonuses tied to responsibility-center financial performance.
Delays in several accounting department activities are caused both by cutbacks in resources allocated to the department and increases in the volume of activity. Many employees in the accounting department complain that staffing levels are at a point where day-to-day operations are barely manageable. These are not considered idle comments. The prevailing corporate philosophy is that the company is a manufacturing company that will spend its money on upgrading its manufacturing facilities before it spends money on a bunch of "bean counters".

Top management basically has a hands-off management style. Top management tends to rely on others to pay attention to the details. They say they focus on the "big picture". But when you meet with them, it becomes painfully obvious that the big picture does not include a thorough review of financial information. When you talk to individual members of top management, you get the feeling that the right hand does not know what the left hand is doing.
Management is predicting a narrow margin of compliance with certain restrictive debt covenants. Even though a waiver has always been obtained, the lender threatened to not be so lenient in the current year. If a waiver is needed and is not granted, a significant amount of long-term debt will become current. Management is taking the lender's threat seriously.

Strong competition has led to increasing marketing costs and lower margins. Nevertheless, management is determined to maintain its market share and is preparing to introduce new products, the initial cost of which will put downward pressure on the company's profits.
Early in the current audit period, a company replaced its chief internal auditor with a better-trained, more aggressive individual. The new chief internal auditor had five years' experience on the audit staff of a Big 8 accounting firm and six years' experience in the internal audit department of a fortune 500 company. At the same time, resources to hire additional internal audit personnel were approved. By the beginning of the second quarter, all additional personnel were hired. The effectiveness of the internal audit function has significantly improved.
As suggested by the audit committee of the board of directors, top management initiated a program where it formally assesses the company's exposure to fraudulent financial reporting. After consulting with the chief internal auditor, this risk assessment is documented biannually in a report from the CEO to the audit committee. The audit committee, the external auditors, and the chief and internal auditor formally review the company's procedures for assessing its exposure to fraudulent financial reporting, management's documentation of this process, and management's responses to identified areas of exposure. This program is taken seriously.

The company has begun a job rotation program at the middle and lower management levels. Upper-level management are involved whenever possible. The assignments outside the employee's traditional area can last up to four months.
D

The company has adopted a written code of corporate conduct, signaling to all employees standards for conducting the company's affairs. Employees are asked to review the code annually. The code is updated when necessary.

E

Top management no longer harbors the philosophy that manufacturing/sales should be emphasized at all costs, even if service departments like accounting have to "make do" with what they have. Over the last eight quarters, the accounting department has received substantial increases in funding.